ECONOMETRIC MODELING BASED ON PANEL DATA

Muminova Maxbuba Abduvafoyevna
Associate Professor of the Department of "Statistics and Econometrics"
of the Tashkent Institute of Finance

Abstract: In this article, panel data, analysis, data representing the activity of a set of enterprises by quarters, annual socio-economic indicators of a certain group of countries, construction of flexible and perfect models of panel data and spatial the importance of panel data in the analysis of the influence of various factors on the development of phenomena over time, econometrics and research of socio-economic processes is considered.

Key words: Panel data, time series, regression model, Production functions, production logarithm, identification.

Panel data are multivariate observational data that are measured repeatedly over time.

If the sample contains data on several objects, each of which is observed at several time intervals, then such data are called panel data.

Panel data is a special type of data widely used in econometrics and socio-economic research. They are a set of observations at different time points for each research object. This data structure makes it possible to analyze the dynamics and changes of events over time, as well as to take into account the individual characteristics of each object.

Panel data is used in economics, sociology, political science and many other fields. They allow researchers to take into account not only temporal changes, but also differences between research objects. This is especially useful when analyzing the influence of various factors on the development of events over time.

For example, panel data can be used to analyze the effect of education on workers' wages. Researchers collect data on wages and educational attainment for each worker over several years to see how these rates change over time and how education affects wages. they can analyze the mystery.

Panel data is a set of observations of the same variable or a set of variables over different time periods or events. This can be data collected over several years, quarters, months or days.

Panel data is a special form of data collected over time. Their main difference from "ordinary" data is that they allow the analysis of changes in variables over time and take these changes into account when running statistical tests and models.

These data are often used in economic and sociological studies to study the dynamics and interdependence of various factors. They help to determine the influence of various factors on a variable, and also predict future values of variables based on past observations.

Examples of panel data include information on household income, commodity prices, unemployment rates, or consumer preferences. For each of these variables, data are collected at different times and there are factors that influence their variation.

In econometrics, panel data often refers to multivariate data that contain measurements over a period of time. A panel dataset can be time-based and contain observations of individuals in a specific sample or information about each individual in the sample.

For example, we can see the sales-to-employee ratio for 50 firms over a five-year period, or the ratio of GDP to money supply over 20 years for 20 countries if we take panel data on a geographic basis.

Panel data consists of data obtained as a result of consecutive observation of one economic entity or objects (households, firms, regions, countries, etc.). An example of this can be the study of the annual budget of several households, the data reflecting the activities of a complex of enterprises by quarters, or the study of the annual socio-economic indicators of a group of countries. Thus, panel data combines information such as time series with spatial data: spatial data is available for each economic entity and corresponds to each entity. The input data consists of one or more time series. This special composition of panel data allows us to create more flexible and sophisticated models and to answer questions that cannot be answered within spatial models. In particular, it is possible to take into account and analyze specific individual differences between economic units that cannot be done within the framework of standard regression models.

For example, by studying the value of GDP per capita for a given country, it is possible to observe the inflation rate, investment volume, money supply, etc. for each period. But in addition to them, there are factors that cannot be observed or expressed in numerical form, but they have a significant impact on the studied indicator (for example, the geographical location of the object, history, cultural 'rifian traditions, etc.). The influence of such factors can be considered constant (that is, independent of time) for each national economy. With spatial data for only a few countries, it is possible to determine the effect of simple economic factors on GDP per capita, but it is not possible to assess differences in specific characteristics between countries. The assessment can be made based on the analysis of the results of several years of observations.

One of the persistent problems at the micro level is the representation of household spending on one type of product, for example: personal care products. By entering the family's economic, socio-demographic living conditions, you can collect spatial data and find the relationship between the cost of the studied product and the family's income. But panel data may show that income is not always important in family spending, but differences in spending are affected by family traditions, level of culture, and other factors that cannot always be measured and observed.

In many cases, individual factors of objects are correlated with other influencing variables. For example, the level of general culture of the family is naturally related to the level of family income. In regression, a certain factor means that the item is important in the model, and ignoring it in the model will lead to incorrect estimation of other parameters. In other words, models based on panel data lead to correct parameter estimates. In this case, panel data includes observation data of the same objects at different times. The non-correlation of these observations can lead to a distortion of reality, so the analysis of these models requires more accurate methods than the EKK method.

Panel data consists of data obtained from observations of n different objects at different times T. When writing cross-object data, indexes can be used to specify the numbers of objects under study. For example, - y is the value of the dependent variable for item i. It is required to specify the object number and time in the panel data. For this, it is necessary to use additional numbering, that is, the index i indicates the observation number, and the index t indicates the observation time. Thus, - *y* denotes the value of the dependent variable at *i* - object at time t (*k*-order vector), - denotes the value of the independent variable at *i* - object at time *t*, - corresponding errors (i=1,2,…,n; t=1,2,..,T).

For panel data, it is possible to show the actual situation even if some data is missing some tracking data. In balanced panels, all observations are known, that is, all variables have an observation value in each subject in each period. If the data for one object is missing for one period, such panels are called unbalanced.

The simplest model is a simple linear regression model:

$$y_{it} = x'_{it} + \varepsilon_{it} \qquad (1)$$

ISSN 2349-7793 (online), Published by INTERNATIONAL JOURNAL OF RESEARCH IN COMMERCE, IT, ENGINEERING AND SOCIAL SCIENCES.,
under Volume: 17 Issue: 08 in August-2023
https://www.gejournal.net/index.php/IJRCIESS

or by specifying as follows

$$y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ . \\ . \\ y_{iT} \end{bmatrix}, \ x_i = \begin{bmatrix} x'_{i1} \\ x'_{i2} \\ . \\ . \\ x'_{iT} \end{bmatrix}, \ \varepsilon = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ . \\ . \\ \varepsilon_{iT} \end{bmatrix};$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{bmatrix}.$$

we write in matrix form:

$$Y = X \cdot \beta + \varepsilon. \tag{2}$$

As mentioned, panel data allows to take into account individual differences between economic units (objects). One of the possible ways to implement such ideas (1) is to express the model in the following form:

$$y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it}. \tag{3}$$

here α_i - size *i* – of the object *t* represents a specific effect that does not depend on time.

Taking into account the unique effect of the object allows to draw more adequate conclusions. For example, a typical approach to the problem of estimating production functions is to estimate Eq.

$$y_{it} = \mu + x'_{it}\beta + \varepsilon_{it}. \tag{4}$$

Here: y_{it} - logarithm of production,

x'_{it} - *k* logarithms of dimensional vector production functions.

In order to have a correct idea of production functions (for small firms), it is appropriate to include management quality in the production factors, i.e.

$$y_{it} = \mu + x'_{it}\beta + q_i\beta_{k+1} + \varepsilon_{it}. \tag{5}$$

Here: q_i - represents the quality of management.

If this variable is significant, then model (4) cannot be estimated with EKKU. But q_i Since it is unobservable, it can be considered a unique quality of management. β_{k+1} (5) since it is not

ISSN 2349-7793 (online), Published by INTERNATIONAL JOURNAL OF RESEARCH IN COMMERCE, IT, ENGINEERING AND SOCIAL SCIENCES.,
under Volume: 17 Issue: 08 in August-2023
https://www.gejournal.net/index.php/IJRCIESS

identifiable in the model, it cannot be estimated. (3) in the model α_i two models can be seen depending on the assumptions about the parameter properties.

Fixed efficiency model: In equation (3). α_i is assumed to be an unknown parameter (or a model with a fixed influencing factor).

Random efficient model: (3) in Eq $\alpha_i = \mu + u_i$ it is assumed that Here μ- parameter is common to all units all the time. u_i - error (this model is recommended for independent study by the student).

The issue of choosing a model is decided individually in each case.

Fixed efficient regression models. The constant effective regression estimation method is a method that allows taking into account the effects of variables that are not included in the regression equation, that differ in terms of the observed objects, but are invariant over time.

A fixed-effects regression model uses *n* different constants (constants), one for each item. These constants are a set of binary (or indicator) variables that take into account the degree of influence of all variables omitted from the regression equation, which are different for different observations, but constant over time. can be imagined.

A regression model with fixed effects.

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 z_i + u_{it} \tag{6}$$

Here: y_{it} – dependent variable;

x_{it} – observed independent variable;

z_i – an unobservable variable that takes different values in individual objects, but does not change over time (for example, a cultural, social or political situation that affects the state of the object).

The main issue here is unobservable, fixed *z* when a variable is involved *x* and *y* showing the effect of β_i consists of estimating coefficients. z_i because it is an expression that has different values for different objects but does not change over time $\alpha_i = \beta_0 + \beta_1 \cdot z_i$ total *n* can be considered as an expression that has different values in different objects. Then expression (6) can be written as follows:

$$y_{it} = \beta_1 x_{it} + \alpha_i + u_{it} \tag{7}$$

This expression represents a regression model with fixed effects. Here $\alpha_1, \alpha_2, \ldots, \alpha_n$ are considered unknown constants to be evaluated. The interpretation of these constants comes from studying the right side of equation (7).

Deviation constant obtained when estimating the regression line β_1 same for all objects. Constants vary depending on the objects.

In expression (7). α_i s "*i* – can be considered as a special characteristic of the object, $\alpha_1, \alpha_2, \ldots, \alpha_n$ and in some cases, they are considered to be the same factors that have a constant influence (effective). Changes in these constants may be related to changes in neglected variables (for example, in equation (6) z_i).

ISSN 2349-7793 (online), Published by INTERNATIONAL JOURNAL OF RESEARCH IN COMMERCE, IT, ENGINEERING AND SOCIAL SCIENCES.,
under Volume: 17 Issue: 08 in August-2023
https://www.gejournal.net/index.php/IJRCIESS

Observable-specific constants above can be modeled using binary variables used to define each observable. To construct a fixed-influence regression model using binary variables, *i*=1 which is equal to one in and is equal to zero in other cases D_{1i} – binary variable, *i* = 2 which is equal to one when it is and equal to zero in other cases D_{2i} –variable and so on. But entering all n binary variables into the regression equation simultaneously with a common free term leads to multicollinearity. Therefore, it is necessary to optionally omit one of the binary variables (for example, for the first group of observations D_{1i} the). Then, the constant effective regression equation equivalent to equation (6) can be written as:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \ldots + \gamma_n D_{ni} + u_{it} \quad (8)$$

Here: $\beta_0, \beta_1, \gamma_2, \ldots, \gamma_n$ –unknown coefficients to be estimated.

To determine the relationship between the coefficients in equation (7) and the coefficients in equation (8), it is necessary to equalize their right sides. The first item in equation (3) is for the regression line $\beta_0 + \beta_1 x_{it}$ the expression can be written, then $\alpha_1 = \beta_0$ will be. Expression for the regression line of the second and subsequent items $\beta_0 + \beta_1 x_{it} + \gamma_i$ and accordingly $i \geq 2$ for $\alpha_i = \beta_0 + \gamma_i$.

Thus, the method in the form of equations (3) and (4) can be used to create a constant effective regression model. In equation (3), the model values depend on the properties of one or another object *n* written by various constants. In equation (4), the model is one general constant and $n-1$ written by binary variables.

In both ways *x* the (angle) coefficients in front of the variable take the same values for all objects. It should be noted that the source of the eigenconstants in equation (3) and the binary variables in equation (4) is a single unobservable that takes different values but is invariant over time. z_i is a variable (character).

Modelni bir nechta *x* o‘zgaruvchilarigacha kengaytirish

Extending the model to multiple x variables

If there are other observables that affect *y*, are correlated with *x*, and are time-varying, they need to be included in the regression equation to avoid the resulting errors because they are neglected.
.

A constant efficient regression model can be expressed as follows:

$$y_{it} = \beta_1 x_{1,i1} + \ldots + \beta_k x_{k,it} + \alpha_i + u_{it} \quad (9)$$

Here: $i = 1, \ldots, n;\ t = 1, \ldots, T$. $x_{1,i1} - i$ – the value of the independent variables entered in the first stage of the object at time t, $x_{k,it} - i$ - the value of the independent variables entered in the k - step of the object at time t; $\alpha_1, \alpha_2, \ldots, \alpha_n$ – constant quantities depending on the object of observation.

The model can be equivalently written using common constants, *x* - independent variables and *n-1* binary variables:

$$y_{it} = \beta_0 + \beta_1 x_{1,i1} + \ldots + \beta_k x_{k,it} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \ldots + \gamma_n D_{ni} + u_{it} \quad (10)$$

ISSN 2349-7793 (online), Published by INTERNATIONAL JOURNAL OF RESEARCH IN COMMERCE, IT, ENGINEERING AND SOCIAL SCIENCES.,
under Volume: 17 Issue: 08 in August-2023
https://www.gejournal.net/index.php/IJRCIESS

Here: $D_{2i} - i = 2$ a binary variable that is equal to one when it is and zero otherwise, etc.

Thus, panel data allows taking into account the dynamics and individual characteristics of research objects in the analysis of various questions. They are a powerful tool for research, allowing for more accurate and reliable results. Using panel data allows for more accurate analysis of historical trends, identification of causal relationships, and more accurate forecasting.

REFERENCES:

1. Muminova M.A. Ekonometrika: O‘quv qo‘llanma.–T.: “Iqtisod-Moliya”, 2022.

2. Habibullayev I.H, UtanovB. Ekonometrika asoslari: O‘quv qo‘llanma. –T.: “Iqtisod-Moliya”, 2018, -172b.

3. Habibullayev I. Ekonometrika: Darslik. –T.: “Iqtisod-Moliya”, 2020, -240 b.

4. Habibullayev I., Jumayev A. Ekonometrika: amaliy mashg‘ulot uchun o‘quv qo‘llanma. –T.: “Iqtisod-Moliya”, 2020, -176 b.

5. Елисеева И.И., Куришева С.В. и др. Эконометрика: Учебник. –М.: Издательство Юрайт, 2018. –288 с.

6. Muminova Makhbuba Abduvafoyevna. (2022). Модели корреляционно-регрессионного анализа социально-экономических процессов. международный журнал социальных наук и междисциплинарных исследований issn: 2277-3630 импакт-фактор: 7.429, 11(11), 441-451. извлечено из https://www.gejournal.net/index.php/ijssir/article/view/1227

7. Mullabayev Baxtiyarjon Bulturbayevich, Muminova Makhbuba Abduvafoyevna. “Development of methodology for managing the activities of vertically integrated industrial enterprises”. International journal of research in commerce, it, engineering and social sciences *issn: 2349-7793 impact factor: 6.876.,* volume: 16 issue: 06 in june 2022.

8. Mullabayev Baxtiyarjon Bulturbayevich, Muminova Makhbuba Abduvafoyevna. IMPROVEMENT OF MODERN ENERGY INDUSTRIES IN THE CONDITIONS OF DIGITAL ECONOMY. ASIA PACIFIC JOURNAL OF MARKETING & MANAGEMENT REVIEW ISSN: 2319-2836 *IMPACT FACTOR: 7.603* Vol 12, Issue 05, 2023