Linear Econometrics for Finance
Lecture 3

Gianni De Nicolò

Fall II, 2019

Overview and reading assignments

1. The multivariate linear regression (MLR) model
(W, Ch.3, 3-1-3-2)
2. Statistical properties of the OLS estimators
(W, Ch.2, 3.3-3.5)
3. The MLR model in matrix form
(W, Appendix D-3,D-4, D-6, E-1,E-2)

1.1 The multiple linear regression (MLR) model

Variable y is explained in terms of variables $x_1, x_2, \ldots, x_k$

Intercept

Slope parameters

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

Dependent variable,
explained variable,
response variable,...

Independent variables,
explanatory variables,
regressors,...

Error term,
disturbance,
unobservables,...

Motivations for multiple regression

1. Incorporate more explanatory factors into the model
2. Explicitly hold fixed other factors that otherwise would be in the residual
3. Allow for more flexible functional forms

1.2 OLS estimation

Random sample

$$\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) : \; i = 1, \ldots, n\}$$

Regression residuals

$$\widehat{u}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \widehat{\beta}_2 x_{i2} - \ldots - \widehat{\beta}_k x_{ik}$$

Minimize the sum of squared residuals

$$\min \sum_{i=1}^{n} \widehat{u}_i^2 \quad \rightarrow \quad \widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_k$$

First order conditions: $k+1$ linear equations in $k+1$ unknown:

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - ..... - \beta_k x_{ik}) = 0$$

$$\sum_{i=1}^{n} x_{ij} (y_i - \beta_0 - \beta_1 x_{i1} - ..... - \beta_k x_{ik}) = 0; j = 1, 2, ..k$$

1.3 Properties of OLS

Fitted values and residuals

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \widehat{\beta}_2 x_{i2} + \ldots + \widehat{\beta}_k x_{ik} \qquad \widehat{u}_i = y_i - \widehat{y}_i$$

Fitted or predicted values

Residuals

Algebraic properties

$$\sum_{i=1}^{n} \widehat{u}_i = 0 \qquad \sum_{i=1}^{n} x_{ij}\widehat{u}_i = 0 \qquad \bar{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \bar{x}_1 + \ldots + \widehat{\beta}_k \bar{x}_k$$

Deviations from regression line sum up to zero

Covariance between deviations and regressors are zero

Sample averages of y and of the regressors lie on regression line

Goodness of fit

$$R^2 \equiv SSE/SST = 1 - SSR/SST$$

Notice that R-squared can only increase if another explanatory variable is added to the regression

Note: R^2 never decreases and usually increases by adding explanatory variables to the regression. This is problematic for the use if R^2 as a variable selection criterion: WHY? *The importance of a variable is gauged by a non-zero partial effect on the dependent variable in the population.*

1.4 Partialling out (Frisch-Waugh theorem)

The estimated coefficients in a multiple regression can be obtained in two steps:

1) Regress the explanatory variable on all other explanatory variables
2) Regress the dependent variable on the residuals from this regression

Example with dependent variable y and two explanatory variables x_1 and x_2:
Step 1: regress x_1 on x_2, and obtain residuals r_{12} ; regress x_2 on x_1 and obtain residuals r_{21}
Step 2: regress y on r_{12} to obtain $\hat{\beta}_1$, and regress y on r_{21} to obtain $\hat{\beta}_2$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} r_{i,12} y_i}{\sum_{i=1}^{n} r_{i,12}^2} ; \hat{\beta}_2 = \frac{\sum_{i=1}^{n} r_{i,21} y_i}{\sum_{i=1}^{n} r_{i,21}^2}$$

- The residual from the first regression is the part of the explanatory variable that is uncorrelated with the other explanatory variables. It represents the impact of one variable after the impact of all other variables has been *partialled out*.
- The slope coefficient of the second regression therefore represents the isolated effect of the explanatory variable on the dependent variable.

1.5 Assumptions of the MLR model (1-2)

Assumption MLR.1 (Linearity in parameters)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

In the population, the relationship between y and the explanatory variables is linear

Assumption MLR.2 (Random sampling)

$$\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) : \; i = 1, \ldots, n\}$$

The data is a random sample drawn from the population

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i$$

Each data point therefore follows the population equation

Note: the random sampling assumption is valid for a cross-section. For time series and panel data, this assumption has to be modified

1.6 Assumptions of the MLR model (3-4)

Assumption MLR.3 (No perfect collinearity)
In the sample (and therefore in the population), none of the independent variables is constant and there are no exact linear relationships among the independent variables.

Remarks on ML3:
a. ML3 only rules out perfect collinearity/correlation between explanatory variables; imperfect correlation is allowed.
b. If an explanatory variable is a perfect linear combination of other explanatory variables it is superfluous and may be eliminated.
c. Constant variables are also ruled out (collinear with intercept).

Assumption MLR.4 (Zero conditional mean)

$$E(u_i | x_{i1}, x_{i2}, \ldots, x_{ik}) = 0$$

The value of the explanatory variables must contain no information about the mean of the unobserved factors

Remark on MLR4: In a multiple regression model, MLR4 is much more likely to hold because fewer things end up in the error.

Assumption MLR.5 (Homoskedasticity)

All explanatory variables are collected in a random vector

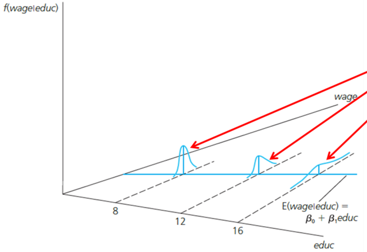
$$Var(u_i|\mathbf{x}_i) = \sigma^2 \quad \text{with} \quad \mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$$

Example: wage equation

$$Var(u_i|educ_i, exper_i, tenure_i) = \sigma^2$$

This assumption may also be hard to justify in many cases

Heteroskedasticity



The variance of the unobserved determinants of wages increases with the level of education

Including irrelevant variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

= 0 in the population

No problem because $E(\widehat{\beta}_3) = \beta_3 = 0$.

However, including irrevelant variables may increase sampling variance.

Omitting relevant variables: the simple case

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$ ← True model (contains x_1 and x_2)

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{u}$$ ← Estimated model (x_2 is omitted)

Omitted variable bias

$$x_2 = \delta_0 + \delta_1 x_1 + v$$ ← If x_1 and x_2 are correlated, assume a linear regression relationship between them

$$\Rightarrow \quad y = \beta_0 + \beta_1 x_1 + \beta_2(\delta_0 + \delta_1 x_1 + v) + u$$

$$= \boxed{(\beta_0 + \beta_2 \delta_0)} + \boxed{(\beta_1 + \beta_2 \delta_1)} x_1 + \boxed{(\beta_2 v + u)}$$

If y is only regressed on x_1 this will be the estimated intercept

If y is only regressed on x_1, this will be the estimated slope on x_1

error term

1.9 Omitted variables bias more generally

Example:

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + u$$

$$abil = \delta_0 + \delta_1 educ + v$$

Will both be positive

$$wage = (\beta_0 + \beta_2\delta_0) + \boxed{(\beta_1 + \beta_2\delta_1)} educ + (\beta_2 v + u)$$

The return to education β_1 will be <u>overestimated</u> because $\beta_2\delta_1 > 0$. It will look as if people with many years of education earn very high wages, but this is partly due to the fact that people with more education are also more able on average.

All estimated coefficients will be biased

More general case

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

True model (contains x_1, x_2, and x_3)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + w$$

Estimated model (x_3 is omitted)

No general statements possible about direction of bias

Analysis as in simple case if one regressor is uncorrelated with others

1.10 Example

$$wage = \beta_0 + \beta_1 educ + \beta_2 abil + u$$

Will both be positive

$$abil = \delta_0 + \delta_1 educ + v$$

$$wage = (\beta_0 + \beta_2\delta_0) + \boxed{(\beta_1 + \beta_2\delta_1)} educ + (\beta_2 v + u)$$

The return to education β_1 will be overestimated because $\beta_2\delta_1 > 0$. It will look as if people with many years of education earn very high wages, but this is partly due to the fact that people with more education are also more able on average.

When is there no omitted variable bias?

If the omitted variable is irrelevant or uncorrelated

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 abil + u$$

If exper is approximately uncorrelated with educ and abil, then the direction of the omitted variable bias can be as analyzed in the simple two variable case.

2.1 Statistical properties of the OLS estimators

Theorem 3.1 - Unbiasedness

$$MLR.1 - MLR.4 \implies E(\hat{\beta}_j) = \beta_j, j = 0, 1, 2..., k$$

Theorem 3.2 - Variances

Variance of the error term

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}, \quad j = 1, \ldots, k$$

Total sample variation in explanatory variable x_j:

$$\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$$

R-squared from a regression of explanatory variable x_j on all other independent variables (including a constant)

- A high σ^2 increases the sampling variance because there is more "noise" in the equation, and does not decrease with sample size.
- In general, more sample variation leads to more precise estimates.

2.2 Variances in misspecified models

The choice of whether to include a particular variable in a regression can be made by analyzing the trade-off between bias and variance.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \longleftarrow \text{True population model}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \longleftarrow \text{Estimated model 1}$$

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 \longleftarrow \text{Estimated model 2}$$

It might be the case that the likely omitted variable bias in the misspecified model 2 is overcompensated by a smaller variance.

Conclusion: Do not include irrelevant regressors

$$\beta_2 = 0 \;\Rightarrow\; E(\hat{\beta}_1) = \beta_1, E(\tilde{\beta}_1) = \beta_1, Var(\tilde{\beta}_1) < Var(\hat{\beta}_1)$$

Trade off bias and variance; Caution: bias will not vanish even in large samples

$$\beta_2 \neq 0 \;\Rightarrow\; E(\hat{\beta}_1) = \beta_1, E(\tilde{\beta}_1) \neq \beta_1, Var(\tilde{\beta}_1) < Var(\hat{\beta}_1)$$

2.3 Estimation of σ^2

The estimate:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k-1}$$

The true sampling variation of the estimated β_j $\longrightarrow$ $sd(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)} = \sqrt{\sigma^2 / \left[SST_j(1-R_j^2)\right]}$

Plug in $\hat{\sigma}^2$ for σ^2 here

The estimated sampling variation of the estimated β_j $\longrightarrow$ $se(\hat{\beta}_j) = \sqrt{\widehat{Var}(\hat{\beta}_j)} = \sqrt{\hat{\sigma}^2 / \left[SST_j(1-R_j^2)\right]}$

Theorem 3.3 - Unbiasedness of $\hat{\sigma}^2$

$$MLR.1 - MLR.5 \implies E(\hat{\sigma}^2) = \sigma^2$$

2.4 Efficiency of the OLS estimator

- Under assumptions MLR.1 - MLR.5, OLS is unbiased. However, under these assumptions there may be many other estimators that are unbiased.
- Which unbiased estimator is the one with the smallest variance?
- Consider linear estimators, i.e. estimators linear in the dependent variable.

$$\tilde{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i$$

May be an arbitrary function of the sample values of all the explanatory variables; the OLS estimator can be shown to be of this form

Theorem 3.4 (Gauss-Markov). Under assumptions MLR.1 - MLR.5, the OLS estimators are the **best linear unbiased estimators (BLUEs)** of the regression coefficients, i.e.

$$Var(\widehat{\beta}_j) \leq Var(\tilde{\beta}_j) \quad j = 0, 1, \ldots, k$$

$$\text{for all } \tilde{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i \quad \text{for which} \quad E(\tilde{\beta}_j) = \beta_j, j = 0, \ldots, k.$$

Note: OLS is the best estimator only if MLR.1 – MLR.5 hold; if there is heteroskedasticity for example, there are better estimators.

3.1 The MLR in matrix form (W, Appendix E-1)

1. y is a $n \times 1$ vector of observations of the dependent variable
2. X is a $n \times k$ matrix of regressors (n observations and $k+1$ regressors (including the constant)
3. β is a k vector of parameters
4. u is a $n \times 1$ vector of errors
5. $Var(u|X) = \sigma^2 I_n$, where I_n an identity matrix of dimension n

The model can be written as:

$$y = X\beta + u$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & \dots & x_{k2} \\ 1 & x_{13} & \dots & x_{k3} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1n} & \dots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \dots \\ u_n \end{bmatrix}$$

3.2 The MLR in matrix form (cont.)

Equivalently,

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \dots + \beta_k x_{k1} \\ \beta_0 + \beta_1 x_{12} + \dots + \beta_k x_{k2} \\ \beta_0 + \beta_1 x_{13} + \dots + \beta_k x_{k3} \\ \dots \\ \beta_0 + \beta_1 x_{1n} + \dots + \beta_k x_{kn} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \dots \\ u_n \end{bmatrix},$$

The error vector satisfies $E(u|X) = 0$.
$Var(u|X) = \sigma^2 I_n$ is the variance-covariance matrix given by

$$Var(u|X) = E(uu') = \sigma^2 I_n = \begin{bmatrix} \sigma^2 & 0 & .. & .. & 0 \\ 0 & \sigma^2 & 0 & .. & .. \\ .. & 0 & \sigma^2 & 0 & .. \\ .. & .. & \dots & \dots & \dots \\ 0 & .. & .. & 0 & \sigma^2 \end{bmatrix}.$$

3.3 Linear independence and rank of a matrix (W, Appendix D-3)

Linear Independence. A set of $n \times 1$ vectors $\{x_1, x_2, , ..., x_r\}$ is **linearly independent** if, and only if,

$$a_1 x_1 + a_2 x_2 + ... + a_r x_r = 0 \tag{1}$$

where $(a_1, a_2, ..., a_r)$ are scalars (constants). Independence implies that $(a_1 = 0, a_2 = 0, ..., a_r = 0)$.)

If (1) holds for a set of scalars that are not all zeros, then the set $\{x_1, x_2, , ..., x_r\}$ is **linearly dependent** (one or more vectors in the set can be written as linear combinations of the other vectors).

Rank of a matrix. Let A be a $n \times m$ matrix

1. $rank(A)$ is the **maximum number of independent columns of** A
2. If $rank(A) = m$, tehn A has full (column) rank
3. Let $n = m = k$ (A is a $k \times k$ square matrix) . If $rank(A) = k$, then A is **invertible** (A^{-1} exists).

3.4 Quadratic forms and positive definite matrices (W, Appendix D-4)

Quadratic form. The **quadratic form** associated with a **symmetric** $n \times n$ matrix A is a real-valued function defined for $n \times 1$ vector x:

$$f(x) = x'Ax = \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=1}^{n}\sum_{j>i}^{n} a_{ij}x_ix_j$$

1. A symmetric matrix A is said to be **positive definite** if for all $n \times 1$ vectors x (except $x = 0$)

$$x'Ax > 0$$

2. If A is positive definite, then A^{-1} exists and it is positive definite.
3. If X is a $n \times k$ matrix and $rank(X) = k$, then $X'X$ is positive definite (and therefore invertible)

3.5 Differentiation of linear and quadratic forms (W, Appendix D-6)

Let a be a $n \times 1$ vector of constant, and x a generic x vector.

Define the linear function:

$$f(x) = a'x = \sum_{i=1}^{n} a_i x_i$$

The derivative of f with respect to x is the vector of partial derivatives:

$$\partial f(x)/\partial x = a'$$

Given a symmetric matrix A, define the quadratic form:

$$g(x) = x'Ax$$

The derivative of g with respect to x is the vector of partial derivatives:

$$\partial g(x)/\partial x = 2x'A$$

3.6 OLS estimation

The vector of residuals is defined as: $\hat{u} = y - Xb$, where b is the vector of parameters to be chosen (We will denote the solution vector $\hat{\beta}$, as usual)

The vector b is chosen to minimize the sum of squared residuals:

$$f(b) = u'u = (y - Xb)'(y - Xb) \tag{2}$$

Using transpose property: $(A + (-)B)' = A' + (-)B'$, we can write (1) as:

$$\begin{aligned} f(b) = (y - Xb)'(y - Xb) = (y' - (Xb)')(y - Xb) = \\ y'y - y'Xb - (Xb)'y + (Xb)'Xb \end{aligned} \tag{3}$$

Using transpose property: $(AB)' = A'B'$, we can write (3) as:

$$\begin{aligned} f(b) = y'y - y'Xb - (Xb)'y + (Xb)'Xb = \\ y'y - y'Xb - y'Xb + b'X'Xb = \\ y'y - 2y'Xb + b'X'Xb \end{aligned} \tag{4}$$

3.7 OLS estimation (cont.)

Differentiating (4) with respect to b and setting the derivative equal to zero, we get the first order condition that the OLS vector $\hat{\beta}$ must satisfy.

$$\partial f(b)/\partial b = -2y'X + 2X'X\hat{\beta} = 0 \implies -2X^{'}(y - X\hat{\beta}) = 0 \implies X'X\hat{\beta} = X'y \tag{5}$$

Assuming that $X'X$ is invertible (non-singular), we can pre-multiply the left hand side of (5) by $(X'X)^{-1}$ to obtain:

$$\hat{\beta} = (X'X)^{-1}X'y \tag{6}$$

Differentiating (5) with respect to $\hat{\beta}$, we obtain the matrix of second partial derivatives, which turns out to be positive (more on this later). Thus, we obtain the second order condition for minimum, which is satisfied:

$$\partial^2 f(\hat{\beta})/\partial\hat{\beta}\partial\hat{\beta}' = 2X'X > 0 \tag{7}$$

3.8 Fitted values and residuals

Fitted values ($\widehat{y}$).

$$
\begin{aligned}
\widehat{y} &= \begin{bmatrix} \widehat{y}_1 \\ \widehat{y}_2 \\ \widehat{y}_3 \\ \dots \\ \widehat{y}_n \end{bmatrix} = \begin{bmatrix} \widehat{\beta}_0 + \widehat{\beta}_1 x_{11} + \dots + \widehat{\beta}_k x_{k1} \\ \widehat{\beta}_0 + \widehat{\beta}_1 x_{12} + \dots + \widehat{\beta}_k x_{k2} \\ \widehat{\beta}_0 + \widehat{\beta}_1 x_{13} + \dots + \widehat{\beta}_k x_{k3} \\ \dots \\ \widehat{\beta}_0 + \widehat{\beta}_1 x_{1n} + \dots + \widehat{\beta}_k x_{kn} \end{bmatrix} \\
&= X\widehat{\beta}.
\end{aligned}
$$

Residuals: $\widehat{u}$.

$$
\widehat{u} = y - \widehat{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{bmatrix} - \begin{bmatrix} \widehat{y}_1 \\ \widehat{y}_2 \\ \widehat{y}_3 \\ \dots \\ \widehat{y}_n \end{bmatrix} = y - X\widehat{\beta}.
$$

3.9 OLS properties

From the first-order conditions of the minimization problem (Eq. (5)) we can write

$$X^{'}(y - X\widehat{\beta}) = 0 \Rightarrow X^{'}\widehat{u} = 0$$

or, equivalently,

$$\left[\begin{array}{c} \sum_{i=1}^{n} \widehat{u}_i \\ \sum_{i=1}^{n} x_{1i}\widehat{u}_i \\ \dots \\ \sum_{i=1}^{n} x_{ki}\widehat{\varepsilon}_i \end{array}\right] = 0 \implies$$

$$X^{'}\widehat{u} = 0$$

$$\sum_{i=1}^{n} \widehat{u}_i = 0$$

The residuals are orthogonal to the X observations and the sum of residuals is 0.

3.10 Statistical properties: Unbiasedness

$$
\begin{aligned}
\widehat{\beta} &= (X'X)^{-1}X'y \\
&\Rightarrow \\
\widehat{\beta} &= (X'X)^{-1}X'(X\beta + u) \\
&= (X'X)^{-1}(X'X)\beta + (X'X)^{-1}X'u \\
&\Rightarrow \\
\widehat{\beta} &= \beta + (X'X)^{-1}X'\varepsilon \implies
\end{aligned}
$$

$$
\begin{aligned}
E(\widehat{\beta}) &= E(\beta + (X'X)^{-1}X'u) \\
&= \beta + (X'X)^{-1}X'E(u|X) \\
&= \beta
\end{aligned}
$$

3.11 Statistical properties: Variances

$$
\begin{aligned}
Var(\widehat{\beta}) &= E[(\widehat{\beta} - E(\widehat{\beta}))(\widehat{\beta} - E(\widehat{\beta}))'] \\
&= E[(\widehat{\beta} - \beta)(\widehat{\beta} - \beta)'] \\
&= E\left[\left((X'X)^{-1}X'u\right)\left((X'X)^{-1}X'u\right)'\right] \\
&= E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}\right] \\
&= (X'X)^{-1}X'E(uu')X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'I_nX(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}.
\end{aligned}
$$

The variance of $\widehat{\beta}$ depends *directly* on the variance of the error σ^2 and *inversely* on the *variability* of the X observations, i.e., $X'X$.

3.12 Gauss-Markov Theorem

The OLS estimator $\widehat{\beta}$ is BLUE (best linear unbiased).
For any estimator $\widetilde{\beta}$ which is linear (in the observations y) and unbiased, it turns out that

$$Var(\widetilde{\beta}) \geq Var(\widehat{\beta}).$$