Estimation

Econometría. ADE.

Estimation

- We assume we have a sample of size T of:
  - The dependent variable (y)
  - The explanatory variables (x_1,x_2, x_3, …, x_k), with x_1 being an intercept (a column vector of 1s)
- We also have a vector of unknown parameters $\beta=(\beta_1, \beta_2, \ldots, \beta_k)'$
- We have a perturbation vector (u) of dimension T
- A set of hypothesis about the relationship between X, β and u.

Estimation

In these circumstances, we can link the dependent and the explanatory variables as follows:

$$Y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t$$

Matrix form: $Y = X\beta + u$

Compact form: $Y_t = x_t' \beta + u_t$

Our key problem is to provide value to the components of the vector β

Estimation

To assign values to the parameters, we have some options.

-We can invent them.

-We can phone Rappel and ask him for the more appropriate values according to his “view”.

- We can use a random procedure.

- We can estimate them.

Estimation

Estimation theory is a branch of statistics that deals with providing values to a set of parameters based on measured/empirical data that has a random component.

An **estimator** attempts to approximate the unknown parameters using the measurements.

Estimation

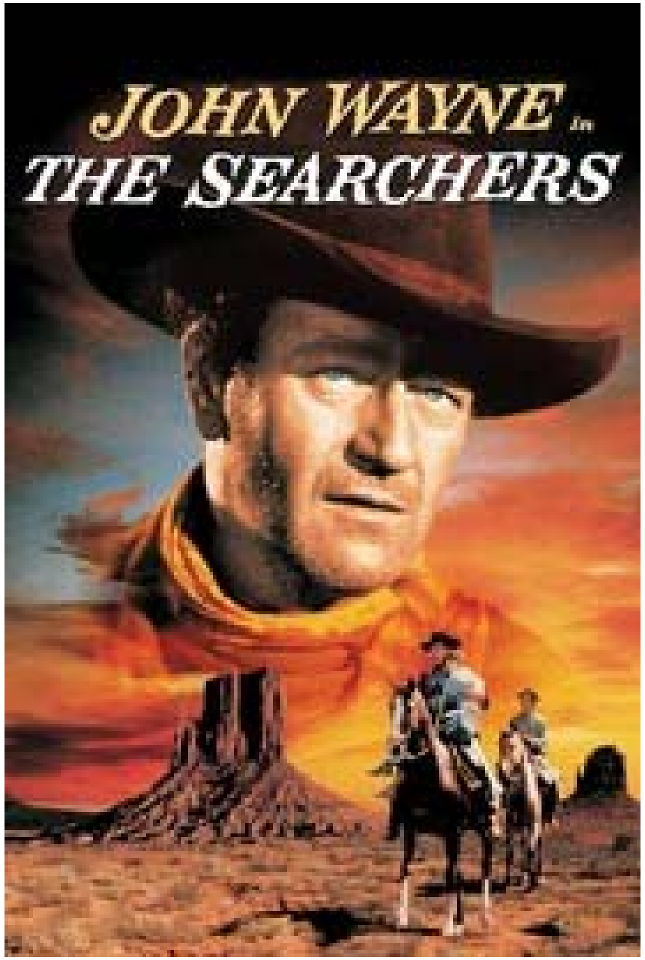
Estimation methods

- Maximum likelihood
- Ordinary Least Squares
- Non-linear methods
- GMM

Which one will we use along the course?

Estimation

- Do you know this John Ford's film?
- There is a very famous scene where cowboys escape from redskins by crossing the Rio Grande river. (28:05)
- Do they use a least squares principle?



Estimation

We have the following the relationship:

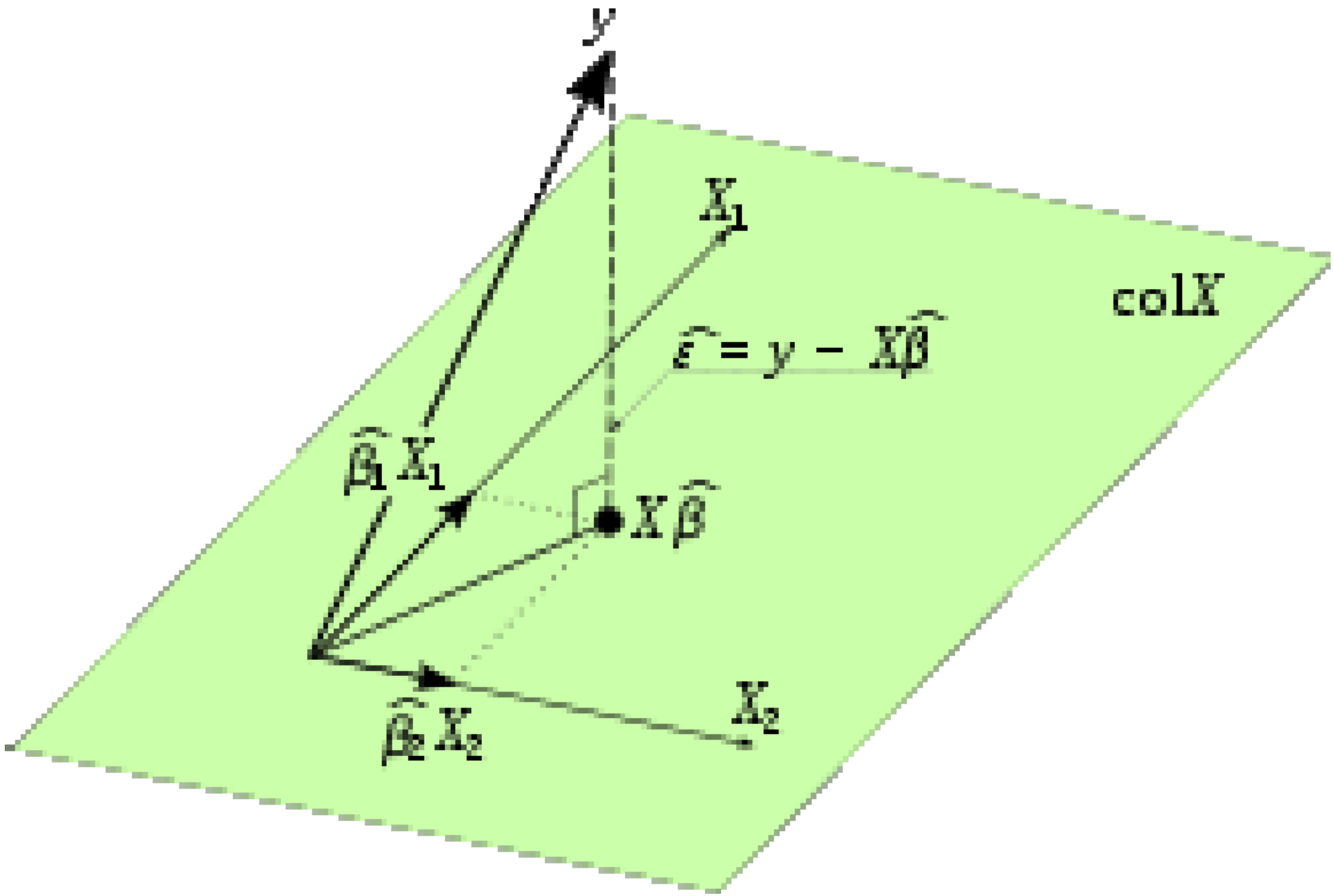
Y = X β + u

Thus, we can express it as follows:

$$Y = X\hat{\beta} + \hat{u} = \hat{Y} + \hat{u}$$

$$\hat{Y}\ is\ the\ estimated\ part\ of\ Y$$

$$\hat{u}\ is\ the\ residual\ part$$



y
X1
colX
ε̂ = y − Xβ̂
β̂1 X1
Xβ̂
X2
β̂2 X2

Estimation

The OLS estimator

$$\hat{\beta} = \arg\min_{b} \left\| y - Xb \right\|$$

Estimation

- Then,
- The projection matrix is: $P_X = X(X'X)^{-1}X'$
- The residual matrix is: $M = I - P_X = I - X(X'X)^{-1}X'$
- P_X and M are orthogonal
- P_X is symmetric ($P_X = P'_X$)
- P_X is idempotent ($P_X P_X = P_X$)

Estimation

- Consequently, we have that:

$$\hat{Y} = P_X Y = X\left(X'X\right)^{-1} X'Y = X\hat{\beta}$$

$$and$$

$$\hat{u} = M\,Y = \left(I - P_X\right)Y = \left[I - X\left(X'X\right)^{-1} X'\right]Y$$

Estimation

- By simply comparing the equation, we obtain the expression of the ordinary least squares (OLS) estimator:

$$\hat{\beta} = \left(X'X \right)^{-1} X'Y$$

Estimation

We can alternatively obtain this same result by simply applying the least squares principle:

- We need to get the combination of values of the estimation vector that minimizes the sum of the squared residuals.
- Analytically, this implies to solve this:

$$min\, \hat{u}'\hat{u} = \min (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

Estimation

$$min\,\hat{u}'\hat{u} = \min\,(Y - X\,\hat{\beta})'(Y - X\,\hat{\beta})$$

$$\frac{\partial \hat{u}'\hat{u}}{\partial \hat{\beta}} = 0 = -2\,X\,'y + 2\,X\,'X\,\hat{\beta}$$

$$X\,'y = X\,'X\,\hat{\beta}$$

$$\hat{\beta} = \left(X\,'X\right)^{-1} X\,'Y$$

Properties of the LS Estimator

We have obtained the LS estimator. But, is good enough this estimator?

Properties

- Unbiased
- BLUE
- Consistent
- Efficient

Properties of the LS Estimator

It is advisable to express the LS estimator as follows:

$$\hat{\beta} = (X'X)^{-1} X'y = (X'X)^{-1} X'(X\beta + u) =$$

$$= (X'X)^{-1} X'X\beta + (X'X)^{-1} X'u =$$

$$= \beta + (X'X)^{-1} X'u$$

Properties of the LS Estimator

Thus, we have that:

$$E(\hat{\beta}) = E(\beta) + E\left[\left(X'X\right)^{-1} X'u\right] =$$

$$= \beta + \left(X'X\right)^{-1} X'E\left[u\right] =$$

$$= \beta$$

The OLS estimator vector is unbiased

Properties of the LS Estimator

The variance of the estimator vector is

$$\mathrm{Var}\left(\hat{\beta}\right)=\mathrm{E}\left\{\left[\hat{\beta}-\mathrm{E}\left(\hat{\beta}\right)\right]\left[\hat{\beta}-\mathrm{E}\left(\hat{\beta}\right)\right]'\right\}=\mathrm{E}\left\{\left[\beta+\left(X'X\right)^{-1}X'u-\beta\right]\left[\beta+\left(X'X\right)^{-1}X'u-\beta\right]'\right\}=$$

$$=\mathrm{E}\left\{\left[\left(X'X\right)^{-1}X'u\right]\left[\beta+\left(X'X\right)^{-1}X'u\right]'\right\}=\mathrm{E}\left[\left(X'X\right)^{-1}X'u\,u'X\left(X'X\right)^{-1}\right]=$$

$$=\left(X'X\right)^{-1}X'\,\mathrm{E}\left[u\,u'\right]X\left(X'X\right)^{-1}=\left(X'X\right)^{-1}X'\,\sigma^2\,I_T\;X\left(X'X\right)^{-1}=\sigma^2\left(X'X\right)^{-1}X'X\left(X'X\right)^{-1}=$$

$$=\sigma^2\left(X'X\right)^{-1}$$

Properties of the LS Estimator

The OLS estimator is BLUE

- Best Linear Unbiased Estimator
- This implies that it has the lowest variance, as compared to other unbiased, linear estimators.
- Gauss-Markov theorem helps us to prove it.

Properties of the LS Estimator

The LS estimator is consistent if we can prove that:

$$\lim_{T\to\infty} \Pr\left[\left|\beta - \hat{\beta}\right| \leq \varepsilon\right] = 1$$

Or equivalently

$$p\lim \hat{\beta} = \beta$$

Properties of the LS Estimator

The use of plim's can help us to prove consistency, given that:

$$p\lim \hat{\beta} = p\lim\left[\beta + \left(X'X\right)^{-1} X'u\right] =$$

$$= p\lim \beta + p\lim\left(X'X\right)^{-1} X'u =$$

$$= \beta + p\lim \frac{\left(X'X\right)^{-1}}{T} p\lim \frac{X'u}{T} =$$

$$= \beta + \Sigma_X \times 0 =$$

$$= \beta$$

Properties of the LS Estimator

We can employ an alternative method to prove consistency. An estimator is consistent if:

• It is asymptotically unbiased

$$\lim_{T \to \infty} E\left(\hat{\beta}\right) = \beta$$

• Its variance goes to zero when the sample grows to infinity.

$$\lim_{T \to \infty} Var(\hat{\beta}) = 0$$

Properties of the LS Estimator

The first condition is held by simply considering that:

$$\lim_{T\to\infty} E\left(\hat{\beta}\right) = \lim_{T\to\infty} \beta = \beta$$

Properties of the LS Estimator

The second condition also holds.

$$\lim_{T\to\infty} Var(\hat{\beta}) = \lim_{T\to\infty} \left(X'X \right)^{-1} =$$

$$= \lim_{T\to\infty} \frac{\sigma^2}{T} \left(\frac{X'X}{T} \right)^{-1} = \lim_{T\to\infty} \frac{\sigma^2}{T} \lim_{T\to\infty} \left(\frac{X'X}{T} \right)^{-1} =$$

$$= 0 \times \Sigma_X = 0$$

Properties of the LS Estimator

We can also follow a very intuitive approach. If an estimator is consistent, we have that:

- The estimator asymptotically collapses towards a value, which is the true value of the parameter.

Thus, the differences between estimator and true value of the parameter vanishes.

- It is also true that the higher the sample size, the more information available to estimate the model.

Properties of the LS Estimator

Let us consider that the observations of a dependent variable are generated by the following model:

$$y_t = 2 + u_t \quad t = 1, 2, ..., T$$

T={10, 50, 100, 500, 1000}

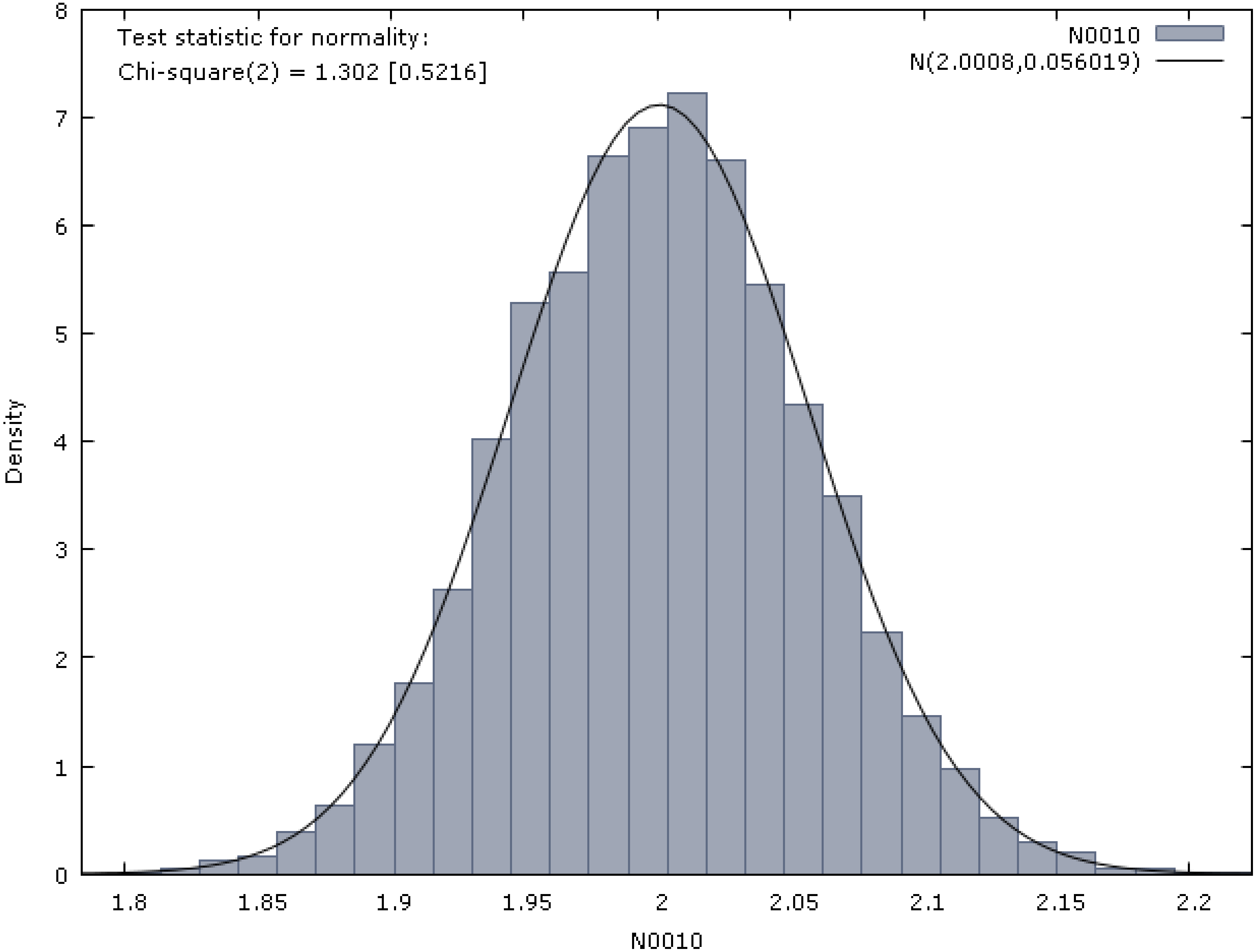
The perturbation is generated by a N(0,1) distribution

Properties of the LS Estimator

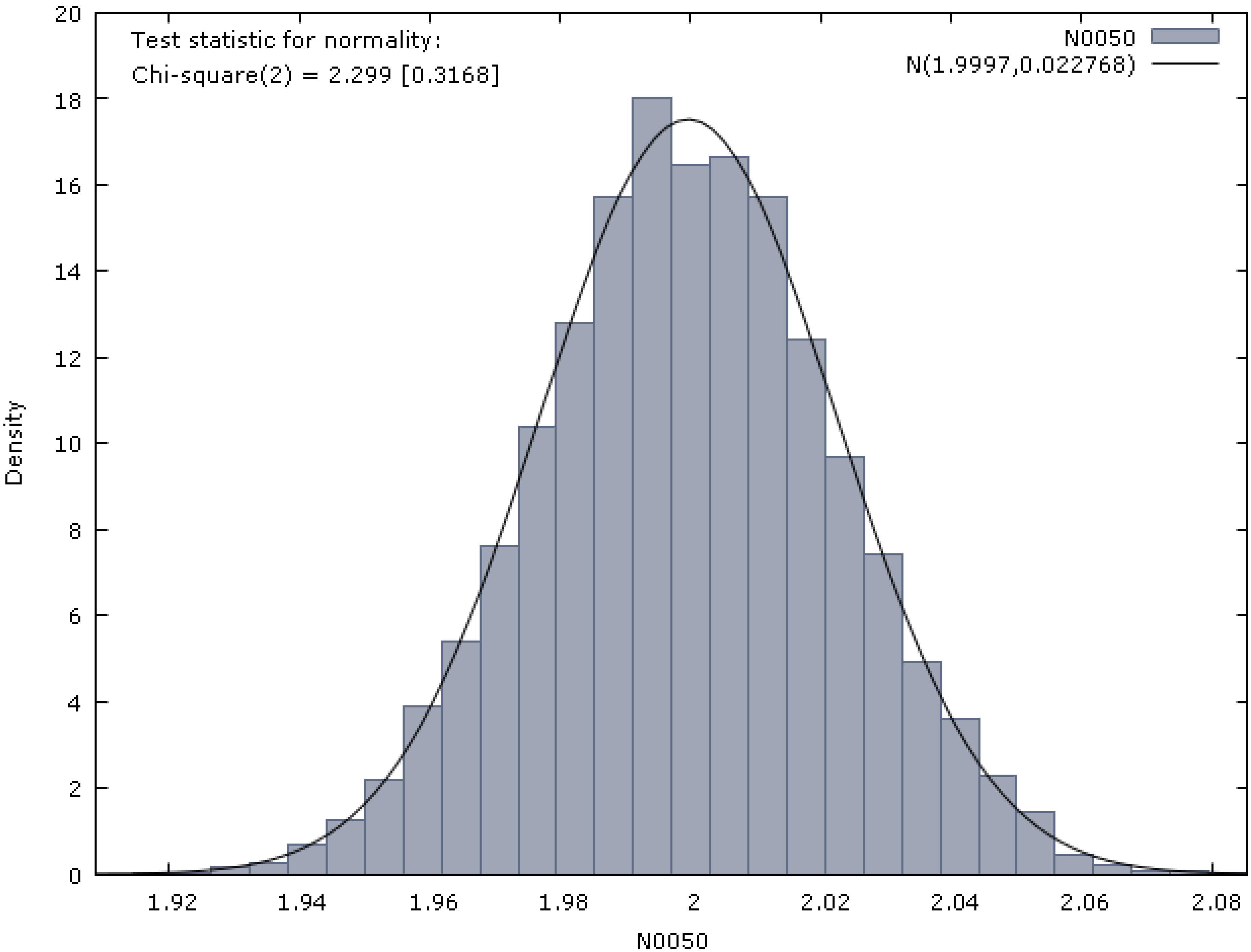
Then, we estimate by OLS the following model

$$y_t = \alpha + u_t \quad t = 1, 2, ..., T$$

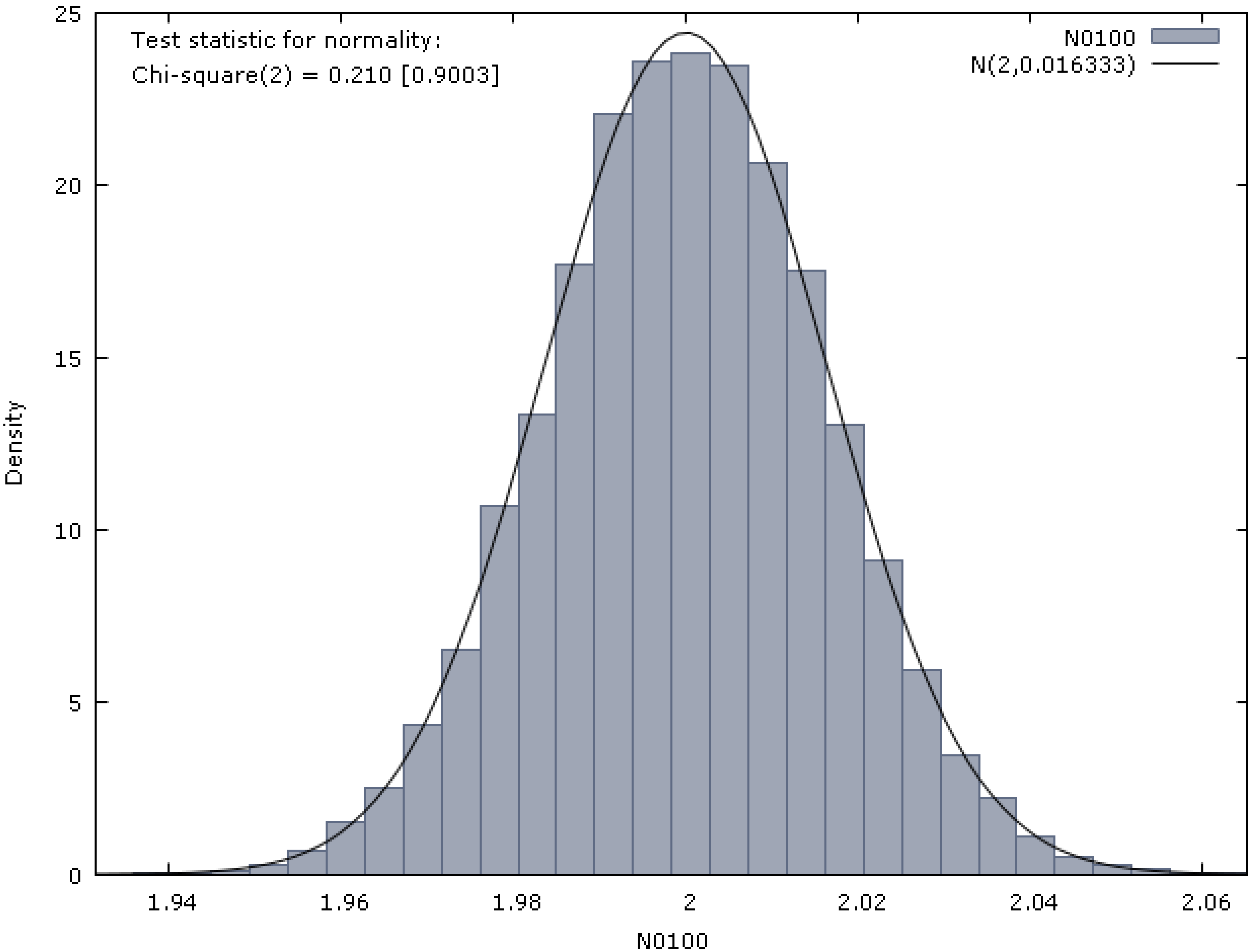
The results we have obtained are reported in next Figures.



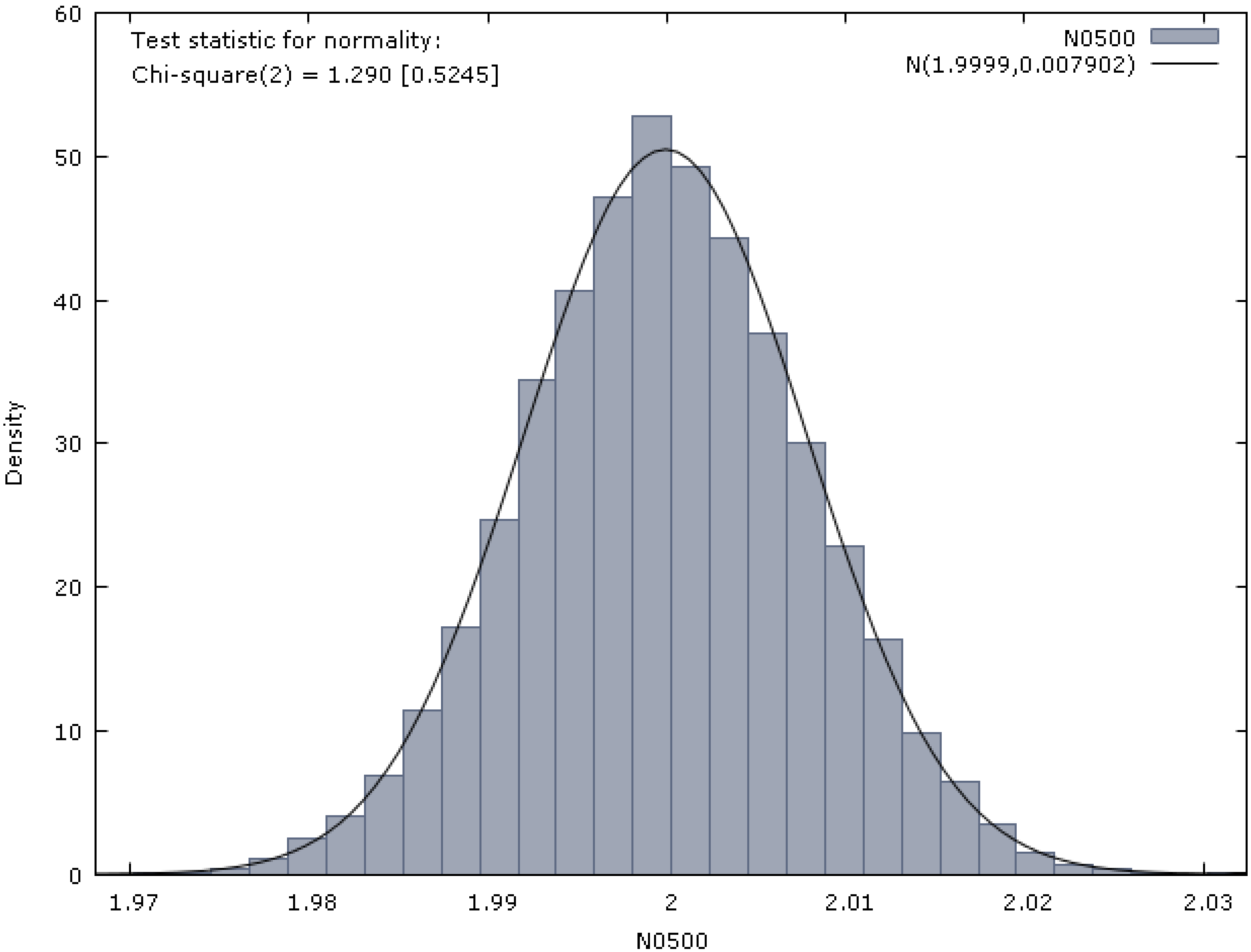
Test statistic for normality:
Chi-square(2) = 1.302 [0.5216]
N0010
N(2.0008,0.056019)
Density
N0010
0
1
2
3
4
5
6
7
8
1.8
1.85
1.9
1.95
2
2.05
2.1
2.15
2.2

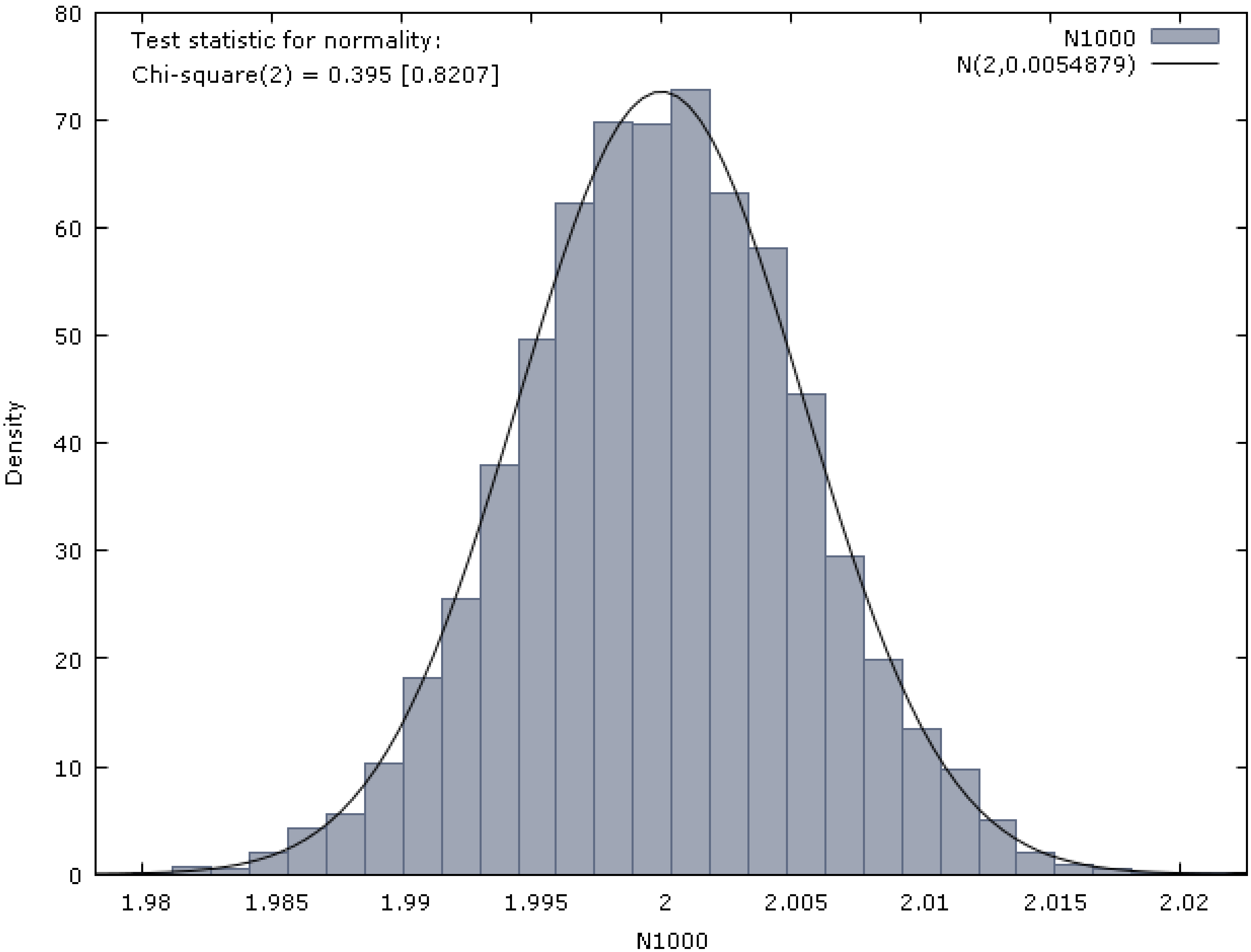


Test statistic for normality:
Chi-square(2) = 2.299 [0.3168]
N0050
N(1.9997,0.022768)
Density
N0050
0
2
4
6
8
10
12
14
16
18
20
1.92
1.94
1.96
1.98
2
2.02
2.04
2.06
2.08



Test statistic for normality:
Chi-square(2) = 0.210 [0.9003]
N0100
N(2,0.016333)
Density
N0100
25
20
15
10
5
0
1.94
1.96
1.98
2
2.02
2.04
2.06


Test statistic for normality:
Chi-square(2) = 1.290 [0.5245]
N0500
N(1.9999,0.007902)
Density
N0500
0
10
20
30
40
50
60
1.97
1.98
1.99
2
2.01
2.02
2.03


Test statistic for normality:
Chi-square(2) = 0.395 [0.8207]
N1000
N(2,0.0054879)
Density
N1000
0
10
20
30
40
50
60
70
80
1.98
1.985
1.99
1.995
2
2.005
2.01
2.015
2.02

Properties of the OLS estimator. Efficiency

An estimator is efficient if it has the lowest variance, compare to other unbiased estimators.

$$Var(\hat{\beta}) \leq Var(\tilde{\beta})$$

Properties of the OLS estimator. Efficiency

To prove it:

• We should impose that the perturbation follows a normal distribution

• We should then obtain the Cramer-Rao bound.

$$Var\left(\tilde{\theta}\right) \geq \frac{1}{I\left(\theta\right)} = \frac{1}{E\left[\left(\frac{\partial}{\partial\theta}\ell\left(X;\theta\right)\right)^2\right]}$$

Properties of the OLS estimator. Efficiency

- We should compare the Cramer-Rao bound with the variance of the OLS estimator.
- For the OLS estimator vector, the Cramer-Rao bound is $\sigma^2 (X'X)^{-1}$.
- It coincides with the covariance matrix of the OLS estimator vector.
- Consequently, the OLS estimator is efficient.

OLS estimator of the variance of the perturbation.

- If we assume that the perturbation vector follows a $N(0, \sigma^2 I)$ distribution, we have that the vector *y* follows a $N(X\beta, \sigma^2 I)$ distribution.
- We already know how the vector β can be estimated
- We need now to get an estimator of the variance of the perturbation (σ^2)
- We can use the OLS approach

OLS estimator of the variance of the perturbation.

The OLS estimator can be defined as follows:

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-k} = \frac{\left(y - X\hat{\beta}\right)'\left(y - X\hat{\beta}\right)}{T-k}$$

With $\hat{u}'\hat{u}$ being the sum of the squared residuals (SSR)

OLS estimator of the variance of the perturbation.

This estimator shows the following properties:

- It is unbiased
- It is NOT linear
- It is consistent
- It is NOT efficient

Maximum Likelihood estimator

The ML estimation selects the set of values of the model parameters that maximizes the likelihood function.

- Assuming all the previous hypothesis, the joint density function of $(y_1, y_2, \ldots, y_T)$ can be expressed as follows:

$$f\left(y / X; \beta, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{T}{2}} \exp\left[-\frac{1}{2}\left(y - X\beta\right)'\left(y - X\beta\right)\right]$$

Maximum Likelihood estimator

Then, the likelihood function can be stated as follows:

$$\mathfrak{L}\left(\beta,\sigma^2; y / X \right) = \left(2\pi\sigma^2\right)^{-\frac{T}{2}} \exp\left[-\frac{1}{2}(y - X\beta)'(y - X\beta)\right]$$

Which is the difference with respect the joint density function?

Maximum Likelihood estimator

To obtain the ML estimators, we should first the log-likelihood function

$$\ell\left(\beta,\sigma^2\right) = -\frac{T}{2}\ln\left(2\pi\right) - \frac{T}{2}\ln\left(\sigma^2\right) - \frac{1}{2}\left(y - X\beta\right)'\left(y - X\beta\right)$$

Then, we should obtain the values of the parameters that maximize this function

$$\frac{\partial\ell}{\partial\beta} = 0$$

$$\frac{\partial\ell}{\partial\sigma^2} = 0$$

Maximum Likelihood estimator

The solutions to this optimization problem are:

$$\tilde{\beta} = \hat{\beta} = (X'X)^{-1} X'Y$$

The ML estimator and the OLS estimator coincide.

Both have the same properties.

Maximum Likelihood estimator

The ML estimator of the variance of the perturbation is:

$$\tilde{\sigma}^2 = \frac{\left(y - X\hat{\beta}\right)'\left(y - X\hat{\beta}\right)}{T}$$

It is consistent

It is not unbiased, linear, efficient

It has asymptotical good properties (it is a MLE)