STRUČNI RAD

CHALLENGES OF PANEL DATA SERIES APPLICATION IN RESEARCH OF VARIOUS ECONOMIC PHENOMENA

Stankov Biljana[1]

Roganović Milijana[2]

Abstract: *In recent decades, in addition to time series data and comparative data, panel data has been intensively used in empirical research. The application of these data series facilitates the implementation of analyses that could not otherwise be carried out using individual time series or comparative data. One of the key advantages of the panel data model in researching numerous economic phenomena and other aspects of social reality is that it provides the maximum amount of information from a limited number of observations per observation unit in a given time period while minimizing the number of degrees of freedom. Greater data variability contributes to greater efficiency of estimates and a reduction of multicollinearity, and given that these models enable the observation of variations by observation units (individual effects) and by time periods (time effects), the analysis of structure and changes in structure over time can be performed simultaneously. In addition, panel models provide greater informativeness as a result of an increased sample size and a higher number of degrees of freedom.*

Keywords: *panel data, panel series, panel data models, research, economic phenomena.*

[1] Visoka poslovna škola strukovnih studija, Vladimira Perića Valtera 4, Novi Sad, biljana.stankov@vps.ns.ac.rs, https://orcid.org/0000-0001-8773-4143

[2] Visoka poslovna škola strukovnih studija, Vladimira Perića Valtera 4, Novi Sad, milijana.roganovic@vps.ns.ac.rs

INTRODUCTION

According to the available scientific and expert literature, the first panel data collection and the implementation of a panel study, which dealt with the income dynamics, was carried out in the sixties of the last century in the USA - *Panel Study of Income Dynamics - PSID* (Sarafidis & Wansbeek, 2021). Since then, the popularity and application of the methodology in the field of panel data analysis has been constantly increasing with the development of a large number of new models, methods and research tests.

Panel series represent a combination of time series data and comparative data as they simultaneously provide temporal and spatial dimensions (Gujarati, 2009). These data represent a special form of composite data where the same cross-sectional unit (for example country, company, etc.) is observed over time (Gujarati, 2004). In the aforementioned context, cross-section data refer to a set of observation values that a variable has in different time periods, and are therefore collected at regular time intervals (e.g. daily, weekly, monthly, quarterly, annually, etc.), while comparative data (Pooling time series) are data of one or more variables that were collected at the same moment in time (e.g. household consumption by country during one year).

The research problem of this paper is oriented towards the challenges and possibilities of applying panel data series in the research of various aspects of economic reality. The subject of the research involves the analysis of the advantages and limitations of the panel data application, as well as the consideration of types and models of panel data which are implementing in empirical research.The aim of this research study refers to finding out what needs to be done in cases where research cannot be conducted using only time series or cross-sectional data. Also, the research aim includes determining the most common areas of application of panel data series and indicating the effectiveness of their application in solving the previously mentioned problem.

The scientific and expert literature is very rich in research studies that apply panel data series. In Europe, the application of panel studies began at the end of the eighties of the last century. Initially, Hujer & Schneider (1989) published the results of a socio-economic panel research in Germany, shortly after that the results of a study dealing with labor mobility in Sweden (Bjorklund, 1989) appeared, and then Alessie, Kapteyn & Melenberg (1989) published the results of a household consumption research in the Netherlands.

At the end of the eighties and during the nineties of the last century, panel models were increasingly applied in numerous scientific researches, mainly in the field of social sciences, such as political sciences (Beck & Katz, 1995), sociological sciences (England, Farkas, Kilbourne & Dou, 1988), and economic sciences (Brown, Keim, Kleidon & Marsh, 1983; Boehmer & Megginson, 1990; Erdem, 1996; Keane, 1997) etc.

In the Republic of Serbia, the first studies based on panel data series were economic studies dealing with foreign trade (Dragutivnoć Mitrović, 2005). More frequent application has been noted over the last decade, in research on factors of business success of enterprises in the Republic of Serbia (Knežević, 2015), then research on foreign direct investments and competitiveness of european developing countries (Stankov, 2017), the difference in earnings and the impact on employment (Ognjenović, 2018) and relationship between competitiveness and coruption (Radulović, 2018). In recent years, the application of panel data series has also been noted in sociological research on household consumption in Serbia (Hanić & Bugarčić, 2021), as well as the connection between investment in education and some macroeconomic elements (Lepojević & Janković-Milić, 2024).

POSSIBLITIES AND LIMITATIONS OF PANEL DATA APPLICATION

Due to the numerous advantages of panel series, over time there has been an intensification of their application in various fields of empirical research. Among the most significant advantages, according to Hsiao (2003) and Klevmarken (1989), the following should be highlighted:

- Controlling individual heterogeneity.

One of the basic assumptions on which panel series are based is that the unit of observation (company, country, etc.) is a heterogeneous category. In the case of time series data and comparative data, there is no possibility of heterogeneity control, so the research results may be biased.

- More information, greater variability, a higher number of degrees of freedom and greater efficiency of panel data with lower collinearity between variables.

The problem of multicollinearity is often present in time series. In the case of the panel series application, the spatial dimension is included, which adds additional variability of the data, so the probability of multicollinearity is significantly reduced.

- More successful research into the dynamics of adaptation.

By using panel data, trends such as periodic unemployment, job changes, residential mobility, income volatility, etc. can be studied much better compared to the use of time series data and comparative data. With panels, it is possible to connect certain experiences and behaviors of the subject at one point in time with other experiences and behaviors of the same subject at another point in time.

- More effectively identifying and measuring effects that cannot otherwise be identified using only comparative data or only time series data.

The research problem can arise if there are large differences in the values of the observation units being compared, making such a comparison meaningless. The above-mentioned can be avoided with panel series because their application ensures the observation of differences in data related to the same unit of observation.

Illustration 1. Panel data scheme
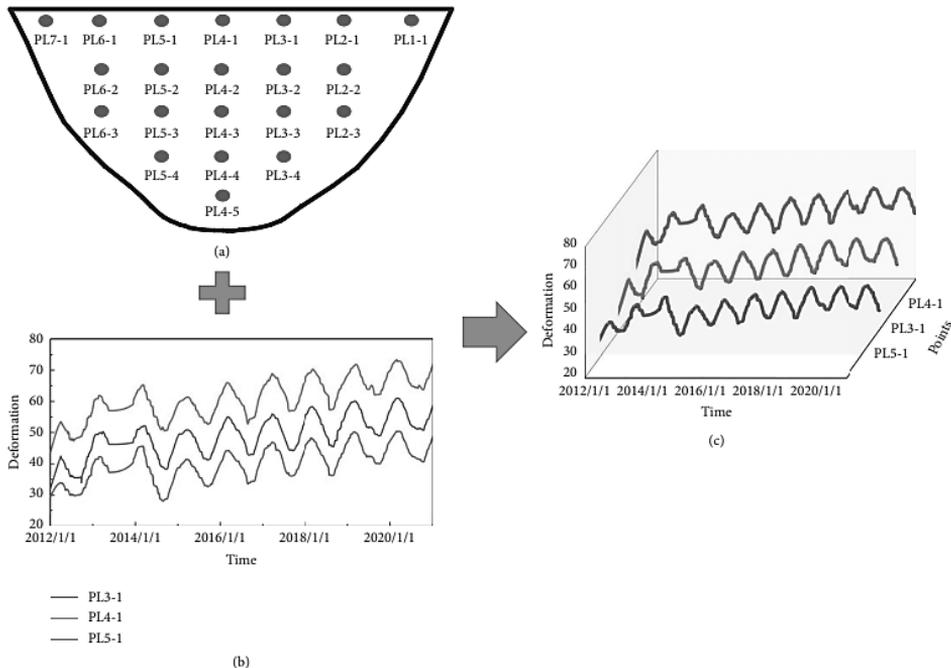
Note. Cross-sectional data (a) + Time-series data (b) = Panel data (c)

Source: Cheng, L., Chen, J., Ma, C., Yang, J., Xu, X., & Yuan, S. (2023). Multipoint Deformation Safety Monitoring Model for Concrete Arch Dams Based on Bayesian Model Selection and Averaging. *Structural Control and Health Monitoring*, 2023(1), 5042882.

One of the important advantages of using this data is related to the dynamism of information and empirical results, which is achieved by using comprehensive panel data modeling instead of single-point modeling. In this way, the validity and accuracy of the model are increased.

Cheng et al. (2003) point out another advantage and state that two-dimensional panel data includes time series and cross-sectional information, as shown in Illustration 1 (c). Cross-sectional data indicate the locations of the observation points (a), and time series data reflect the time effects of the prototype observation data (b).

Ajmani (2009) states that another significant advantage of using panel series is reflected in the increase in sample size, i.e. increasing the number of observations. If N represents the number of observation units, and T the number of time periods, the research will include NT observations.

In addition to the evident advantages, by applying the panel series analysis, the authors of empirical studies also encounter certain problems and face the following limitations of this data:

- Data collection and design issues.

According to Bailar (1989), it is a comprehensive problem that combines the problem of coverage (complete data on the population of interest is missing), the problem of non-response (due to the lack of cooperation of the respondent or errors made by the interviewer), the problem of recall (the respondent does not remember the answer exactly), and problems related to the nature of the interview (frequency of interviews, interval between interviews, reference period, etc.).

- Measurement errors.

When conducting research, errors can arise due to incorrect understanding of questions, insufficient or incomplete recall of content, intentional provision of incorrect answers, inadequacy of available information, errors in recording information, interviewer influence, etc. (Kalton, Kasprzyk & McMillen, 1989).

- Selection problems (self-selection, non-response and attrition).

For example, if labor force employment is investigated, a problem arises when certain individuals refuse employment because, for example, they are not satisfied with the offered wage (self-selection). The problem can also arise if the respondent is not available and fails to give an answer, or when they refuse to give an answer during the survey (non-response). Non-response can represent a very serious problem in this case, which

persists not only in the current, but also in all subsequent research cycles. Attrition usually increases when moving from one round of data collection to another, but over time the growth rate becomes lower.

- Time series dimension.

Since typical micro-panels include annual data and thus cover a short time period for each individual subject, the short dimension of the time series can be a problem. If, in this case, the number of time units were to increase, it could lead to the emergence of new problems related, for example, to data spillage or difficulty in computational processes.

- Dependency of comparative data.

This problem can affect the research results and lead to the occurrence of errors in inference. Most often, with macro panels of countries or regions, where very long time series are used, the possibility of comparative data dependence is not taken into account.

Researchers must be aware of the mentioned limitations and shortcomings of panel data. It is unreasonable to expect that the application of panel data will solve all the problems that normally arise when using comparative data and time series data (Baltagi, 2005).

TYPES AND MODELS OF PANEL DATA IN EMPIRICAL RESEARCH

Types of panel data

Since panel studies in recent decades have begun to be intensively applied in the research of various social phenomena, there are more and more theoretical and empirical studies in the scientific literature that talk about their types and ways of grouping. According to one of the criteria, panel data can be viewed as balanced and unbalanced. In the case of balanced panel data, each comparative observation unit has the same number of time series observations, that is, the time series are of the same length. In these circumstances, there are no problems caused by a lack of data. Gujarati (2004) points out that if the number of observations differs from one subject to another in the panel, then it is an unbalanced panel where data for certain observation units are missing in some time periods, and modified classical estimation methods must be applied.

If a large number of units N is observed in a short time period T, this is referred to as classical panels, while longitudinal panels analyze a small number of observation units in a large number of time periods.

The difference between macro panels and micro panels is found in the size of the observation unit. Namely, macro panel data refer to countries, regions, industrial branches and the like, while the analysis of certain companies, individuals, etc., is considered micro panel research.

Depending on how the problem of missing research data is solved, panel data can be viewed as rotating and pseudo panels. Biørn (1981) points out that this problem can be solved by replacing, i.e., "rotating," the group of units whose data are missing in a given period with an equal number of new units for which data are available, in order to maintain the same number of observation units throughout the entire research period. In this way, a new series of rotating panel data would be obtained. Baltagi (1995) states that for models based on rotational panel series, there are well-developed estimation and testing methods. Since it is not always feasible to ensure sample consistency from period to period in terms of the presence of exactly the same observation units, groups of observation units and their average values can be used instead of individual data for estimating panel models. These series are therefore called pseudo-panel series.

Models of panel data

Panel data models can be viewed as linear and non-linear models. Linear panel models can be static (estimation and testing are carried out under the conditions of violation of the initial assumptions) and dynamic models, as well as non-stationary panels (unit root tests and cointegration).

The general form of the panel data regression model can be presented as:

$$y_{it} = \beta_{1it} + \sum_{k=2}^{K} \beta_{kit} x_{kit} + u_{it}$$

where:

y_{it} – the value of the dependent variable for the i observation unit in period t;

x_{kit} – the value of the k independent variable for the i observation unit in period t;

$x_{1it} = 1$, for each i and t;

β_{1it} - unknown regression parameters, which in the general form of the panel data model are variable by observation units and time periods, i.e., they are not constant as in classical regression analysis;

u_{it} – random error with arithmetic mean equal to zero and constant common variance for each i and t.

The model includes both structural and temporal dimensions, which raises the question of modeling the structural and temporal dependence of the selected variables in the model. The general specification of the linear panel data model implies that for each observation unit there is a different response of the dependent variable to variations in the independent variables and that this response differs for each time period. Therefore, the regression parameter of each observation unit is specific to each time period. Such a model cannot be estimated because the number of regression parameters exceeds the number of data points in the sample. To estimate the model, assumptions are introduced regarding the properties of the model regressors, the characteristics of the random error, the statistical dependence between the regressors and the random error, and the degree of variability of the regression parameters.

Depending on the degree of variability of the regression parameters, the following situations may occur:

- all regression parameters in the model are constant;
- the intercepts differ by observation units, while the regression parameters for the independent variables are constant;
- regression parameters with independent variables are variable by observation units and time periods.

If the intercepts vary while the regression parameters are constant, a distinction is made between models with individual effects (the intercept varies only by observation units) and models with both individual and time effects (the free member varies across both observation units and time periods).

When the coefficients vary for each i and each t, there is a total of NT intercepts and $k-1$ slope coefficients for each i as well as for each t. Since it is impossible to estimate such a model, it is necessary to introduce various restrictions.

One restriction can be based on the assumption that all regression parameters are constant; in that case, a single parameter β_1 and a single slope parameter β_k would be estimated for the entire sample, meaning that all observation units react in the same way. Such a model is called the **model with constant regression parameters (Pooled-effects Model).**

The second restriction, which is most commonly applied to classical panels where **N** is large and **T** is small, is based on the assumption that the intercepts β_1 vary, while the parameters of the regressors remain constant. A model in which the intercepts β_1 vary only across observation

units is called the **individual effects model**, while a model in which the intercepts β_1 vary across both observation units and time periods is called the **individual and time effects model**. Furthermore, the individual and time effects can be treated as fixed parameters, in which case we refer to the **Fixed-effects Model,** or they can be treated as stochastic variables, in which case we we are talking about the **Random-effects Model.**

CONCLUSION

It is concluded that panel data series contain data on a large number of the same observation units at a given number of specific points in time or time periods, and each observation simultaneously contains both a spatial and a temporal dimension. This allows all relevant information from the sample to be used in the estimation of regression parameters (Matyas & Sevestre, 1996). Panel series are becoming very popular and are increasingly used in the quantification of economic relations because they are by nature disaggregated by observation units, as well as because they combine information on both the structure and dynamics of the observed phenomenon (Dragutinović Mitrović, 2002).

In cases where research cannot be conducted using only time series or cross-sectional data, panel data series offer an efficient solution. Among the key advantages of applying panel models are the successful control of observation unit heterogeneity, the reduced possibility of multicollinearity problems, more efficient research into adjustment dynamics, as well as the identification and measurement of effects that cannot otherwise be identified using only cross-sectional or only time series data. Certainly, one of the most significant advantages of using panel data series lies in the increase in sample size, i.e., the increase in the number of observations.

In the study of macroeconomic phenomena, panel models are most commonly applied in research on international trade, foreign trade exchange, exchange rates, and other socio-economic phenomena. At the micro level, these models are most often used in research on GDP or unemployment trends across different countries, as well as in dynamic studies of income distribution among different categories of the population or company stock prices.

REFERENCES

1. Ajmani, V., B. (2009). *Applied econometrics using the SAS system*. John Wiley & Sons, Hoboken, New Jersey.
2. Alessie, R., Kapteyn, A., & Melenberg, B. (1989). The effects of liquidity constraints on consumption Estimation from household panel data. *European Economic Review, 33*(2-3), 547-555. https://doi.org/10.1016/0014-2921(89)90134-7
3. Bailar, B., A. (1989). Information needs, surveys, and measurement errors. In *Panel surveys* (1-24). John Wiley, New York.
4. Baltagi, B., H. (1995). *Econometric Analysis of Panel Data*. John Wiley & Sons, London.
5. Baltagi, B., H. (2005). *Econometric Analysis of Panel Data*. John Wiley & Sons, England.
6. Beck, N., & Katz, J., N. (1995). What to do (and not to do) with time-series cross-section data. *American Political Science Review, 89*(3), 634-647. https://doi.org/10.2307/2082979
7. Biørn, E. (1981). Estimating economic relations from incomplete cross-section/time-series data. *Journal of Econometrics, 16*(2), 221-236. http://dx.doi.org/10.1016/0304-4076(81)90109-3
8. Bjorklund, A. (1989). Potentials and pitfalls of panel data: The case of job mobility. European Economic Review, 33(2-3), 537-546.
9. Boehmer, E., & Megginson, W., L. (1990). Determinants of secondary market prices for developing country syndicated loans. *The Journal of Finance, 45*(5), 1517-1540.
10. Brown, P., Keim, D., Kleidon, A., & Marsh, T. (1983). Stock return seasonalities and the tax-loss selling hypothesis: Analysis of the arguments and Australian evidence. *Journal of Financial Economics, 12*(1), 105-127.
11. Dragutinović Mitrović, R. (2002). *Analiza panel serija*. Zadužbina Andrejević, Beograd.
12. Dragutinović Mitrović, R. (2005). Ograničenja gravacionog modela u ekonometrijskoj analizi spoljnotrgovinske razmene. *Economic Annals*, *50*(167), 77-106.
13. England, P., Farkas, G., Kilbourne, B., S., & Dou, T. (1988). Explaining occupational sex segregation and wages: Findings from a model with fixed effects. *American Sociological Review*, 544-558.
14. Erdem, T. (1996). A dynamic analysis of market structure based on panel data. *Marketing Science*, *15*(4), 359-378.
15. Gujarati, D., & Porter, D. (2009). *Basic Econometrics*. Mc Graw-Hill International Edition.

16. Gujarati, D. (2004). *Basic Econometrics*. United States Military Academy, West Point.
17. Hanić, H., & Bugarčić, M. (2021). Determinants of household consumption in Serbia - methodological approaches and empirical results. *In Marketing Challenges and Perspectives* (pp. 9-10). Serbian Marketing Association.
18. Hsiao, C. (2003). *Analysis of panel data*. Cambridge University Press, Cambridge.
19. Hujer, R., & Schneider, H. (1989). The analysis of labor market mobility using panel data. *European Economic Review, 33*(2-3), 530-536.
20. Kalton, G., Kasprzyk, D., & McMillen, D., B. (1989). Nonsampling errors in panel surveys. *Panel Surveys*, 249 – 270.
21. Keane, M. P. (1997). Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business & Economic Statistics, 15*(3), 310-327. https://doi.org/10.1080/07350015.1997.10524709?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle
22. Klevmarken, N., A. (1989). Panel studies: What cena we learn from them?. *European Economic Review, 33*(2-3), 523-529. https://doi.org/10.1016/0014-2921(89)90131-1
23. Knežević, A. (2015). *The Panel model application in indentifying factors of business success of manufacturing companies* (Doctoral dissertation). University Novi Sad, Faculty of Technical Sciences.
24. Lepojević, V., & Janković-Milić, V. (2024). Investment in Education and Economic Growth: Panel Data Analysis. In *Circular Economy: Trends and Perspectives* (pp. 171-176). University of Niš, Faculty of Economics.
25. Matyas, L., & Sevestre, P. (1996). *The econometrics of panel data: a handbook of the theory with applications*. Kluwer Academic Publisher.
26. Ognjenović, K. Đ. (2018). *Analiza razlika u zaradama i višestrukog izbora zaposlenja primenom parametarskih i poluparametarskih ekonometrijskih metoda panela* (Doctoral dissertation). University of Belgrade.
27. Radulović, M. (2018). The influence of national competitiveness on the perception of corruption in the countries of the European Union: Panel Vec Model. *Business Economy, 13*(1), 73-92. PROVERITI REFERENCU
28. Sarafidis, V., & Wansbeek, T. (2021). Celebrating 40 years of panel data analysis: Past, present and future*. Journal of Econometrics, 220*(2), 215-226. https://doi.org/10.1016/j.jeconom.2020.06.001

29. Stankov, B. (2017). *Panel analysis of foreign direct investment and competitiveness in emerging and developing European countries* (Doctoral dissertation). University Business Academy, Faculty of Economics and Engineering Management Novi Sad.

IZAZOVI PRIMENE PANEL SERIJA PODATAKA U ISTRAŽIVANJIMA RAZLIČITIH EKONOMSKIH POJAVA

Biljana Stankov

Milijana Roganović

Sažetak: *Poslednjih decenija se u empirijskim istraživanjima, pored podataka vremenskih serija i uporednih podataka, intenzivno koriste i panel podaci. Primena ovih serija podataka olakšava sprovođenje analiza koje se inače ne bi mogle realizovati pojedinačnim korišćenjem vremenskih serija ili uporednih podataka. Jedna od ključnih prednosti modela panel podataka u istraživanjima brojnih ekonomskih pojava i drugih aspekata društvene stvarnosti, je ta što i pored ograničenog broja opservacija sa jedne strane i minimum broja stepeni slobode možemo dobiti maksimum informacija u datom periodu. Porast efikasnosti ocena i smanjenje multikolinearsnoti nastaje zahvaljujući većoj varijabilnosti podataka. Pored navedenog, modeli panel podataka beleže još jednu značajnu prednost koja se ogleda u objedinjavanju individualnih i vremenskih efekata. Varijacije određene pojave se mogu posmatrati po jedinicama posmatranja, ali i po periodima, pa se pored analize strukture može izvršiti i analiza promena u strukturi tokom vremena. Zahvaljujući rastu veličine istraživačkog uzorka i broja stepeni slobode, panel modeli obezbeđuju i veću informativnost.*

Ključne reči: *panel podaci, panel serije, modeli panel podataka, istraživanje, ekonomske pojave.*