Stanisław M.S. Halkiewicz[37]

# 18. What are Difference-in-Difference models? An overview of history, theory and applications

## 18.1. Origins of Difference-in-Differences models

### Epidemiological roots

Difference-in-Differences (DD) models, also referred to as DiD, D-i-D, or Diff-in-Diff, originate from epidemiological studies and are widely used to evaluate the effectiveness of disease prevention measures. John Snow is considered a precursor of the DD method. His analysis of the 1854 cholera outbreak in London's Soho district laid the groundwork for DD models [E.C. Caniglia and E.J. Murray 2020]. Snow identified contaminated water from the Broad Street pump as the source of the outbreak, leading authorities to seal the pump, significantly slowing the disease spread [T.H. Tulchinsky 2018]. Despite initial scepticism, his findings were later confirmed by reverent Henry Whitehead in 1855 [S.W.B. Newsom 2006].

In subsequent research, Snow further developed the DD method by comparing cholera cases in households receiving contaminated and uncontaminated water before and after exposure. This approach, known as the "before and after" study, remains central to DD models today [J. Snow 1857; T.H. Tulchinsky 2018]. Snow is widely regarded as the father of modern epidemiology and anaesthesiology, due to his extensive work on chloroform [D. Cameron and I.G. Jones 1983; N.G. Snowise 2023]. While his contributions to medicine are well recognized, his impact on econometrics and data analysis is often overlooked.

### Further implementation of DD methods in public economics by Richard Lester and Orley Ashenfelter

In the field of economic sciences, methodologies akin to Difference-in-Differences (DD) models began to emerge in the 1950s. A notable early example is R. Lester's [1946] study, which compared employment growth between northern and southern U.S. states before and after the introduction of the minimum wage.

[37] Wroclaw University of Economics and Business.

However, Lester did not perform the requisite calculations to derive an econometric model, precluding his work from being classified as a true DD model.

O. Ashenfelter's [1978] work represents a closer approximation to the contemporary DD method. Although Ashenfelter did not use the term DD until a later paper co-authored with D. Card [1985], his approach assessed the impact of federally sponsored training on worker earnings in similar manner. He delineated two groups: the control group, which consisted of individuals ineligible for training, and the study group, comprising beneficiaries of the training program. The critical event, i.e., the receipt of training, was identified as occurring in 1964 for the study group. This research demonstrated that public funding for training programs could, in the long run, equalize earnings between the study and control groups. G. Imbens and J. Woolridge [2009] attribute the widespread adoption of DD models in empirical economics to Ashenfelter's contributions.

### Further implementations of DD models in public economics

Ashenfelter was followed by a group of prominent economists, including J.D. Angrist [1990, 1991], who studied, among other things, the labour market situation of Vietnam War veterans. He also conducted methodological research that directly or indirectly contributed to the development of DD models. Among other things, he developed models based on instrumental variables (abbreviated as IV models), which are closely related to DD models [J.D. Angrist 2006; J.D. Angrist et al. 1996, 2000; J.D. Angrist and G. Imbens 1995; J.D. Angrist and M. Kolesár 2023]. Together with A.B. Krueger [1991] also studied the impact of the establishment of compulsory schooling on education and earnings. In their study they used the DD model based on Ashenfelter's work. Angrist and Krueger [1999] also authored a chapter in the *Handbook of Labor Economics*, where they formalised the notation of the DD model and described it from the theoretical view.

Nobel laureate Esther Duflo also contributed to the methodology of DD models. She conducted an evaluation of the effect of a school-building programme in Indonesia, called INPRES, as observed on the change on education and wages of graduates [E. Duflo 2001]. This was the first of many studies using DD models in area of education [S.M.S. Halkiewicz 2023]. Together with a research staff from the National Bureau of Economic Research (NBER), she also attempted to construct a test for the significance of DD model error [M. Bertrand et al. 2004].

## 18.2. Applications of DD models

Although DD models have not historically been among the most popular, their use was initially confined to clinical research, epidemiological studies and the evaluation of public programmes. However, in recent times, their application has expanded considerably, with an array of scientific publications now available on the subject (including articles outlining the methodology and examining its technical aspects), along with numerous reviews of the literature to systematise knowledge on DD models. Nevertheless, the majority of articles addressing DD models are case studies, predominantly from the health sciences.

A good introduction to Difference-in-Differences (DD) models can be found in dedicated S. Cunningham's [2020] monograph chapter, which covers theoretical foundations, increasing popularity, and comparative applicability. Additional foundational sources include J.D. Angrist and J.S. Pischke [2008] and J. Wooldridge [2012].

A. Fredriksson and G.M. Oliveira [2019] explore the computational aspects of DD models, including standard errors and applications in management science. D. Fougère and N. Jacquemet [2023] discuss DD models in the context of public program evaluation. C. Wing et al. [2018] provide methodological guidance for designing DD experiments, particularly in clinical and medical research.

## Applications in medicine and health sciences

As mentioned previously, DD models have their origins in medical science. It is therefore in this field that DD models are most commonly in use today. Examples include studies on the effectiveness of pneumococcal vaccines [S.M. Jung et al. 2018], the impact of vision problems in Koreans on socioeconomic status [H. Kim et al. 2021], and the impact of dementia monitoring programmes in the health system [S.J. Lee et al. 2019]. The model has also been applied to studies in geriatrics [Z. Ye and Y. Jiang 2022] and nutrition [C.E. Caspi et al. 2021].

A lot of research has focused on the evaluation of Medicaid, the attempt at public financing of health care in the US. Medicaid's effectiveness was studied:

- in the reduction of gynaecological cancer mortality [S.P. Huepenbecker et al. 2022],
- in providing better access to medical transportation for patients in the primary care setting [K.H. Chaiyachati et al. 2018],
- in smoking prevention [K.E. Hilts et al. 2021],
- or in reducing the financial burden on beneficiaries [H. Gotanda et al. 2020].

It is also worth mentioning a study that aimed to determine whether the English approach to organising the public health system reduces inequalities in access to care [Y. Hu et al. 2016]. Likewise, a study examined whether a comprehensive health insurance policy, introduced on an experimental basis in the state of Massachusetts in 2006, improves the overall health of the population [C.J. Courtemanche and D. Zapata 2014].

## Applications in managerial sciences and business

It was not until the second decade of the 20th century that DD models began to be used in management science. One of the least conventional topics studied using DD models was comparing the propensity to take banking risks between different board compositions of a firm [A.N. Berger et al. 2014] or comparing the acquisition predisposition of male-majority and female-majority boards [G. Chen et al. 2016].

L. Pierce et al. [2015] investigated the impact of workplace surveillance technology on employee productivity and employee theft rates. DD models have also been used to study events affecting innovation in the firm [V.A. Aggarwal and D.H. Hsu 2014; C. Flammer and A. Kacperczyk 2016; J. Singh and A. Agrawal 2011] or the impact of lean methods on labour standards [G. Distelhorst et al. 2017], the relationship between consultant hiring and manager salaries [M.J. Conyon et al. 2019], and the impact of obtaining government certification on productivity [V. Bruno et al. 2016]. Furthermore, DD models have been used to analyse the impact of introducing an environmental tax on industrial air pollution [P. He and B. Zhang 2018].

## 18.3. DD models from an econometric perspective

The premise of DD models is to compare the magnitude of the metric of interest in states before and after the event of interest to the researcher. At the same time, a comparison is made between the trend under study and another trend that is not directly affected by the event, but which followed a similar time course up to the event. In this way it is possible to observe what values the series under study would have taken if the event had not occurred. The control group (or control series) is this parallel, unaffected time series. In practice, it is common to create two dichotomous variables, respectively:

- 0 if the observation took place before the event and 1 if after ($Time$ variable),
- 0 if the observation is in the control group and 1 if it is in the experimental group ($Affected$ variable).

Other slightly different approaches are also used, such as the replacement of the control group and the experimental group by a variable, also dichotomous, which describes the intensity of the phenomenon in the group. For example, E. Duflo [2001] makes such a distinction. A modified version of the DD model also allows examining situations where observations were affected by an event at different times [A. Goodman-Bacon 2021].

### Calculations using intragroup averages

The conventional approach to modelling is to compare intragroup means. The mean value of a metric is compared between the control and study groups in the pre- and post-event periods, yielding four different means. The differences between the means, differentiated by the values of the time and affected variables respectively, are then calculated. The final value of the $DD_e$ coefficient (spelling taken from the work of O. Ashenfelter [1978], where $DD_e$ is the empirical coefficient, to be distinguished from DD as the name of the method) is the difference of the previously calculated differences. The way in which this value is calculated is also the source of the name of the method - Difference-in-Differences. Table 18.1 provides an intuitive illustration of the calculation process.

**Tab. 18.1. Demonstration of $DD_e$ calculation process via differences table**

| Time | Affected | | Difference |
|---|---|---|---|
| | 1 | 0 | |
| 1 | $\bar{y}_1$ | $\bar{y}_2$ | $\bar{y}_1 - \bar{y}_2$ |
| 0 | $\bar{y}_3$ | $\bar{y}_4$ | $\bar{y}_3 - \bar{y}_4$ |
| Difference | $\bar{y}_1 - \bar{y}_3$ | $\bar{y}_2 - \bar{y}_4$ | $DD_e$ |

Source: own calculations.

The value of the $DD_e$ coefficient can be determined in two ways, dependent upon the order in which the averages are subtracted. However, different values for the component differences will be yielded, yet the difference between them will remain constant. This can be demonstrated through the use of equation 18.1:

$$DD_e = (\bar{y}_1 - \bar{y}_3)\text{-}\ (\bar{y}_2 - \bar{y}_4) = \bar{y}_1 - \bar{y}_3 - \bar{y}_2 + \bar{y}_4 = (\bar{y}_1 - \bar{y}_2) - (\bar{y}_3 - \bar{y}_4) = DD_e. \tag{18.1}$$

The formula for the $DD_e$ coefficient 18.1 can therefore be formally written as the differences in the conditional expected values of the variable $y$ (the measure being studied), as shown in equation 18.2 and equation 18.3 (respectively, depending on the order of subtraction):

$$DD_\mathrm{e} = [E(y|Affected = 1 \wedge Time = 1) - E(y|Affected = 1 \wedge Time = 0)] - [E(y|Affected = 0 \wedge Time = 1) - E(y|Affected = 0 \wedge Time = 0)], \tag{18.2}$$

$$DD_e = [E(y|Affected = 1 \wedge Time = 1) - E(y|Affected = 0 \wedge Time = 1)] - [E(y|Affected = 1 \wedge Time = 0) - E(y|Affected = 0 \wedge Time = 0)]. \tag{18.3}$$

## Assumptions of DD model

The validity and meaningfulness of a DD model are contingent on the fulfilment of certain assumptions. These assumptions serve as the foundation for the model's conceptualisation and interpretation.

Among the assumptions, the assumption of trend parallelism is of paramount importance, as it underpins the very notion of a DD model. This implies that it should be guaranteed that the time series under examination in the test group would have exhibited a comparable trend to that of the time series in the control group had the economic occurrence under investigation not occurred. Regrettably, this assumption can never be tested with absolute certainty, since it would necessitate an insight into the trend that would prevail in the absence of the aforementioned occurrence, which would effectively render any further modelling superfluous. In most cases, therefore, trend parallelism is tested not in the period following the event but in the period preceding it. For this purpose, separate statistical tests have been constructed [P. Basu and D.S. Small 2020; G.W. Imbens and J.M. Wooldridge 2009; J.M. Riveros-Gavilanes 2023]. Another formal testing approach is to build a placebo model, calculating the $DD_e$ coefficient using only data from before the event under study. The parallelism assumption can then be assumed to be satisfied if the coefficient is insignificantly different from zero. This approach was employed by P. Schnabl [2012], and a similar solution was also utilised by C.J. Courtemanche and D. Zapata [2014]. In fact, in the majority of studies, the trends are not perfectly parallel, therefore excessive scrupulousness in testing their parallelism is not indicated, as this may lead to the rejection of the time series unnecessarily. It is not uncommon for the literature to omit any attempt to quantify the significance of trend parallelism. Instead, the justification for the selection of the time series to be tested is often provided in a substantive manner, or the parallelism of the trends is tested visually [P.J. Gertler et al. 2016]. It is worth noting that if the discrepancies in parallelism are the result of known one-time events that will not occur in the future, it is acceptable to omit such disturbances in testing [J.M. Riveros-Gavilanes 2023].

In the DD model, it is assumed that the groups under consideration, designated as control and experimental, are stable over time. It is further assumed that observations within these groups are not transferred from one group to another [M. Lechner 2011]. In other words, it is not possible for an individual who is in the control group to move to the experimental group, and conversely. In the case of field studies conducted among local residents, this implies that an individual in the control group must be entirely omitted from the observation record upon their relocation to the affected area. A comparable exclusion must be applied when observations from otherwise delineated groups are combined. The final assumption unique to DD models pertains to the exogeneity of dichotomous variables such as time and affected, as well as the lack of influence of the studied event on their structure and distribution, which is most often evaluated in a substantive manner. A comprehensive examination of the underlying assumptions of DD models and the methods by which they can be validated is presented by M. Lechner [2011], among others.

## $DD_e$ coefficient as a structural parameter of the OLS model

In theory, the assumptions of the DD model can be directly translated into the assumptions of the OLS regression model and are sufficient to carry out modelling with (due to the dichotomous nature of the explanatory variables, the conditions for sphericity and normality of the residuals do not apply). Nevertheless, in practice, rigorous autocorrelation tests of the residuals in DD estimation are often omitted [M. Bertrand et al. 2004], thus rendering the assumptions necessary to construct a valid multivariate regression model untested. Consequently, it is possible that a model constructed on such data may be erroneous from a substantive standpoint. Furthermore, as previously stated, the most crucial and pivotal assumption in estimating a Difference-in-Differences model by any method is that the trends of the experimental and control groups are parallel. This assumption is not required in ordinary least squares (OLS). Nevertheless, it is a common practice to estimate the $DD_e$ coefficient using such a regression. One of the advantages of this approach is that it allows for additional aspects of interpretation to be provided by the model. For instance, the Student's t-test statistic for the statistical significance of the estimator of the structural parameter of the model equation can be taken directly as a reflection of the test for the statistical significance of the effect of the event under study. In order to perform such an estimation, it is necessary to construct a model in which the metric under study is designated as the explanatory variable, while the explanatory variables are $Time$, $Affected$, and their product interaction (which is also a dichotomous variable). The form of the model thus defined is shown in equation 18.4:

$$y_t = \alpha_0 + \alpha_1 \cdot Time + \beta \cdot (Time \cdot Affected) + \alpha_2 \cdot Affected. \quad (18.4)$$

The calculated coefficients are equal to the average change in the value of the dependent variable when the corresponding independent variable increases by one, under the assumption of *ceteris paribus*. The value of the free expression $\alpha_0$ is equal to the average value of the metric if all variables in the model are equal to zero, which is equal to $\overline{y}_4$ shown in table 18.1:

$$\alpha_0 = E(y|Affected = 0 \wedge Time = 0 \wedge Time \cdot Affected = 0) = E(y|Affected = 0 \wedge Time = 0) = \overline{y}_4. \quad (18.5)$$

The values of the coefficients $\alpha_1$ and $\alpha_2$ can be interpreted as a change in the average value of the metric relative to the level expressed by the constant if the variable $Time = 1$ or $Affected = 1$:

$$\alpha_1 = E(y|Affected = 1 \wedge Time = 0) - E(y|Affected = 0 \wedge Time = 0) = \overline{y}_3 - \overline{y}_4, \quad (18.6)$$

$$\alpha_2 = E(y|Affected = 0 \wedge Time = 1) - E(y|Affected = 0 \wedge Time = 0) = \overline{y}_2 - \overline{y}_4. \quad (18.7)$$

The coefficient $\beta$ represents the change in the mean value of a metric relative to a constant value if $Time = 1$, $Affected = 1$, minus the variance already explained by coefficients $\alpha_0$, $\alpha_1$, and $\alpha_2$. The relationship can thus be represented by an equation, namely 18.8, which is equal to coefficient $DD_e$, as demonstrated in equation 18.9, using notation from tab. 18.1:

$$\beta = E(y|Affected = 1 \wedge Time = 1) - E(y|Affected = 0 \wedge Time = 0) - [E(y|Affected = 1 \wedge Time = 0) -$$

$$E(y|Affected = 0 \wedge Time = 0)] - [E(y|Affected = 0 \wedge Time = 1) - E(y|Affected = 0 \wedge Time = 0)], \tag{18.8}$$

$$\beta = \overline{y}_1 - \overline{y}_4 - (\overline{y}_3 - \overline{y}_4) - (\overline{y}_2 - \overline{y}_4) = \overline{y}_1 - \overline{y}_4 - \overline{y}_3 - \overline{y}_2 + 2 \cdot \overline{y}_4 = \overline{y}_1 - \overline{y}_3 - \overline{y}_2 + \overline{y}_4 = (\overline{y}_1 - \overline{y}_3) - (\overline{y}_2 - \overline{y}_4) = DD_e. \tag{18.9}$$

### Interpreting the $DD_e$ coefficient

The $DD_e$ coefficient is interpreted as the average change in the value of the studied metric as a result of the occurrence of the studied event, with the influence of other factors shaping its value, such as trend or other unstudied events, being neglected (under the assumption of trend parallelism) [J.A. Martínez 2021]. This value is also referred to in the literature as the ATE (Average Treatment Effect) or ATET (Average Treatment Effect for the Treated), which is sometimes also abbreviated as ATT. This terminology is derived from randomised trials.

### Significance testing of $DD_e$ coefficient

The evolution of the method and its growing application to diverse scientific disciplines necessitated the development of a systematic approach to assess the impact of a given event on a quantifiable metric. As previously discussed, the initial solution to this challenge involved the utilization of the Student's t-test for determining the statistical significance of the estimator of the structural parameter of the $Time \cdot Affected$ interaction dichotomous variable. Nevertheless, this approach proved to be flawed due to the lack of meaningful insight gained from interpreting the results in the context of the event itself. The initial step in the evolution of an alternative method for assessing the significance of the $DD_e$ coefficient was therefore to evaluate the statistical significance of its standard error and autocorrelation. A research team at the National Bureau of Economic Research (NBER) conducted a comprehensive review of methods for testing and correcting model errors in over 90 different research papers [M. Bertrand et al. 2004]. The aforementioned methods employed bootstrapping [B. Efron and R.J. Tibshirani 1994], variance and covariance matrix testing [G. Kezdi 2011], or the use of placebo intervention [R.M. Bell and D.F. McCaffrey 2002], among others. The NBER research team employed a placebo intervention approach in a Monte Carlo study to quantify the relative impact of model standard error [M. Bertrand et al. 2004]. This investigation led to the formulation of generalised assumptions for an iterative test of the statistical significance of an event's effect on a study metric [S.M.S. Halkiewicz 2023, 2024] and can be further studied in the future.

## 18.4. Conclusions

The DD models represent an effective method for analysing the effects of events on the metrics under study. The methodology can be applied in a multitude of scientific disciplines, extending beyond the field of economics. The intensive development of DD models in the 21st century has resulted in the constant emergence of statistical and econometric tools, as well as software, to improve the analyses that can be conducted using them [S. Cunningham 2020]. Moreover, the models are relatively straightforward to utilise, which may prove advantageous for researchers lacking expertise in quantitative techniques. The aim of this work was to explain the DD methodology, its application and history to a wider academic audience. By extensive literature review and explanation of the underlying methodology, this aim was accomplished.

## Bibliography

Aggarwal V.A., Hsu D.H., 2014, *Entrepreneurial exits and innovation*, Management Science, vol. 60(4), pp. 867-887.

Angrist J.D., 1990, *Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records*, American Economic Review, vol. 80(3), pp. 313-336.

Angrist J.D., 1991, *The draft lottery and voluntary enlistment in the Vietnam Era*, Journal of the American Statistical Association, vol. 86(415), pp. 584-595.

Angrist J.D., 2006, *Instrumental variables methods in experimental criminological research: What, why and how*, Journal of Experimental Criminology, vol. 2(1), pp. 23-44.

Angrist J.D., Graddy K., Imbens G.W., 2000, *The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish*, Review of Economic Studies, vol. 67(3), pp. 499-527.

Angrist J.D., Imbens G.W., 1995, *Two-stage least squares estimation of average causal effects in models with variable treatment intensity*, Journal of the American Statistical Association, vol. 90(430), pp. 431-442.

Angrist J.D., Imbens G.W., Rubin D.B., 1996, *Identification of causal effects using instrumental variables*, Journal of the American Statistical Association, vol. 91(434), pp. 444-455.

Angrist J.D., Kolesár M., 2023, *One instrument to rule them all: The bias and coverage of just-ID IV*, Journal of Econometrics, vol. 240(2), 105398.

Angrist J.D., Krueger A.B., 1991, *Does compulsory school attendance affect schooling and earnings?*, The Quarterly Journal of Economics, vol. 106(4), pp. 979-1014.

Angrist J.D., Krueger A.B., 1999, *Empirical strategies in labor economics*, [in:] *Handbook of Labor Economics (3rd volume, part A)*, O. Ashenfelter i D. Card (eds), Elsevier, Amsterdam.

Angrist J.D., Pischke J.S., 2008, *Parallel worlds: fixed effects, Differences-in-Differences, and panel data*, [in:] *Mostly harmless econometrics: An empiricist's companion*, J.D. Angrist and J.S. Pischke (eds), Princeton University Press, Princeton, NJ.

Ashenfelter O., 1978, *Estimating the Effect of Training Programs on Earnings*, The Review of Economics and Statistics, vol. 60(1), pp. 47-57.

Ashenfelter O., Card D., 1985, *Using the longitudinal structure of earnings to estimate the effect of training programs*, The Review of Economics and Statistics, vol. 67(4), pp. 648-660.

Basu P., Small D.S., 2020, *Constructing a more closely matched control group in a difference-in-differences analysis: Its effect on history interacting with group bias*, Observational Studies, vol. 6(1), pp. 103-130.

Bell R.M., McCaffrey D.F., 2002, *Bias reduction in standard errors for linear regression with multi-stage samples*, Survey Methodology, vol. 28(12), pp. 169-181.

Berger A.N., Kick T., Schaeck K., 2014, *Executive board composition and bank risk taking*, Journal of Corporate Finance, vol. 28, pp. 48-65.

Bertrand M., Duflo E., Mullainathan S., 2004, *How much should we trust Differences-In-Differences estimates?*, The Quarterly Journal of Economics, vol. 119(1), pp. 249-275.

Bruno V., Cornaggia J., Cornaggia K.J., 2016, *Does regulatory certification affect the information content of credit ratings?*, Management Science, vol. 62(6), pp. 1578-1597.

Cameron D., Jones I.G., 1983, *John Snow, the broad street pump and modern epidemiology*, International Journal of Epidemiology, vol. 12(4), pp. 393-396.

Caniglia E.C., Murray E.J., 2020, *Difference-in-Difference in the time of cholera: a gentle introduction for epidemiologists*, Current Epidemiology Reports, vol. 7(4), pp. 203-211.

Caspi C.E., De Marco M., Durfee T., Oyenuga A., Chapman L., Wolfson J., Myers S., Harnack L.J., 2021, *A Difference-in-Difference study evaluating the effect of minimum wage policy on body mass index and related health behaviors*, Observational Studies, vol. 7(2), pp. 1-19.

Chaiyachati K.H., Hubbard R.A., Yeager A., Mugo B., Shea J.A., Rosin R., Grande D., 2018, *Rideshare-based medical transportation for Medicaid patients and primary care show rates: A Difference-in-Difference Analysis of a pilot program*, Journal of General Internal Medicine, vol. 33(6), pp. 863-868.

Chen G., Crossland C., Huang S., 2016, *Female board representation and corporate acquisition intensity*, Strategic Management Journal, vol. 37(2), pp. 303-313.

Conyon M.J., Hass L.H., Peck S.I., Sadler G.V., Zhang Z., 2019, *Do compensation consultants drive up CEO pay? Evidence from UK public firms*, British Journal of Management, vol. 30(1), pp. 10-29.

Courtemanche C.J., Zapata D., 2014, *Does universal coverage improve health? The Massachusetts experience*, Journal of Policy Analysis and Management, vol. 33(1), pp. 36-69.

Cunningham S., 2020, *Difference-in-Differences*, [in:] *Causal Inference: The Mixtape*, S. Cunningham (ed.), Yale University Press, New Haven, CT.

Distelhorst G., Hainmueller J., Locke R.M., 2017, *Does lean improve labor standards? Management and social performance in the Nike supply chain*, Management Science, vol. 63(3), pp. 707-728.

Duflo E., 2001, *Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment*, American Economic Review, vol. 91(4), pp. 795-813.

Efron B., Tibshirani R.J., 1994, *An Introduction to the Bootstrap*, Chapman and Hall/CRC, New York, NY.

Flammer C., Kacperczyk A., 2016, *The impact of stakeholder orientation on innovation: Evidence from a natural experiment*, Management Science, vol. 62(7), pp. 1982-2001.

Fougère D., Jacquemet N., 2023, *Méthode des doubles différences (difference-in-differences)*, [in:] *Méthodes et approches en évaluation des politiques publiques*, A. Revillard (ed.), Éditions science et bien commun, Quebec City.

Fredriksson A., Oliveira G.M., 2019, *Impact evaluation using Difference-in-Differences*, RAUSP Management Journal, vol. 54(4), pp. 519-532.

Gertler P.J., Martinez S., Premand P., Rawlings L.B., Vermeersch C.M.J., 2016, *Impact evaluation in practice (2nd edition)*, Inter-American Development Bank, Washington, DC.

Goodman-Bacon A., 2021, *Difference-in-differences with variation in treatment timing*, Journal of Econometrics, vol. 225(2), pp. 254-277.

Gotanda H., Jha A.K., Kominski G.F., Tsugawa Y., 2020, *Out-of-pocket spending and financial burden among low income adults after Medicaid expansions in the United States: quasi-experimental difference-in-difference study*, British Medical Journal, vol. 368(40), pp. 1-9.

Halkiewicz S.M.S., 2023, *Modele Difference-in-Differences w ewaluacji wpływu zdarzeń ekonomicznych na poprawę edukacji*, [in:] *Współczesne problemy społeczno-ekonomiczne w ujęciu analitycznym*, A. Stanimir (ed.), Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.

Halkiewicz S.M.S., 2024, *Simulated test for the significance of DD estimator with proper randomization*, http://dx.doi.org/10.13140/RG.2.2.18067.49443 (access date: 22.05.2024).

He P., Zhang B., 2018, *Environmental tax, polluting plants' strategies and effectiveness: Evidence from China*, Journal of Policy Analysis and Management, vol. 37(3), pp. 493-520 .

Hilts K.E., Blackburn J., Gibson P.J., Yeager V., Halverson P., Menachemi N., 2021, *Impact of Medicaid expansion on smoking prevalence and quit attempts among those newly eligible, 2011–2019*, Tobacco Prevention & Cessation, vol. 7(8), pp. 1-9

Hu Y., van Lenthe F.J., Judge K., Lahelma E., Costa G., de Gelder R., Mackenbach J.P., 2016, *Did the English strategy reduce inequalities in health? A difference-in-difference analysis comparing England with three other European countries*, BMC Public Health, vol. 16(865), pp. 1-12.

Huepenbecker S.P., Fu S., Sun C.C., Zhao H., Primm K.M., Giordano S.H., Meyer L.A., 2022, *Medicaid expansion and 2-year survival in women with gynecologic cancer: a difference-in-difference analysis*, American Journal of Obstetrics and Gynecology, vol. 227(3), pp. 482-e1.

Imbens G.W., Wooldridge J.M., 2009, *Recent developments in the econometrics of program evaluation*, Journal of Economic Literature, vol. 47(1), pp. 5-86.

Jung S.M., Lee H., Nishiura H., 2018, *The impact of pneumococcal vaccination on pneumonia mortality among the elderly in Japan: a difference-in-difference study*, PeerJ, vol. 6(e6085), pp. 1-17.

Kezdi G., 2011, *Robust standard error estimation in fixed-effects panel models*, SSRN Electronic Journal, special number 9, pp. 95-116.

Kim H., Koo H., Han E., 2021, *Socioeconomic and physical health status changes after visual impairment in Korea using difference-in-difference estimations*, Scientific Reports, vol. 11(820).

Lechner M., 2011, *The estimation of causal effects by Difference-in-Difference methods*, Foundations and Trends® in Econometrics, vol. 4(3), pp. 165-224.

Lee S.J., Seo H.J., Lee D.Y., Moon S.H., 2019, *Effects of a dementia screening program on healthcare utilization in South Korea: A Difference-In-Difference analysis*, International Journal of Environmental Research and Public Health, vol. 16(20), 3837.

Lester R.A., 1946, *Shortcomings of marginal analysis for wage-employment problems*, The American Economic Review, vol. 36(1), pp. 63-82.

Martínez J.A., 2021, *Interpretability for conditional average treatment effect estimation*, University of Cambridge, Cambridge.

Newsom S.W.B., 2006, *Pioneers in infection control: John Snow, Henry Whitehead, the Broad Street pump, and the beginnings of geographical epidemiology*, Journal of Hospital Infection, vol. 16(3), pp. 210-216.

Pierce L., Snow D.C., McAfee A., 2015, *Cleaning house: The impact of information technology monitoring on employee theft and productivity*, Management Science, vol. 61(10), pp. 2299-2319

Riveros-Gavilanes J.M., 2023, *Testing parallel trends in Differences-in-Differences and event study designs: A research approach based on pre-treatment period significance*, Journal of Research, Innovation and Technologies, vol. 2(4), pp. 226-237.

Schnabl P., 2012, *The international transmission of bank liquidity shocks: Evidence from an emerging market*, Journal of Finance, vol. 67(3), pp. 897-932.

Singh J., Agrawal A., 2011, *Recruiting for ideas: How firms exploit the prior inventions of new hires*, Management Science, vol. 57(1), pp. 129-150.

Snow J., 1857, *On the origin of the recent outbreak of cholera at west ham*, British Medical Journal, vol. 45(1), 934.

Snowise N.G., 2023, *Memorials to John Snow – Pioneer in anaesthesia and epidemiology*, Journal of Medical Biography, vol. 31(1), pp. 47-50.

Tulchinsky T.H., 2018, *John Snow, cholera, the Broad Street pump; waterborne diseases then and now*, [in:] *Case Studies in Public Health*, T.H. Tulchinsky (ed.), Elsevier, Amsterdam.

Wing C., Simon K., Bello-Gomez R.A., 2018, *Designing Difference in Difference studies: Best practices for public health policy research*, Annual Review of Public Health, vol. 39, pp. 453-469.

Wooldridge J., 2012, *Pooling cross sections across time: Simple panel data methods*, [in:] *Introductory econometrics: a modern approach (5th edition)*, J. Sabatino, E. Joyner and M. Worls (eds), Cengage Learning, Andover.

Ye Z., Jiang Y., 2022, *The impact of a pilot integrated care model on the quality and costs of inpatient care among Chinese elderly: a difference-in-difference analysis of repeated cross-sectional data*, Cost Effectiveness and Resource Allocation, vol. 20(1), pp. 1-9.