

GRAPHICAL APPROACH TO AGE-PERIOD-COHORT ANALYSIS

Adam Sagan✉

Cracow University of Economics, Poland

ABSTRACT

The paper presents the graphical approach to decomposition of APC effect in cohort studies (mainly applied to demographic phenomena) using multilevel or accelerated longitudinal design. The aim of the paper is to present and visualize the pure age, period and cohort effects based on simulated data with an increment of five for each successive age, period and cohort variation. In cohort analysis on real data all of the effects are usually interrelated. The analysis shows basic patterns of two-variate APC decomposition (age within period, age within cohort, cohort within period, period within age, cohort within age, period within cohort) and reveals the trajectory of curves for each of the pure effects. The APC plots are developed using `apc` library of R package.

Key words: longitudinal analysis, cohort tables, APC plots

JEL codes: C83, J11

INTRODUCTION

Repeated surveys and cohort studies are one of the most frequently used in longitudinal economic, public health, social, finance and marketing research [Fu 2018]. They can be conducted on the same sample as in panel studies or on different samples selected from a target population as repeated independent surveys. The purpose of using both repeated market surveys and panel research is to study and explain changes in the studied phenomena. In the case of panel studies, control of changes covers individual units, but in the case of repeated surveys, the age cohorts are subject to controls and explanations. Especially in the area of management and economics, cohort studies and the analysis of age-period-cohort (APC) effects is a crucial and insightful stream of research. They are related to labour economic studies [Hanoch and Honig 1985, Heckman and Robb 1985, Strzelecki 2019], analysis of savings and consumption [Browning et. al. 2016],

inequality [Kalwij and Alessie 2007], migration [Benstock et al. 2010], and the life cycle [Deaton and Paxson 1994, Karonen and Niemela 2020].

The choice between panel surveys and repeated surveys is driven by the objectives of the survey. In studies with independent samples, it is not possible to assess the actual total changes occurring at the unit level, but only the “net” changes (net effect of all changes) can be assessed. In panel studies, it is possible to observe changes at the individual unit level and thus gross and net changes within individual cohorts can be measured. Since in panel studies the sample is drawn before the first wave of research, these samples are not appropriate for assessing the net effect of the entire population. Contrary to studies with repeated independent surveys, in panel studies, new age cohorts cannot be included in the sample in subsequent waves of studies, and thus the population net effect cannot be estimated in these studies [Firebaugh 1997, p. 3].

Research on independent samples allows for the control of age cohorts. The term cohort is defined as a group of individuals experiencing the same socially defined event at the same time. The most common event in the lives of individuals is the fact of their birth, which makes it possible to distinguish age cohorts (people born at the same period of time). Repeated surveys conducted over longer periods allow to track changes in the behaviour of individual age cohorts and thus to assess the generation effects related to the affiliation of respondents to specific cohorts. Despite the popularity of cohort analysis, the application of graphical tools in decomposition of the effects are still limited [Borowski 1976, Spyrka 2012]. The objective of the paper is to provide a framework for graphical decomposition of age-period-cohort (APC) effect based on simulated data representing pure age, period and cohort effects provided by Rentz and Reynolds [1981], and Glenn [2005].

Graphical decomposition is a popular and user-friendly way of performing APC analysis without using advanced statistical modeling.

METHOD OF APC DECOMPOSITION

Taking into account the criterion of the year of birth in defining a cohort leads to the identification of three effects in conducting repeated studies on independent samples: 1) age effect, 2) period effect and 3) cohort effect. Analysis of these APC (Age-Period-Cohort) effects allows to assess the impact of maturity (the effect of differences over the age of the respondents), general historical and social conditions (the effect of differences in the measurement periods) and generational factors (the effect of differences in age cohorts) on studied phenomena. APC effects are illustrated in Table 1.

Table 1. Age (a), Period (p) and Cohort (c) effects

	p1970	p1980	p1990	p2000
a20	c1950	c1960	c1970	c1980
a30	c1940	c1950	c1960	c1970
a40	c1930	c1940	c1950	c1960
a50	c1920	c1930	c1940	c1950

Source: Author's own elaboration.

Comparing the analyzed phenomena against the columns of the table allows for the assessment of differences in the age cross-section (e.g. differences in shopping preferences between people aged 30 and 50) in a given measurement period (e.g. in 1980 or 1990). The row-wise analysis makes it possible to determine the influence of contextual-historical factors affecting the analyzed rates for a given age category (e.g. differences in shopping preferences of 50-year old consumers between 1970 and 1980). The analysis related to the values on the diagonals of the table concerns the assessment of the influence of generational factors for the given birth year of the respondents (e.g. differences in shopping preferences for those born in 1940 compared to those born in 1960).

Statistical identification of individual effects is difficult due to the occurring collinearity of APC effects. Each effect is intertwined as a perfect linear combination of the other two, which is referred to as the age-period-cohort identification problem. The respondent's birth year is a function of the age and period of the survey ($C = P - A$). The respondent's age is a function of the cohort to which he/she belongs (year of birth) and the study time ($A = P - C$). The study period is a function of the year of birth and age ($P = C + A$). One of the simple methods of identifying the effects, proposed by Mason, Mason, Winsborough, and Poole [1973], consists of transforming the variables determining the age, study period and cohort into $n-1$ dummy variables (e.g. if there are four study periods, they are coded into three dummy variables). The transformed variables are introduced into the model as variables explaining the studied phenomenon.

The decomposition of APC effects depends on the longitudinal study design applied. There are two basic systems for longitudinal data analysis: 1) repeated cross-sectional studies, 2) prospective cohort panels and sequential ALD (Accelerated Longitudinal Design) cohort analysis [Yang and Land 2013].

Repeated cross-sectional studies are characterized by the fact that the compared age cohorts in subsequent study periods come from independent samples randomly selected from the same population. In this type of research, it is not possible to observe intra-unit changes over time, because a new research sample is drawn in each period. The layout of the cohort analy-

sis is based on placing successive study periods in the columns of the table, the compared age ranges in the rows, and the cohort effects on the diagonals of the table. This layout is presented in Table 2 and treats all three dimensions as elements of one level of analysis. For unsustainable cohort analysis plans where, for example, the respondent's age is measured in years, study periods at two-year intervals, and five-year age cohorts, multi-level data designs are used. In such a situation, the observations regarding the age of the respondents are nested in the period of measurement and in the respondent's age cohort. The measurement period and cohort define the social and historical context of the measurement made at the level of the individual respondent and his/her age. This type of data structure is presented in Table 2.

Table 2. Multilevel design of cohort data

	Period 1990	Period 1992	Period 1994	Period 1996
Cohort 1975	15	×	×	×
Cohort 1970	20	22	×	×
Cohort 1965	25	27	29	×
Cohort 1960	30	32	34	36
Cohort 1955	×	37	39	41
Cohort 1950	×	×	44	46
Cohort 1945	×	×	×	51

Source: Author's own elaboration.

Rows in Table 2 represent the individual age cohorts and the columns represent the measurement periods. The multi-level data layout also allows to solve the problem of multi-collinearity of effects, because the data analysis uses generalized linear mixed regression rather than classical linear regression for single-level data.

Prospective cohort panels rely on the use of panel studies to assess the cohort effect. Repeated studies in

the same cohort in subsequent periods allow the assessment of intra-unit changes. The disadvantage of this approach is the dependence of the analysis of the effects on only one initiating sample and thus conducting the analysis for one randomly selected cohort. Repeated measurements on dependent samples also cause the occurrence of the panel conditioning effect, consisting in the adaptation of the same respondents participating in repeated surveys to the survey conditions (they may, for example, remember the answers given in the previous waves of the survey).

The ALD sequential design allows multiple cohorts to be tracked across multiple periods of time. This solution enables the evaluation of both age and cohort effects due to the use of multiple cohorts in the analysis. This makes it possible to assess the intra-unit change within the cohort in subsequent study periods as well as inter-cohort comparisons. The advantage of such a solution is the acceleration of the cohort data collection process over long study periods. Instead of observing a given age cohort over a long study period, a selection of multiple age cohorts, randomized at different time periods and ages, is used to reduce the observation time for the effects of intra-unit changes. The data structure is presented in Table 3.

The table shows the four age cohorts and seven study periods. The „age” variable is also a time variable. In order to shorten the process of collecting data for the cohorts, each of the cohorts is examined in only four periods (for the other periods, data is structurally “missing”). APC analysis is performed using a number of approaches. These include techniques of graphical presentation of effects, cohort tables [Jackowska 2017] and the family of regression methods like survival analysis [Jackowska 2015], generalized linear models, generalized mixed linear models and latent growth curve models.

Table 3. Sequential ALD design

	Age 20	Age 21	Age 22	Age 23	Age 24	Age 25	Age 26
Cohort 1970	Period 1990	Period 1991	Period 1992	Period 1993	×	×	×
Cohort 1971	×	Period 1992	Period 1993	Period 1994	Period 1995	×	×
Cohort 1972	×	×	Period 1994	Period 1995	Period 1996	Period 1997	×
Cohort 1967	×	×	×	Period 1996	Period 1997	Period 1998	Period 1999

Source: Author's own elaboration.

GRAPHICAL APPROACH TO DECOMPOSITION OF PURE APC EFFECTS

Graphical presentation of the APC effects consists of presenting the values of dependent variables (rates) for the age and period or age systems and the cohort in the cross-section of either cohorts or study periods. If the dependent variable is expressed as an index (e.g., mortality rate), then its log transformation is most often performed to identify proportional increases in the rate values. Table 4 presents the cohort data illustrating the development of the studied phenomenon in four waves of studies (in the years 1950–1980) and the cross-section of six age groups (20–70 years). It only contains the effect of age: regardless of the waves of the study, the elderly reveal a stronger occurrence of a given phenomenon. The data depicts pure age effects, offsetting period and cohort effects. R program syntax for graphical analysis of APC is presented in the Figure 1 [Nielsen 2015].

Table 4 shows that for the established age values, the studied phenomenon (rate) does not change across the study periods, while in the cross-section of study periods, the intensity of the phenomenon increases proportionally with age by five units. A graph in which the age-specific index values are presented across study periods and cohorts is presented in Figure 1.

The upper panel of Figure 2 shows the relationship between rate and age within given periods (within period) and cohorts (within cohorts), and the relationship between the phenomenon and the cohort within periods (within period). In the lower panel, there are relationships between period and age dimension, age and cohort dimension, and period and cohort dimension.

For example, a graph showing the relationship between age and rate in the ‘within period’ is represented by a line representing the same values of the phenomenon under study in subsequent periods. On the other hand, the graph showing the relationship between the cohort and the rate in the ‘within period’ is presented by five functions for each period. For example, in 1950, respondents who were born in 1880 (age 70) had a value of 65, born in 1900 (50 years), value

```

library(apc)

#Cohort effect====
v.rates <- c(50, 40, 30, 20,
60, 50, 40, 30,
70, 60, 50, 40,
80, 70, 60, 50,
90, 80, 70, 60,
100, 90, 80, 70)

v.cases <- c(1,1,1,1,
1,1,1,1,
1,1,1,1,
1,1,1,1,
1,1,1,1,
1,1,1,1)

rates <- matrix(data=v.rates,nrow=6, ncol=4,byrow=TRUE)
cases <- matrix(data=v.cases,nrow=6, ncol=4,byrow=TRUE)

cohort <- apc.data.list(response=cases,
dose=cases/rates,data.format="AP",
age1=20,per1=1950,unit=10,
label="Response")

graphics.off()
apc.plot.data.within.all.six(cohort,"m")

#===Age effect=====
v.rates <- c(40, 40, 40, 40,
45, 45, 45, 45,
50, 50, 50, 50,
55, 55, 55, 55,
60, 60, 60, 60,
65, 65, 65, 65)

v.cases <- c(1,1,1,1,
1,1,1,1,
1,1,1,1,
1,1,1,1,
1,1,1,1,
1,1,1,1)

rates <- matrix(data=v.rates,nrow=6, ncol=4,byrow=TRUE)
cases <- matrix(data=v.cases,nrow=6, ncol=4,byrow=TRUE)

cohort <- apc.data.list(response=cases,
dose=cases/rates,data.format="AP",
age1=20,per1=1950,unit=10,
label="Response")

graphics.off()
apc.plot.data.within.all.six(cohort,"m")

#===Period effect=====
v.rates <- c(70, 60, 50, 40,
70, 60, 50, 40,
70, 60, 50, 40,
70, 60, 50, 40,
70, 60, 50, 40)

v.cases <- c(1,1,1,1,
1,1,1,1,
1,1,1,1,
1,1,1,1,
1,1,1,1)

rates <- matrix(data=v.rates,nrow=6,
ncol=4,byrow=TRUE)
cases <- matrix(data=v.cases,nrow=6,
ncol=4,byrow=TRUE)

cohort <- apc.data.list(response=cases,
dose=cases/rates,data.format="AP",
age1=20,per1=1950,unit=10,
label="Response")

graphics.off()
apc.plot.data.within.all.six(cohort,"m")

```

Fig. 1. R syntax for graphical APC decomposition
Source: Own elaboration.

Table 4. Pure age effect in APC analysis

Age	Period 1990	Period 2000	Period 2010	Period 2020
20	40	40	40	40
30	45	45	45	45
40	50	50	50	50
50	55	55	55	55
60	60	60	60	60
70	65	65	65	65

Source: Author’s own elaboration on basis of Rentz, Reynolds [1981] and Glenn [2005].

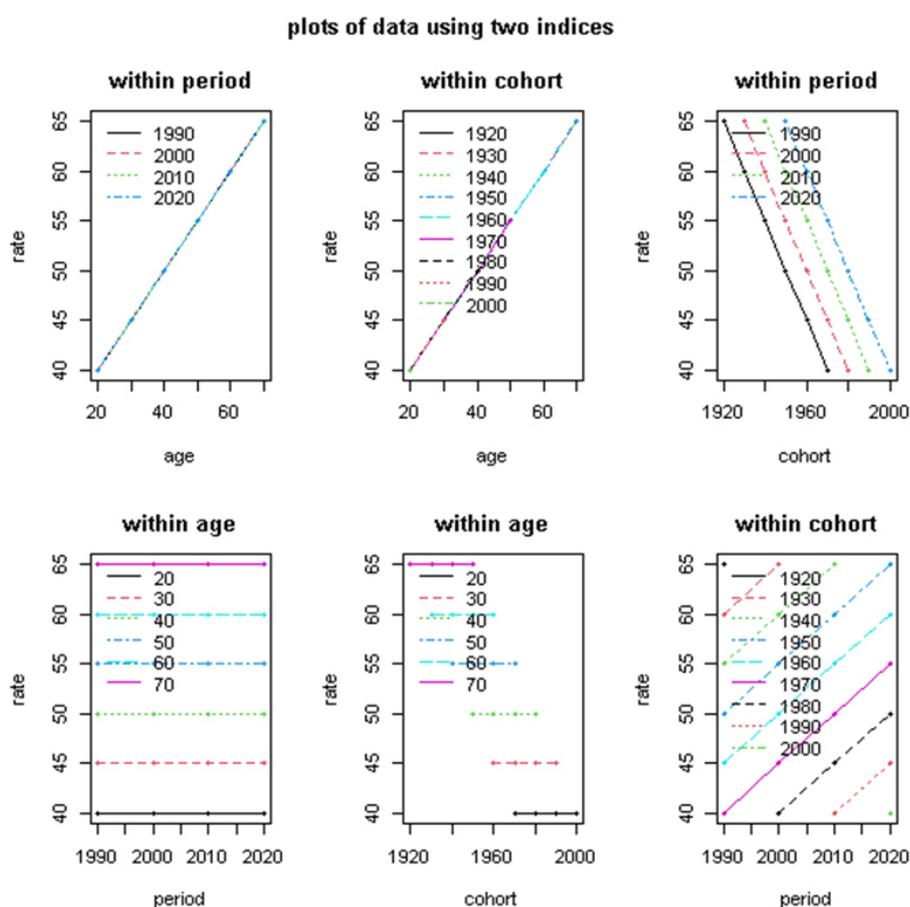


Fig. 2. APC plot with pure age effect and intertwined period and cohort effects

Source: Author’s own elaboration based on apc library of R package.

of 55, born in 1920 (30), value of 45, and born in 1930 (20), value of 40. However, in subsequent periods of the study, the same cohorts already revealed different sets of values of the studied phenomenon.

Table 5 includes values that illustrate pure period effects, offsetting age and cohort effects. The table

shows that with the next wave of research, the intensity of the phenomenon under study decreases in all age groups.

Graphs in which the indicator values specific for the study period are presented by age and cohorts, are shown in Figure 3.

Table 5. Pure period effect in APC analysis

Age	Period 1990	Period 2000	Period 2010	Period 2020
20	70	60	50	40
30	70	60	50	40
40	70	60	50	40
50	70	60	50	40
60	70	60	50	40
70	70	60	50	40

Source: Author’s own elaboration on basis of [Rentz and Reynolds 1981, Glenn 2005].

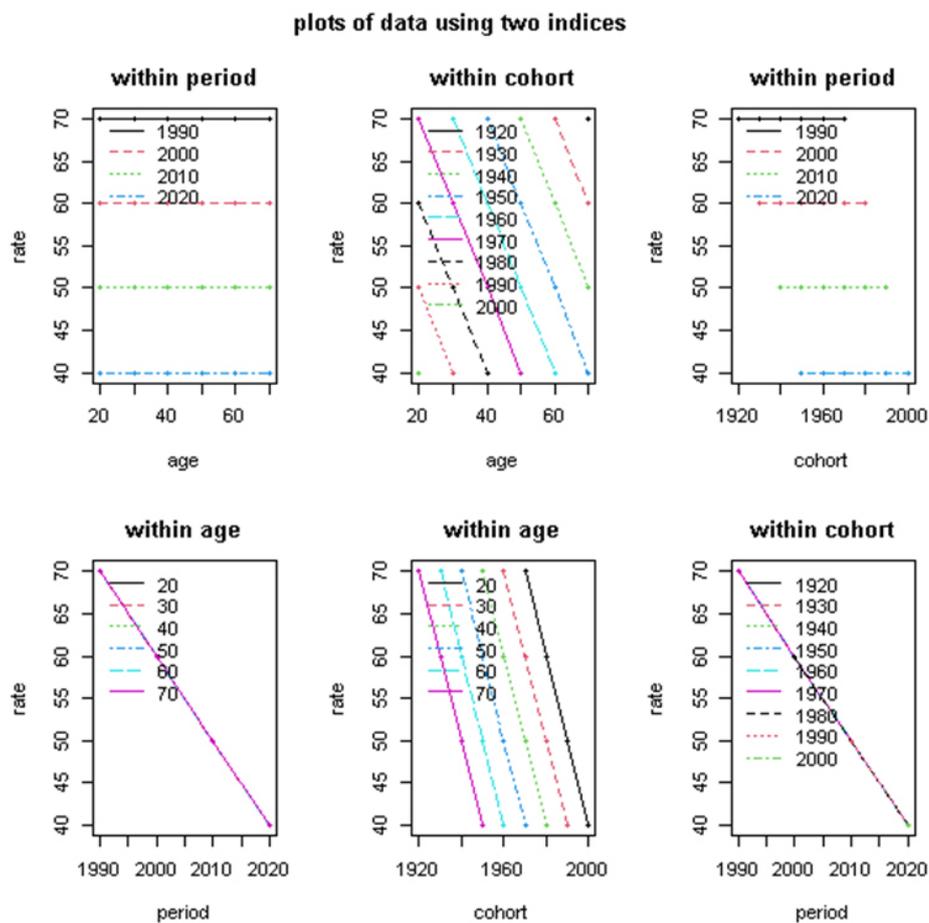


Fig. 3. APC plot with pure period effect intertwined age and cohort effects

Source: Author’s own elaboration based on apc library of R package.

For example, the graph showing the relationship between age and rate in the ‘within period’ is represented by four horizontal lines appropriate for each study period. Regardless of the age level, the value of

the index in the first period of the study is 70, in the second, 60, in the third – 50, and in the fourth – 40. On the other hand, the graph specifying the relationship between the cohort and the phenomenon (rate) with-

Table 6. Pure cohort effect in APC analysis

Age	Period 1990	Period 2000	Period 2010	Period 2020
20	50	40	30	20
30	60	50	40	30
40	70	60	50	40
50	80	70	60	50
60	90	80	70	60
70	100	90	80	70

Source: Author’s own elaboration on the basis of [Rentz and Reynolds 1981, Glenn 2005].

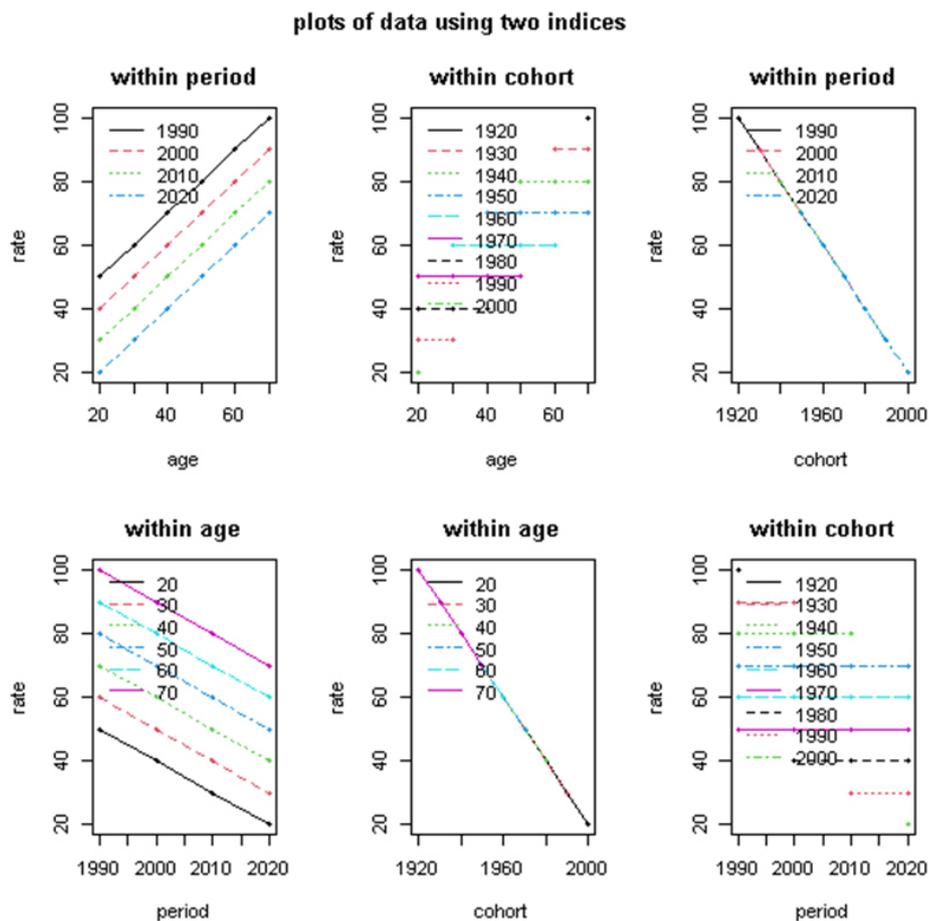


Fig. 4. APC plot with pure cohort effect intertwined period and age effects

Source: Author’s own elaboration based on *apc* library of R package.

in the periods (within period) is represented by five parallel relationships for each period. For example, in 1950, respondents born between 1880 and 1930 had a value of 70, those born between 1890 and 1940 had

a value of 60, those born between 1900 and 1950 had a value of 50, and those born between 1910 and 1960 had a value of 40. Table 6 shows the pure cohort effect, offsetting age and period effects.

Taking into account the values on the diagonals, the table shows that the younger the age cohort, the lower the intensity of the phenomenon under study. Graphs showing cohort-specific index values by age and study period are shown in Figure 4.

For example, a graph showing the relationship between age and rate in the ‘within period’ is represented by four increasing functions specific to each study period. Within the age of the respondents, the value of the phenomenon increases in each study period. On the other hand, the graph defining the relationship between the cohort and the age ‘within period’ is presented with one straight line. This relationship is negative (the younger the cohort, the lower the intensity of the response). For example, the responses for a cohort born in 1880 and tested in 1950 (70 years) is 100, born in 1900 and measured in 1960 (60 years) is 80, etc.

SUMMARY

The decomposition of pure APC effect using graphical analysis reveals the dominant pattern of each effect nested within each of two other effects. It enables to make clear interpretation of each of the net effects. The most interesting is the interpretations of parallelism of specific functions. For pure age effect (cross-sectional differences with no trend trajectories) parallel lines exhibit period ‘within cohort’ and cohort ‘within period’ effects. For pure period effect (trend differences with no cross-sectional variability) parallel lines show age ‘within cohort’ and cohort ‘within age’ effects. And for pure cohort effect (with both trend and cross-sectional variability) parallel lines depict age ‘within period’ and period ‘within age’ effect. The specificity of pure effects is that they provide all effects in a linear form with perfect separation of APC effect. It may serve as a model for comparative analysis of APC effects identified on real data, where all effects are usually intertwined and difficult to separate in graphical analysis.

ACKNOWLEDGEMENTS

The paper was financed by Cracow University of Economics based on POTENCJAL program

REFERENCES

- Borowski, S. (1976). *Analiza kohortowa i jej zastosowanie*. Wydawnictwo Naukowe PWN, Warszawa
- Browning, M., Crossley, T.F., Lüthmann, M. (2016). Durable Purchases over the Later Life Cycle. *Oxford Bulletin of Economics and Statistics* 78, 145–169.
- Deaton, A.S., Paxson, C.H. (1994). Saving, Growth, and Aging in Taiwan. [In:] D.A. Wise (Ed.), *Studies in the Economics of Aging*. Chicago University Press, Chicago, 331–361
- Glenn, N.D. (2005). *Cohort analysis*. Thousand Oaks, Sage, CA.
- Firebaugh, G. (1997). *Analyzing Repeated Surveys*, Sage University Paper Series on Quantitative Applications in the Social Sciences, no. 07-115. Thousand Oaks, Sage, CA.
- Fu, W. (2018). *A Practical Guide to Age-Period-Cohort Analysis. The Identification Problem and Beyond*. CRC Press, Boca Raton.
- Hanoch, G., Honig, M. (1985). True Age Profiles of Earnings: Adjusting for Censoring and for Period and Cohort Effects. *The Review of Economics and Statistics* 67, 384–394.
- Heckman, J., Robb, R. (1985). Using Longitudinal Data to Estimate Age, Period and Cohort Effects in Earnings Equations. [In:] W.M. Mason, S.E. Fienberg (Eds), *Cohort Analysis in Social Research*. Springer, New York, 137–150.
- Jackowska, B. (2015). Analiza kohortowa czasu istnienia mikroprzedsiębiorstw w Gdańsku. *Zarządzanie i Finanse Journal of Management and Finance* 13(4/2), 127–145.
- Jackowska, B. (2017). Analiza przekrojowa i kohortowa w badaniu umieralności a ubezpieczenia na życie. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 500, 48–61.
- Kalwij, A.S., Alessie, R. (2007). Permanent and Transitory Wages of British Men, 1975–2001: Year, Age, and Cohort Effects. *Journal of Applied Econometrics* 22, 1063–1093.
- Karonen, E., Niemela, M. (2020). Life Course Perspective on Economic Economic Shocks and Income Inequality Through Age-Period-Cohort Analysis: Evidence from Finland. *The Review of Income and Wealth* 2(66), 287–310, <https://doi.org/10.1111/roiw.12409>
- Mason, K.O., Mason, W.M., Winsborough, H.H., Poole, K.W. (1973). Some Methodological Issues in the Cohort Analysis of Archival Data. *American Sociological Review* 38, 242–258.

- Nielsen, B. (2015). *apc: An R Package for Age-Period-Cohort Analysis*, *The R Journal* 7/2, 52–64.
- Rentz J.O., Reynolds, F.D. (1981). Separating Age, Cohort and Period Effects in Consumer Behavior. *Advances in Consumer Research* 8(1), 596–601.
- Spyrka, S. (2012). Wykorzystanie analizy kohortowej do badania procesów retencji zatrudnienia. *Edukacja Ekonomistów i Menedżerów: problemy, innowacje, projekty* 1(23), 127–136.
- Strzelecki, P.A. (2019). Health, disability and labor force participation trends in Poland. *Gospodarka Narodowa* 3(299), 31–53.
- Yang, Y., Land K.C. (2013). *Age-Period-Cohort Analysis New Models, Methods, and Empirical Applications*. Chapman and Hall/CRC, Boca Raton

ANALIZA GRAFICZNA EFEKTÓW APC W BADANIACH KOHORTOWYCH

STRESZCZENIE

W artykule przedstawiono graficzne podejście do dekompozycji efektu APC w badaniach kohortowych z wykorzystaniem wielopoziomowego lub przyspieszonego wzdłużnego planu badań wzdłużnych. Celem artykułu jest przedstawienie i wizualizacja czystych efektów wieku, okresu i kohorty na podstawie symulowanych danych z przyrostem pięcioletnim dla każdego kolejnego wieku, okresu i kohorty. W analizie kohortowej na rzeczywistych danych wszystkie efekty są zwykle ze sobą powiązane. Analiza pokazuje podstawowe wzorce dwuwymiarowej dekompozycji APC (wiek w przekroju okresu, wiek w przekroju kohorty, kohorta w przekroju okresu, okres w przekroju wieku, kohorta w przekroju wieku, okres w przekroju kohorty) i ujawnia trajektorie krzywych dla każdego z czystych efektów. W analizie wykorzystano bibliotekę *apc* programu R.

Słowa kluczowe: badania wzdłużne, tabela kohortowa, wykresy APC