CENTRALITY-ORIENTED CAUSALITY.
A STUDY OF EU AGRICULTURAL SUBSIDIES
AND DIGITAL DEVELOPMENT IN POLAND

DANIEL KOSIOROWSKI[1], JERZY P. RYDLEWSKI[2*]

[1]Department of Statistics, Cracow University of Economics,
ul. Rakowicka 27, 31-510 Kraków, Poland

[2]Faculty of Applied Mathematics, AGH University of Science and Technology,
al. Mickiewicza 30, 30-059, Kraków, Poland

Results of a convincing causal statistical inference related to socio-economic phenomena are treated as an especially desired background for conducting various socio-economic programs or government interventions. Unfortunately, quite often real socio-economic issues do not fulfil restrictive assumptions of procedures of causal analysis proposed in the literature. This paper indicates certain empirical challenges and conceptual opportunities related to applications of procedures of data depth concept into a process of causal inference as to socio-economic phenomena. We show how to apply statistical functional depths to indicate factual and counterfactual distributions commonly used within procedures of causal inference. Thus, a modification of Rubin causality concept is proposed, i.e., a centrality-oriented causality concept. The presented framework is especially useful in the context of conducting causal inference based on official statistics, i.e., on the already existing databases. Methodological considerations related to extremal depth, modified band depth, Fraiman-Muniz depth, and multivariate Wilcoxon sum rank statistic are illustrated by means of example related to a study of an impact of EU direct agricultural subsidies on digital development in Poland in the period 2012–2018.

Keywords: *causal inference, counterfactuals, statistical depth function, robust statistical inference, digital development*

1. Introduction

In the history of scientific and economic thought, in particular, one may find various statistical frameworks developed for expressing the necessary assumptions under which statistical results can be endowed with causal interpretation. There is an agreement that causes should tell us not only that two phenomena are related, but also how they are

---

*Corresponding author, email address: ry@agh.edu.pl

related. They allow us to make a reliable prediction about the future, explain the relationships between and occurrence of events, and enable one to develop effective policies intervention. Causal knowledge should exhibit a certain kind of stability, resistance concerning circumstances under which a phenomenon is observed, and parameters of the environment. Causality alone is an insufficient basis for undertaking further scientific or decisive acts. To use causes to predict, explain or alter behaviour effectively, we must also know the time over which a relationship takes place, the probability with which it will occur, and how other factors alter its efficacy. We need to know when the effect starts and how long it persists. Causal relationships depend on time scales as relationships may persist a while or a long time.

In economics, for several content-matter justified reasons, causal relations, which are characteristic of an influential majority of objects or agents, are of prime importance [8]. An effective way of defining this influential majority of economic objects is provided by statistical depth function [35], which is a basic notion used within a discipline of multivariate statistics called data depth concept (DDC). Using it, one may propose a very useful multivariate [19, 20, 35] or functional [9, 23–25] generalisations of one-dimensional statistical procedures based on order statistics and ranks and, in consequence, robust causal inferential procedures. Statistical depth function provides a centre-outward order of multivariate or functional objects. It expresses the centrality of an object as a number from an interval [0, 1], with values close to 0 treated as peripheral, and values close to 1 as central. Thus, one may consider causal relationships of objects, for which measures of departure from centres are not bigger than prefixed thresholds. This is a natural way of focusing our attention on the influential majority of economic objects.

DDC provides a variety of useful statistical procedures [15] but despite excellent findings obtained in recent years [6, 22] its procedures still exhibit significant limitations in the context of conducting empirical research based on relatively small, sparse datasets drawn from resources from statistical offices, with observations often exhibiting "malicious configurations".

One of the aims of this paper is to critically discuss selected chances and limitations related to applications of certain tools of DDC in the case of robust empirical causal analysis concerning two spatial economic phenomena, namely agricultural subsidies (treatment), and digital development of a country (effect). In the considered empirical example, there is a functional dataset as an input set, and a multivariate dataset describing digital development as an output set. Causal inference scheme should be adjusted to these datasets. As it is virtually impossible to repeat the experiment, we treat the typical observations as factual and atypical observations as counterfactual.

Note that knowledge of such "centre focused causality" seems to be especially desirable in the context of performing social programs dedicated to typical objects (e.g., middle class stimulating programme) or for decreasing a fraction of certain atypical objects (e.g., avoiding poverty programme).

In statistical literature, one of the most widely recognised conceptualisations of causal inference is Rubin's potential outcome representation [4, 5, 30]. Rubin postulates that causal statements can only be derived if one additionally considers what would have happened if an object had experienced something different than its experience. Another especially influential concept in economic and econometric literature is Granger's non-causality (GNC) [12] and its variants or generalisations. Its empirical version, in essence, may be expressed as a prediction error approach. Assuming a specific implementation of Granger's non-causality, one can easily indicate depth-based estimation or testing procedure replacing least squares or maximal likelihood principles used by default. Next influential approach to causality analysis is probabilistic approaches [27, 14]. The basic probabilistic theory of causality is that C is a cause of E if

$$P(E\,|\,C) > P(E\,|\sim C) \tag{1}$$

where $P(E\,|\,C)$ denotes the conditional probability of E under condition C, $\sim$ denotes complement of an event.

Referring to the DDC, it is straightforward to consider this condition within a central region of a certain degree, and hence to focus on a certain majority of objects.

2. Applying DDC in the causal inference

Nowadays, for accepting or rejecting the existence of a certain causal relationship between phenomena, one needs to present convincing empirical evidence obtained in the process of statistical inference. Generally, to identify a phenomenon X as a cause of a phenomenon Y, one has to demonstrate that variability in X produces variability in Y. In this context, one may consider various kinds of variations: across time, across individuals, across characteristics, across groups, and across intervention versus observation [4, 12].

Currently, it seems to be more and more commonly recognised that properties of the statistical inference strongly depend on the quality of empirical data used in the inference as well as on fulfilling assumptions underlying the procedures [33]. Using depths, we can demonstrate how robustly measured variability of X produce robustly measured variability of Y, where X and Y are expressed as multivariate of functional random variables.

In the spirit of Rubin [30], the causal effect of one treatment E over another C for a particular unit and time interval from t_1 to t_2 there is a difference between what would have happened at a time t_2 if the unit had been exposed to E initiated at t_1 and what would have happened at a time t_2 if the unit had been exposed to C initiated at t_1.

Rubin [30] considers $2N$ units (e.g., small regions of a country), half being exposed to a control treatment C and a half to a treatment E. If treatments E and C were randomly assigned to units, $Y(E)$ denotes the value of Y measured at t_2 for the unit given the unit received experimental treatment E initiated at t_1, similarly $Y(C)$, then $Y(E) - Y(C)$ is a causal effect.

The problem in measuring $Y(E) - Y(C)$ is that we can never observe both $Y(E)$ and $Y(C)$ since we cannot return to time t_1 to give other treatment. Next problematic issue arises, as the use of truly random samples may be difficult in practice [14].

Although robustness and causality conceptually seem to be very closely tied, surprisingly, considerations binding them are rare. On the other hand, one may get an impression that authors conducting researches on causality in the area of economics implicitly assume some kind of "sample-to-sample stability" of their empirical argumentation even if in the centre of their statistical considerations stands non-robust least-squares method [7]. For proposing truly robust causality analysis, it is natural to source from concepts and developments of modern robust statistics [33]. For several merit-justified reasons related to the fact that economic systems are multidimensional by default, we focus our attention on the DDC methods.

Many excellent papers present particular DDC tools, which are of minor importance for the aims of this paper. Here, we would like to stress opportunities related to applications of depths for functional objects (e.g., subsidies trajectories [10] and depth-induced multivariate rank tests (e.g., a comparison of control and treatment groups in terms of various multivariate rank tests based on ranks induced by depths [13, 17, 19]). A choice of a specific depth in this context is a statistical as well as a merit issue, as the DDC offers a rich overview of depths in terms of a balance between effectiveness, robustness, and computational complexity [6, 22], and thanks to a locality concept [26] options as to "resolution" at which we compare phenomena. Moreover, instead of using the propensity scores [28], one may consider conducting a family of local Wilcoxon tests for a certain sequence of locality parameters $\beta_1, ..., \beta_k$ [17]. In the case of a significant difference, we expect that majority of the tests reject equality of distributions in the control and intervention group.

2.1. Depth-based outlyingness and causality

Formula (1) gives us some causality inference in a rather simple setup, where C or $\sim C$ occurs. If the probability space is more complex, (1) should be a base for some more complex inference. We are interested in assessing whether the event $\mathbf{C} = \{C$ is one of the central observations in the considered dataset with the locality parameter $\beta\}$ is a cause of A, where

locality parameter β is understood in the sense of [26], and hence it is a parameter combining probability and outlyingness. For clarity, note that $\sim\mathbf{C} = \{C$ is not one of the central observations in the considered dataset with the locality parameter $\beta\}$.

Look at the random variable, which is the difference

$$E(A\,|\,\mathbf{C}) - E(A\,|\sim\mathbf{C}) \tag{2}$$

The realisation of the above random variable is a function. If

$$\mu\{x : E(A(x)\,|\,\mathbf{C}) - E(A(x)\,|\sim\mathbf{C})\} = \mu(D)$$

then C is a cause of A, where $\mu(D)$ is a Lebesgue measure of the domain D of the random function A, and we consider only the situation where

$$E(A(x)\,|\,\mathbf{C}) - E(A(x)\,|\sim\mathbf{C}) > 0$$

We can choose a level, e.g., it can be equal to 99, 98 or 95%. It seems reasonable to treat the quantity

$$\mu\{x : E(A(x)\,|\,\mathbf{C}) - E(A(x)\,|\sim\mathbf{C})\} = \beta\mu(D)$$

as a measure of strength of causal relationship restricted to majority of objects on level β. On a theoretical level, we use a notion of conditional expected value, which is often quite difficult to calculate and estimate for functional objects [3]. As robustness stands in the centre of our considerations, we propose to use a difference between sample functional medians as estimator of the above quantity. DDC offers robust measures of multivariate location and scatter, robust regression, as well as robust goodness of fit, prediction error measures, and hence it provides natural statistical tools of causal inference.

Let $\mathbf{Y} = \{Y_1, ..., Y_k\}$ denote k-dimensional vector of our interest. The DDC enables, among others, to consider the following causal reasoning schemes:

1. Assessing the difference $\mathbf{Y}(\mathbf{E}) - \mathbf{Y}(\mathbf{C})$ using depths (see (1)).

2. For two samples of multivariate or functional objects $\mathbf{Y}^n(\mathbf{E})$ and $\mathbf{Y}^m(\mathbf{C})$, consider a difference between two depth induced sample medians $\mathrm{MED}\,[\mathbf{Y}^n(\mathbf{E})] - \mathrm{MED}\,[\mathbf{Y}^m(\mathbf{C})]$. One may expect that estimated distribution of a difference between two depth induced medians should be informative in the context of causal inference.

3. For two multivariate or functional "regression samples" $\mathbf{Z}^n(\mathbf{E})$ and $\mathbf{Z}^m(\mathbf{C})$ consider distribution of a difference between two vectors of regression parameters Θ estimated via depth-based procedure, e.g., using depth trimmed samples we would like to

estimate the distribution $\Theta[\mathbf{Y}^n(\mathbf{E})] - \Theta[\mathbf{Y}^m(\mathbf{C})]$. The distribution should inform us about the details of the treatment effect.

4. Depths enable comparisons of control and treatment groups on sets $C(\beta_i)\backslash C(\beta_j)$, where $C(\beta_i)\backslash C(\beta_j)$ denotes set subtraction of central regions $C(\beta_i)$, $C(\beta_j)$, and β_i, β_j ($\beta_i > \beta_j$) denote both the level of centrality and locality levels (i.e., central regions covering the smallest depth region with probability equal or larger than β in the sense of Paindavaine and van Bever [26], which can be interpreted as a sort of comparison on levels appealing to the comparison using propensity scores of Rosenbaum and Rubin [28].

To effectively use the above approaches in the causal inference we need an operationally feasible theory for the listed tools. Unfortunately, with several exceptions, depth-based inference procedures are conceptually more demanding than classical ones, and they require asymptotic or resampling machinery. The issue is especially evident in the case of the theory of the DDC tools dedicated to functional data [1, 2].

2.2. A framework for causality analysis using functional and multivariate depths

Without doubts, our perception of an economic phenomenon often relates to the evaluation of properties of a function over a certain continuum. One may consider probability density function of the random variable describing an income of a household, consider GDP per capita trajectory of a country during a decade, day and night number of visits of Internet users in the Internet service, or behaviour of an investor optimism indicator within a month. Reducing the whole function to a certain set of scalars (e.g., mean, variance) very often implies a significant loss of valuable information on the phenomenon and in consequence may lead to an inappropriate perception of the phenomenon. A "shape" of the consumer price index (CPI) during a month may better express an investor optimism during the considered period as a specific sequence of peaks and valleys in CPI trajectory may denote a sequence of an activity bursts and consumer hesitations, hence "a spectrum of moods" called optimism. Considering economic phenomena as functions is natural and enables us to take into account more information than in the classical scalar approach. One may associate values of economic aggregates with shapes and other properties of other functional aggregates (e.g., yield curves, Lorenz curves, fertility curves, life expectancy curves).

2.3. Proposal

We have V objects, for which characteristics $\{S^v(t)\}_{v=1,\dots,V}$ are given for $t \in (t_0, t_1)$, and thus are treated as functional objects. Moreover, a multidimensional vector representing the treatment effect for each object is given, namely $A^v \in \mathbf{R}^l$. The centrality of

an object v_0 is expressed in terms of the value of empirical functional depth for $S^{v_0}(t)$ among $\{S^v(t)\}_{v=1,...,V}$. Then, we divide original V objects into two groups, according to higher ($v \in F$), or lower ($v \in C$) value of calculated functional depth. Subsequently, rank is calculated for each vector A^v concerning a chosen multidimensional depth. In the last step, the distribution of ranks for both groups, i.e., A^v where $v \in F$ or $v \in C$, is compared, using the Wilcoxon rank-sum test. This centrality-oriented causality is a modification of the Rubin causality concept in this sense that it enables to use a DDC and its tools to define factuals and counterfactuals. Application of DDC concept to discriminate functional elements of groups F and C is the equivalent to using propensity scores of Rosenbaum and Rubin [28]. Note, however, that our approach is blind to the direction of the outlyingness of the atypical observations. Consequently, the problem is that the centrality-oriented causality scheme does not differentiate shape outliers, amplitude outliers, outliers concerning the covariance structure (for discussion of functional outliers in the economics [16, 18]. It merely separates central observations from the atypical ones for the chosen functional depth.

Note that, similarly, we can define an outlyingness-oriented causality. Namely, we divide the original V objects into two groups. The first group ($v \in F$) consists of observations with a low value of calculated functional depth, which at the same time is frequently much above (or below) a functional median. The second group ($v \in C$) consists of the remaining observations or some chosen subset of them. Note also that it seems reasonable to fix the outlyingness parameter defining groups C and F so that the number of observations in the group C is significantly smaller than the number of observations in the group F. The outlyingness parameter should depend on the considered depth.

2.4. Considered depths

In this paper, we consider a projection depth for multidimensional data as well as three different notions of functional depth. In this subsection, we give definitions of the depth functions that are applied in the empirical example.

2.4.1. Projection depth

Let X^n be a sample consisting of n observations. The projection depth $D(x; X)$ of a point $x \in \mathbf{R}^d$ coming from a distribution X is

$$D(x; X) = \frac{1}{1 + \sup_{\|u\|=1} \dfrac{|u^T x - \mathrm{Med}(u^T X)|}{\mathrm{MAD}(u^T X)}}$$

where Med denotes median and median absolute deviation

$$\text{MAD}\,(X) = \text{Med}\left(\left| X - \text{Med}\,(X) \right|\right)$$

Sample projection depth denoted by $D(x; X^n)$ is obtained by replacing X a probability distribution in $\mathbf{R}^d$ with its empirical counterpart calculated from the sample X^n [21, 34]. This depth is one of the best multivariate depths. It has many desirable properties, i.e., $D(x; X^n)$ converges uniformly and strongly to $D(x; X)$.

2.4.2. Selected functional depths

Consider a sample of N functions $\{x_1(t), ..., x_N(t)\}$ defined on the interval $[0, 1]$. Note that if the functions' domain is an interval $[T_1, T_2]$, they should be transformed. The Fraiman–Muniz depth of the function $x(t)$ is the integral [9]:

$$FM\left(x(t)\right) = \int_0^1 D_N(x(t))dt$$

where $D_N(x(t)) = 1 - \left| \dfrac{1}{2} - F_{N,t}(x(t)) \right|$, and $F_{N,t}(x(t)) = \dfrac{1}{N} \sum_{k=1}^N I(x_k(t) \le x(t))$, $I(\cdot)$ is the indicator function.

Modified band depth (MBD) [23] of the curve x concerning functional sample $\{x_1, ..., x_N\}$ estimates the curves' frequency of being in the centre of the sample. Consider sets of the following form

$$A(x; x_{i_1}, x_{i_2}) = \{t \in [0, 1] : \min_{r=i_1, i_2} x_r(t) \le x(t) \le \max_{r=i_1, i_2} x_r(t)\}$$

Finally, MBD can be defined:

$$\text{MBD}(x \mid \{x_1, ..., x_N\}) = \frac{2}{N(N-1)} \sum_{1 \le i_1 < i_2 \le N} \lambda(A(x; x_{i_1}, x_{i_2}))$$

where $\lambda(\cdot)$ is a Lebesgue measure.

To define extremal depth (ED) [25] we need to define for $t \in [0, 1]$ the pointwise depth of $x(t)$ concerning the functional sample $\{x_1, ..., x_N\}$:

$$D_x(t; \{x_1, ..., x_N\}) = 1 - \frac{\left| \sum_{i=1}^{N} I\{x_i(t) < x(t)\} - I\{x_i(t) > x(t)\} \right|}{N}$$

Let **D** denote the union of D_x overall functions x. Let Φ_x be the cumulative distribution of the different values taken by $D_x(t; \{x_1, ..., x_N\})$ as t varies in the interval [0, 1]. In other words, define

$$\Phi_x(r) = \int_0^1 I\{D_x(t; \{x_1, ..., x_N\}) \le r\} dt$$

for each $r \in$ **D**. It is called the depth CDF or d-CDF.

We now consider two functions x and y with their corresponding d-CDFs. Let $0 \le d_1 \le d_2 \le ... \le d_M \le 1$ be the ordered elements of their combined depth levels. If $\Phi_x(d_1) > \Phi_y(d_1)$, then we say that x is more extreme then y and denote the relation $x \prec y$. If $\Phi_x(d_1) = \Phi_y(d_1)$, then we consider element d_2 and compare $\Phi_x(d_2)$ and $\Phi_y(d_2)$. If $\Phi_x(d_i) = \Phi_y(d_i)$ for $i = 1, ..., M$, then the functions are considered to be equivalent in terms of depth and we denote the relation $x \sim y$. Extremal depth of function x concerning the functional sample $\{x_1, ..., x_N\}$ is:

$$\text{ED}(x \mid \{x_1, ..., x_N\}) = \frac{\#\{i : x \succeq x_i\}}{N}$$

where $x \succeq x_i$ means that either $x \succ x_i$ or $x \sim x_i$.

No single notion of functional depth will perform uniformly better than others. FM depth is the classical one, while MBD is designed to deal with time-varying functions. On the other hand, ED has some advantages that are not shared by other depth notions. Central regions obtained with ED achieve the nominal simultaneous coverage probability, and there is a correspondence between central regions based on ED and appropriate pointwise central regions. Moreover, ED is quite robust to some classes of functional outliers (for details see [25]).

3. Empirical application

In the below example we underline certain difficulties appearing in real applications of DDC-based procedures used in robust causal inference applied to phenomena responsible for the development of a country consisting of smaller units, e.g., voivodships or

counties. These issues are of prime importance for designing various government programs and evaluating their effects. Generally speaking, the conditions that we face in the practice of economic reasoning significantly depart from the conditions postulated in the theory of causal inference [5, 31].

We consider the issue of an impact of selected agricultural subsidies granted to farmers in Polish voivodships within EU funds on the digital development of Poland in the period 2012–2018. We use the data on the amount of direct payment for campaign 2007–2017 in single area payment scheme; the considered subsidies involve (1) single area payment, (2) subsidies to legumes, (3) subsidies to tomatoes, and (4) subsidies to soft fruits. The subsidies are expressed in currency (PLN). The digital development of Poland is measured via selected available annual indicators of the information society, innovation activity, and a degree of dissemination of a telecommunication net, namely (1) expense on R & D in each voivodship expressed in PLN, (2) the percentage of companies that possess their web portal, (3) the percentage of companies obtaining purchase orders through their web portals, (4) percentage of companies equipping their employees with mobile computers, (5) main telephone lines expressed in kilometres, and (6) number of production lines controlled with computers. The annual data for each voivodship is retrieved from Statistics Poland [32].

Note that Polish voivodships are considered to be homogeneous, i.e., since their socio-economical characteristics are very similar, and they are subject to the same legal system. That is why we assume that stable unit-treatment value assumption is fulfilled as far as it is possible [5, 31]. Methods used within our considerations may easily be broadened to similar issues. In other words, to show an application of the proposed approach, we consider a problem of evaluation of the impact of EU agricultural subsidies (an indirect intervention/treatment) for a certain kind of regional development of a country, namely digital development of Poland.

In "the closest to ideal" circumstances we could observe a representative random sample drawn from all the smallest sub-regions of Poland (i.e., 2477 communities) concerning an objective measure of digital development at the time t_0, when a subsidy is taken (in a situation of being exposed to a treatment) and at a certain time $t_1 > t_0$, and could observe the same collection of communities concerning the same measure of digital development at moments t_0, t_1 when subsidies are not taken (control group). In practice, however, as based on Statistics Poland services [32], we have a pair of short already realised time series of treatment variables and response variables concerning bigger sub-regions of higher administrative level of a country (i.e., 16 Polish voivodships). In a general spirit of an approach of Gill and Robins [11], we decide to propose the following robust causal inference scheme. For several organisational and ethical reasons, our degree of control over the "experiment" is very low. As very often in the practice of economic studies, we are given data, and then we are looking for an appropriate model for them. In the causal inference, we additionally have to take into account

discrimination between a factual (what happens due to the treatment) and the counter-factual (what would have happened if an alternative treatment had been taken) distribution of a variable expressing an effect of a treatment. One of the main merit tasks in this context is to indicate a reliable alternative for the treatment.

In the considered empirical example, other problems appear. Firstly, the number of objects is relatively small. Secondly, the number of observations per unit is very small. Namely, we are given the data consisting of 4 types of agricultural subsidies obtained from each of 16 Polish voivodships (regions) every year in the period 2012–2018, and we consider 6 available variables representing the digital development of a region. According to the proposal described in 2.3, in the first step, we define a certain synthetic variable $S^v = h(S_1^v, ..., S_4^v)$ representing all 4 types of agricultural subsidies and calculate its value for each voivodship, where h in our example is a sum of all subsidies divided by a voivodship's population in the year. We treat S^v as a functional object, and hence we have 16 trajectories of the functional variable S. Furthermore, for each voivodship $v = 1, ..., 16$ we have a vector variable $A^v = (A_1, ..., A_6)$, where $A_i \in \mathbf{R}$, and where coordinates of the vector represent measures of certain aspects of digital development of the voivodship. Note that components of the aggregate A^v are incomparable. In our case, they are expenses expressed in currency (PLN), the number of some goods or fraction of companies fulfilling certain conditions connected with digital development. In this context, an application of multidimensional rank tests seems to be a reasonable solution. Following our general conceptual scheme, we replace the multidimensional observations for each voivodship by their ranks induced by multivariate depth (in the example we use the projection depth [21] calculated with the algorithm implemented in DepthProc R package [15].

In the second step, we calculate functional depth for S^v, $v = 1, ..., 16$. We use modified band depth, Fraiman–Muniz depth and extremal depth for the purpose. As we consider a centrality-oriented causality scheme, we divide voivodships into two groups (subsets of indices F, C) according to higher $(v \in F)$ and lower $(v \in C)$ value of calculated functional depth. In our simulation study regarding properties of the extremal depth, we use a depth value of $\alpha = 0.5$ to separate sets C and F. In the third step, we calculate the multivariate Wilcoxon sum rank statistic for the two samples $A^v, v \in F$ or $v \in C$.

Note that for calculating functional empirical depth we need a sufficiently larger number of observations per unit (voivodship) than the number of units (voivodships). This requirement is sometimes difficult to fulfil, so we need to apply a certain "replication of data" strategy. The application of the replication strategy is undesirable, but it is necessary, as without the strategy in our empirical example we cannot use the extremal depth, which possesses relevant statistical properties concerning coverage properties of depth-induced central regions.

Since in the considered empirical example increasing linear trends for each voivod-ship are quite evident, we decide to substitute the original dataset by simulated artificial dataset consisting of significantly more observations per voivodship than the original data by generating observations from a mechanism, which does not change the qualita-tive properties of the whole sample. For this reason, in the empirical example we use a simple linear model with parameters estimated with deepest regression method [29] and error characteristics estimated from original data with the assumption of normality, namely replicated data is generated in the following manner:

$$y_{rep} = \beta_0 + \beta_1 t + \varepsilon \qquad (3)$$

where t is drawn from a uniform distribution on the time interval [2014, 2018], and ε from a zero-mean Gaussian distribution with standard deviation equal to empirical standard de-viation of the original data. The parameters β_0 and β_1 are estimated with the deepest re-gression method (for details see [29] and computed with deepReg2d function of DepthProc R package [15]). Several parameters estimating methods including least squares, polyno-mial and constrained polynomial regressions are considered as well.

For instance, Fig. 1 presents an original sequence of consolidated subsidies in the period 2012–2018 (data for 2015 is missing in Statistics Poland database) for Opolskie voivodship (left panel) and a simulated sequence of 500 observations (right panel) from a model imitating the left panel.
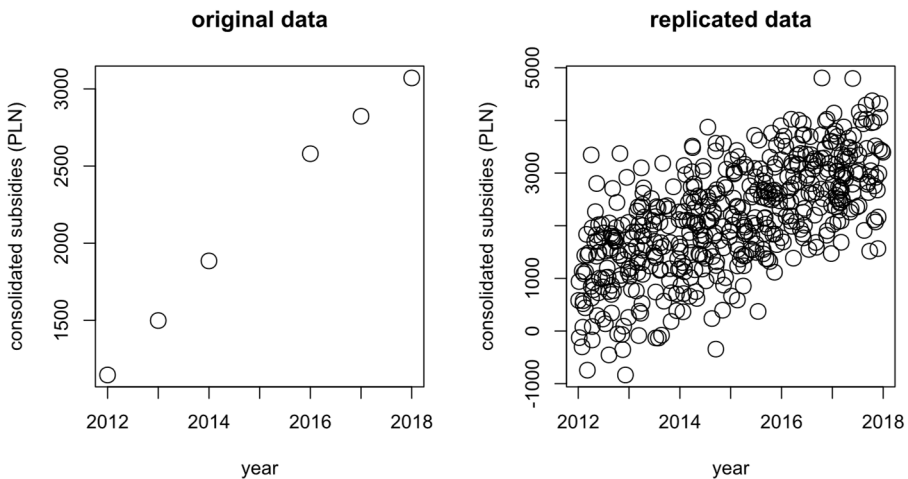

Fig. 1. Original and replicated data for Opolskie voivodship

To consider the treatment as a cause of the aggregate A we have to show that a con-ditional distribution of A under condition of the factual subsidies differs from a condi-tional distribution of A under certain alternative treatment plan, an alternative sequence

of subsidies. A fundamental difficulty here is that we have only access to the database provided by the Statistics Poland and we do not observe counterfactual distribution, i.e., distribution of A under different sequence of subsidies. To overcome this difficulty and taking into account a postulate stating that causal inference should possess a certain stability property, we propose to compare the distribution of the aggregate A estimated as based on voivodships with indices $v \in F$, which are central (i.e., typical) in terms of subsidies trajectory with a distribution of the aggregate A estimated as based on voivodships with indices $v \in C$, which are peripheral in terms of subsidies trajectory. The first distribution is treated as factual, whereas the second distribution as counterfactual.

Note that we have a functional dataset as an input set and a multivariate dataset as an output set. We must adapt a causal inference scheme to these datasets. The proposed approach is arguable but note that we assume that there exists a feasible digital development path for the considered homogeneous units, i.e., the regional digital development path can be achieved in the real world. Moreover, it is virtually impossible to repeat the experiment. That is why we treat typical or central observations as factual and atypical observations as counterfactual. In other words, we compare the digital development aggregate A conditioned on a treatment identified with subsidies trajectories with a high degree of centrality with aggregate A conditioned on subsidies trajectories with a low degree of centrality. In practice, first, we divide voivodships concerning specified levels of centrality, i.e., probability coverage, and then we estimate appropriate conditional distributions of aggregate A using the data replication scheme if it is required.

As we use random data replication procedure, we propose to repeat the second step 1000 times for 1000 artificially replicated datasets and infer on differences between factual and counterfactual distributions as based on estimated distribution of multivariate Wilcoxon sum rank test applied to groups of objects indicated in the second step of the procedure [17]. For measuring the centrality of the subsidies trajectory, we use a modified band depth (MBD) [23], Fraiman–Muniz depth (FM) [9] and extremal depth (ED) [25].

In the considered empirical example, we repeated 100 times a whole sequence of the second and the third steps repeated 1000 times and we obtained averages of 1000 average values of the multivariate Wilcoxon sum rank statistic equal to 37.72 (3.46) when ED was used, 33.23 (0.084) when FM was used and 24.212 (0.0734) when MBD was used; in brackets, standard deviations of 100 means are given (see Table 1, the first row).

Additionally, to strengthen conclusions drawn from our causal inference procedure, we repeated 1000 times an experiment, in which we compare two independent random samples A^j and A^k, where j is the same as a number of observations in F, and k is the same as a number of observations in C, and we obtain an average value of the multivariate Wilcoxon sum rank statistic equal to 32.281 (5.960) when ED is used, 31.761 (6.69) when FM is used, and 29.2692 (6.49) when MBD is used (see Table 1, the second row).

A difference between the average Wilcoxon sum rank statistic for samples representing factual and counterfactual distributions chosen via the proposed functional outlyingness criterion and those chosen at random together with significantly smaller standard deviations in the first case justify the validity of the proposal. Moreover, MBD behaves better than ED and FM in this sense that it enables better discrimination between factual and counterfactual distribution, and hence MBD gives the strongest arguments that EU subsidies influence the digital development of Poland.

Table 1. Average values of the multivariate Wilcoxon sum rank statistic for factuals and counterfactuals defined by outlyingness and defined completely at random

Type of samples	MBD	FM	ED
Sample defined by outlyingness	24.212 (0.0734)	33.23 (0.084)	37.72 (3.46)
Sample defined at random	29.2692 (6.49)	31.761 (6.69)	32.281 (5.960)

In Table 2, a comparison of depth values for 16 Polish voivodships is presented. No replication strategy is applied in MBD and FM case, so only the original data from Statistics Poland is used. As previously indicated and implicitly stated in [25], in ED case we had to replicate the original data. In Table 2 we see that six voivodships (denoted with the superscript f) are assigned typical observations for each of the functional depths, while six other voivodships (denoted with the superscript c) are always assigned atypical observations. For each of the three considered functional depths, eight voivodships are classified as typical representing the factual, while other eight are classified as atypical representing the counterfactual.

Table 2. Comparison of depth values for 16 Polish voivodships

Voivodship	MBD	FM	ED
Dolnośląskie	0.425	0.5249	0.8125
Kujawsko-Pomorskie	0.5	0.4986	0.375
Lubelskie[c]	0.341(6)	0.2937	0.25
Lubuskie[f]	0.591(6)	0.9702	0.625
Łódzkie[f]	0.575	0.9320	0.875
Małopolskie[c]	0.241(6)	0.2967	0.3125
Mazowieckie[f]	0.541(6)	0.7181	1
Opolskie[c]	0.475	0.4971	0.5
Podkarpackie	0.341(6)	0.4179	0.6875
Podlaskie[c]	0.125	0.1844	0.125
Pomorskie[f]	0.491(6)	0.5410	0.9375
Śląskie[c]	0.125	0.2458	0.0625
Świętokrzyskie[f]	0.591(6)	0.9702	0.75
Warmińsko-Mazurskie[c]	0.241(6)	0.2796	0.1875
Wielkopolskie[f]	0.48(3)	0.4994	0.5625
Zachodniopomorskie	0.575	0.5480	0.4375

We considered data replication algorithms including several parameters estimating methods, i.e., the deepest regression method, least squares, polynomial, and constrained polynomial regressions.

Although the maximal depth object changed from repetition to repetition, allocations of voivodships within groups F and C exhibited a high level of stability. We conclude therefore that agricultural subsidies may be treated as one of the reasons of digital development in the regions (voivodships) of Poland.

4. Summary and conclusions

Causal inference is commonly treated as the essence of a scientific comprehension of empirical reality. This kind of reasoning has many variants supported by different schools of economic thought. Although the GNC concept seems to be the most popular in economics, we have focused our attention on less popular in economics but very prominent in other sciences concept proposed by Donald Rubin. We have modified the concept to obtain a centrality-oriented causality reasoning scheme. As in economics it is difficult to conduct a truly randomised experiment postulated by Rubin's theory, we have proposed a certain kind of implementation of his theory. A "trick" based on the application of functional data depth which is possible to apply in practice has been used. In other words, we have proposed a novel DDC-based method of indicating factual and counterfactual distributions in causal inference scheme of Rubin. This seemingly quite simple "trick" is of prime importance in the context of difficulties we face in a real empirical causal analysis based on official statistics.

We have applied the centrality-oriented causality scheme to study the impact of agricultural subsidies, which may be treated as an indirect intervention or treatment, on a degree of digital development in the regions of Poland in the period 2012–2018. We have obtained arguments for the existence of causal relation between these economic phenomena. The proposed scheme may be easily generalised and adjusted to other studies. Having richer datasets, one can obtain more evident conclusions.

Thus, we have shown that the DDC offers valuable possibilities for conducting robust causal inference in economics, especially in multivariate and functional cases. The DDC is not a remedy for solving fundamental issues of causal inference of how to perform causal inference based on already existing datasets. To treat procedures of the DDC as real alternatives to classical multivariate of functional methods, further theoretical studies on sample properties of the procedures are required.

As a result of our studies, we conclude that MBD and FM functional depths are more useful than ED in the context of analysing relatively small and sparse empirical

datasets. As ED has other desired statistical properties, further studies on their modifications for small datasets are required. Original dataset and simple R code used in the paper are available upon request.

Acknowledgements

References

[1] BONGIORNO E.G., GOIA A., *Classification methods for Hilbert data based on surrogate density*, Comp. Stat. Data Anal., 2016, 99, 204–222.

[2] BONGIORNO E.G., GOIA A., *Describing the concentration of income populations by functional principal component analysis on Lorenz curves*, J. Mult. Anal., 2019, 170, 10–24.

[3] BOSQ D., *Linear Processes in Function Spaces*, Springer-Verlag, New York 2000.

[4] COX D.R., WERMUTH N., *Causality: a statistical view*, Int. Stat. Rev., 2004, 72 (3), 285–305.

[5] DAWID A.P., *Causal inference without counterfactuals*, J. Am. Stat. Assoc., 2000, 95 (450), 407–424.

[6] DYCKERHOFF R., MOZHAROVSKYI P., *Exact computation of the halfspace depth*, Comp. Stat. Data Anal., 2016, 98, 19–30.

[7] ENGLE R.F., WHITE S., *Cointegration, causality, and forecasting. A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press, Oxford 1999.

[8] FLEURBAEY M., MANIQUET F., *A Theory of Fairness and Social Welfare*, Cambridge University Press, Cambridge 2011.

[9] FRAIMAN R., MUNIZ G., *Trimmed means for functional data*, Test, 2001, 10 (2), 419–440.

[10] GIJBELS I., NAGY S., *On a general definition of depth for functional data*, Stat. Sci., 2017, 32 (4), 630–639.

[11] GILL R.D., ROBINS J.M., *Causal inference for complex longitudinal data. The continuous case*, Ann. Stat., 2001, 29 (6), 1785–1811.

[12] HENDRY D.F., *Granger causality*, Eur. J. Pure Appl. Math., 2017, 10 (1), 12–29.

[13] JURECKOVA J., KALINA J., *Nonparametric multivariate rank tests and their unbiasedness*, Bernoulli, 2012, 18 (1), 229–251.

[14] KLEINBERG S., *Causality, Probability, and Time*, Cambridge University Press, New York 2013.

[15] KOSIOROWSKI D., ZAWADZKI Z., *DepthProc: An R package for robust exploration of multidimensional economic phenomena*, J. Stat. Soft. (forthcoming), 2020.

[16] KOSIOROWSKI D., RYDLEWSKI J.P., ZAWADZKI Z., *Functional outliers detection by the example of air quality monitoring*, Stat. Rev., 2018, 65 (1), 81–98.

[17] KOSIOROWSKI D., RYDLEWSKI J.P., SNARSKA M., *Detecting a structural change in functional time series using local Wilcoxon statistic*, Stat. Papers, 2019, 60, 1677–1698.

[18] KOSIOROWSKI D., MIELCZAREK D., RYDLEWSKI J.P., *A critical study of usefulness of selected functional classifiers in economics*, Acta Univ. Lodz.. Folia Oecon., 2020, 2 (347), 71–90.

[19] LIU R.Y., SINGH K., *A quality index based on data depth and multivariate rank tests*, J. Am. Stat. Assoc., 1995, 88, 252–260.

[20] LIU R.Y., PARELIUS J.M., SINGH K., *Multivariate analysis by data depth. Descriptive statistics, graphics and inference (with discussion)*, Ann. Stat., 1999, 27 (3), 783–858.

[21] LIU X., ZUO Y., *Comppd: A Matlab package for computing projection depth*, J. Stat. Soft., 2015, 65 (2), 1–21.

[22] LIU X., ZUO Y., WANG Z., *Exactly computing bivariate projection depth contours and median*, Comp. Stat. Data Anal., 2013, 60, 1–11.

[23] LOPEZ-PINTADO S., ROMO J., *On the concept of depth for functional data*, J. Am. Stat. Assoc., 2009, 104 (486), 718–734.

[24] NAGY S., GIJBELS I., HLUBINKA D., *Depth-based recognition of shape outlying functions*, J. Comp. Graph. Stat., 2017, DOI: 10.1080/10618600.2017.1336445.

[25] NARISETTY N.N., NAIR V.N., *Extremal depth for functional data and applications*, J. Am. Stat. Assoc., 2016, 111 (516), 1705–1714.

[26] PAINDAVAINE D., VAN BEVER G., *From depth to local depth: a focus on centrality*, J. Am. Stat. Assoc., 2013, 105, 1105–1119.

[27] PEARL J., *Causality-Models, Reasoning, and Inference*, Cambridge University Press, Cambridge 2000.

[28] ROSENBAUM P.R., RUBIN D.B., *The central role of the propensity score in observational studies for causal effects*, Biometrika, 1983, 70 (1), 41–55.

[29] ROUSSEEUW P.J., HUBERT M., *Regression depth*, J. Am. Stat. Assoc., 1999, 94 (446), 388–433.

[30] RUBIN D.B., *Estimating causal effects of treatment in randomized and nonrandomized studies*, J. Edu. Psych., 1974, 66, 688–701.

[31] RUBIN D.B., *Causal inference using potential outcomes. Design, modeling, decisions*, J. Am. Stat. Assoc., 2005, 100 (469), 322–331.

[32] *Statistics Poland*, https://stat.gov.pl/, 2019, URL (access date: July 8, 2019).

[33] WILCOX R., *Introduction to Robust Estimation and Hypothesis Testing*, Academic Press, 2014.

[34] ZUO Y., *Projection based depth functions and associated medians*, Ann. Stat., 2003, 31 (5), 1460–1490.

[35] ZUO Y., SERFLING R., *General notions of statistical depth function*, Ann. Stat., 2000, 28, 461–482.