ISSN 2045-3345

RESEARCH JOURNAL OF ECONOMICS, BUSINESS AND ICT

COUNTERFACTUAL IMPACT EVALUATION OF HUMAN RESOURCES DEVELOPMENT

Oto Potluka[1], Jan Brůha[2]

[1]University of Economics in Prague, [2]The Institute for Structural Policy - IREAS; Czech Republic

ABSTRACT

The methods of Counterfactual Impact Evaluation (CIE) have not been extensively applied in the case of the Structural Funds assistance in the EU.

The main research question is whether the Counterfactual Impact Evaluation methods are applicable in the case of OP HRE (Operational program Human Resources and Employment), area of support 1.1 in the Czech Republic and what requirements have to be met. In case of this intervention, the "treatment" investigated by this paper is the enterprise support to training of employees.

The crucial part of conducting Counterfactual Impact Evaluations is the quality of data set. The OP HRE was tested for the compliance with the data requirements (sample size, randomization for approval of the assistance, homogeneity of the assistance).

The area of support 1.1 in the OP HRE met the basic requirement for the Counterfactual Impact Evaluation. There is enough assisted firms and the assistance is homogenous (i.e. there are trainings in just a few type of educational themes). The OP HRE support 1.1 offers possibility for the Counterfactual Impact Evaluation.

JEL CLASSIFICATION & KEYWORDS

■ C31 ■ C33 ■ D61 ■ COUNTERFACTUAL ■ IMPACT EVALUATION ■ STRUCTURAL FUNDS

INTRODUCTION

There has started few years ago a discussion concerning use of econometric methods for evaluation of impacts of the EU Cohesion Policy assistance. Efficiency of public expenditures is stressed as one of the main topic in economic policy during the time of economic crisis. The Counterfactual Impact Evaluation becomes an important tool for evaluation of the effectiveness. Massive subsidizing by EU funds shows the lack of methodology in the field of evaluation of such interventions in the whole European Union. Until now, the conducted evaluations have so-far utilized mainly qualitative methods, which cannot answer some important questions.

The quantitative evaluation methods are the reaction to pitfalls of qualitative methods - mainly the Counterfactual Impact Evaluation (CIE) methods. The combination of both qualitative and quantitative methods can answer a combination of questions necessary for successful implementation of expenditure programmes. Those questions are what is the impact, of what on whom and why. The combination of the qualitative with quantitative methods is necessary (Mouqué 2011).

The main research question of this paper is whether the Counterfactual Impact Evaluation methods are applicable in the case of the Czech Operational Programme Human Resources and Employment (OP HRE), area of support 1.1 "Adaptability". There are also additional questions to the main question: what requirements have to be met? The second question is: what are the preliminary results of testing of the Counterfactual Impact Evaluation methods. To answer these questions, we inquire whether it is feasible to perform a research of the impact of OP HRE on firms (especially on turnover, employment and profitability).

The paper is organized as follows. First, the CIE methods and data requirements are introduced. The description of OP HRE follows. Then, the appropriate CIE methods and their application to the OP HRE are described together with tests on turnover of firms. The last part concludes.

Counterfactual Impact Evaluation Methods

Generally, the ideal Counterfactual Impact Evaluation of public support programmes for firms would be based on a comparison of treated observations with a counter-factual situation of non-treaded observation of the same firms at the same time. Obviously, this is not possible.

This methodological obstacle can be overcome by using of two different groups of firms at the same time. The first one is composed of firms with the assistance; the second one is a control group. Both groups must have similar characteristics. With a large number of observations in each group, the mean differences of characteristics (i.e. the mean difference) are small.

This approach enables to – under some assumptions and a careful treatment – consider those two groups as almost identical and to use them as the ideal situation described above. The way of how this is exactly done is specific to a particular Counterfactual Impact Evaluation method.

Criteria for Counterfactual Impact Evaluation

Applicability of CIE requires meeting two basic requirements:

- A large number of observations. Ensuring large number of cases leads to working mainly at the level of individuals or organizations, but other approaches are also available (regions). Meeting of this requirement ensures statistical significance of estimates and the contribution of the reliability of results. But it is not a strict condition. There might be significant estimates of the impact even with a small sample size (up to hundreds). It depends on how the support works on firms. For the discussion see White (2011).
- Homogeneous cases. In order to use the Counterfactual Impact Evaluation methods, it is necessary to use cases representing the same or very similar situation.

Data from OP HRE for Counterfactual Impact Evaluation

The OP HRE, support area 1.1 "Adaptability" aims at preventing unemployment. It is done by investment into human capital. The main part of the support is a subsidy to training of employees in firms and to creation and implementation of business systems of human resources development (flexible forms of employment; systems of further development of employees´ skills and knowledge;

appraisal and development of human resources in firms) (MoLSA 2007). The projects with the trainings aimed at improving skills and knowledge of firms´ employees as key activities have been used for the purpose of the evaluation in question.

There are some applicants and beneficiaries who do not satisfy the requirement of homogeneity of the assistance. Those are associations, training agencies and ministries as the types of applicants in calls in the OP HRE, support area 1.1. Those beneficiaries will not used in our study as the support goes to firms through different ways in comparison with the situation when a firm is a final beneficiary. Moreover, those organizations will not probably participate on data collection.

Three main types of calls have been identified. The first one, grant calls, is suitable for Counterfactual Impact Evaluation. So are some of the system projects as the second type. The third type, are calls not suitable for the purpose of the evaluation. Those are calls oriented to other fields of eligible activities than trainings of private companies´ employees.

The first suitable data set represents grant calls. Grant calls are calls with private institutions as final beneficiaries. There are 1481 firms supported by the ESF in this data sample[1]. The amount of the assistance varies from 1 million to 10 millions of CZK. There are 1663 supported applications and 1907 rejected applications as some firms applied more than once.

The second suitable data set is a system project from the call Nr. 34 "Vzdělávejte se!" In this call, firms are in a position of target groups. They do apply just for funding of a particular training, not for the management of the supported project. There is open also new call Nr. 71 "Vzdělávejte se pro růst". The call Nr. 71 is quite new and has not been yet included into the data sample. There are 3357 firms supported by ESF in this part of the data sample (Nr. 34). The amount of the assistance varies from few hundred CZK to 4 million CZK. The mean of the assistance is 65 267 CZK in the case of this type of support.

There are suitable two sets of data for CIE according to the above-mentioned information. The first one is from grant calls. The second one is based on system projects. The main difference between those two data sets is the existence of points from the appraisal process in the case of grant calls. It enables to test regression discontinuity design and instrumental variables on this data set. We will test propensity score matching for both data sets.

Proposed methodological application of Counterfactual Impact Evaluation

The following text describes the methods which the authors would use for CIE in the case of OP HRE.

For discussion of CIE in OP HRE, we start with the characterization of the distribution of the application and actually obtained support (its size and type) by firms depending on their location, sector (NACE[2]) and other characteristics (i.e., size). This question could help us in constructing models below. From the econometric point of view, the distribution of the support size will be modelled non-parametrically (a non-parametric density function estimation based on the kernel estimator). To do that, we propose a non-parametric spatial model, where the number or ratio of firms supported in the region will be explained by the geographic location and firm characteristics (size, NACE, legal form). We suggest using radial basis functions (Buhmann 2003), which is a non-parametric method. The strength of suggested method is that the geographic location of firms may not be modelled using variables 0-1 (depending on the county or district where the firm is located), but it can be understood as a continuous variable (longitude and latitude location of the company). This method allows verifying whether it is possible to model geographic location e.g. depending on the distance from the nearest economic centre, from a major road or from the border (for exporting firms). Non-parametric method has the advantage that it identifies whether there is some easy-to-grasp-regularity in the location of economic activity (or its type). If so, it will be used for answering other types of research in CIE (especially the application of the method propensity score matching).

Datasets requirements:

(I) a list or a random selection of applicants for support from the relevant calls along with a description of basic characteristics (NACE, location, type, and basic economic data, an indicator of whether or not awarded aid, if so how much). Characteristics based on the available resources (no questionnaire, here), what could limit us, but hopefully not entirely.

(II) Random selection (probably according to the Identification number)

Then, we will proceed to questions comparing successful applicants. Here, we suggest using three independent methods. In case the results are similar, we can be sure about their robustness.

First, we apply the regression-discontinuity approach. We examine whether we should use the sharp or the fuzzy variant of the model. This will be decided according to the type of choosing the supported companies.

First, it is necessary to treat the situation that the evaluation of proposals was done in several rounds, which may cause the following two problems:

1. generally - different amount of points was necessary for support;
2. Various macroeconomic environment (support in the time of economic boom may have a different impact than support during the recession).

We solve the first problem by considering the deviation from the border of gaining support relatively to the round. The

[1] The data relates to 1th of July 2011; Number of realized projects includes the following categories: Project recommended / approved; Project with a decision; Project in realization; Realization finished. Thus there are projects with different phase of realization. Other projects are classified as not-realized.

[2] There are different approaches how to grasp statistically the industry sector of company. One of them is the introduction of dummy variables for individual sectors. An alternative method is not to use dummy variables, but to use variables describing the characteristics typical for firms in the industry (export status, size, etc.) in the regression. These properties could be derived either from the microdata, or some from the national accounts (e.g., share of exports in value added in the sector). The advantage of the second approach is that results can be more informative: it is easier to interpret the coefficients on observed characteristics than dummy variables. On the other hand, this approach exposes a greater risk than of incorrect specification than dummy variables. Alternatively, you can combine both approaches: the first stage to use dummy variables in the regression itself, and then in the second stage to explain the coefficients of dummy variables with the above-mentioned characteristics of sectors. This approach allows avoiding the wrong specification in the first instance, and - if the second phase of the model is well specified - allows a better interpretation of results. At the same time any incorrect specification of the second phase will not affect the results of the first phase. A similar approach was used in the model of travel expenses, see Murdock (2006).

second problem will be solved through the inclusion of time dummy among regressors (see below). An alternative is to use an index of economic performance for the relevant industry sector.

We begin with the local-linear model where we will control for important firm characteristics (such as sector, geographic location, the size and time when the support was obtained).

By using a local linear model with extra regression variables (business characteristics or time dummy or cycle indexes) we can address the problem of heterogeneity of support impact. A care must be made to deal with heterogeneity of the type of education. One – easier – possibility is to include the type of education among the control regressors. Second possibility (tougher on data and statistic) is to consider the multiple treatment variant (Papay, Willet and Mumane 2011). We pursue this alternative, only if it is possible to meaningfully categorize the support obtained from ESF OP HRE.

We plan to use Imbens and Lemieux (2007) to make a set of robustness tests (especially the length of the window selection for local linear model, testing whether there is indeed a regression discontinuity, the use of alternative methods - sieve estimator – instead of local linear regression, etc).

As demands for data, we need information about successful and unsuccessful applicants - scoring obtained in the process and in the evaluation round (especially how many points were needed in the round). Most data about the companies are available in public resources. Probably, the most serious problem is the data on employment, which will probably need a survey.

Then, we attempt at using the instrumental variable estimator. As the instrument, we plan to use the identification of appraisal experts. This may be a valid instrument, because the appraisal expert probably determines the outcome (kind/strict), but presumably does not influence the outcome. In case appraisal experts do not vary enough (equally strict), this method would be unusable.

We start with a linear instrumental variable model, where we explain the percentage change in the indicator (profit, sales, employment) as a function of control variables (size, location, and sector of firms) instrumenting by the appraisal experts. Controlling for these regressors will enable us to address some questions concerning turnover, employment and profitability. Again by adding extra regressors it is possible to address the problem of impact heterogeneity on different companies.

As an alternative, we will consider a semi-nonparametric estimator[3]. This would make the results robust to the functional form, but also enables keeping the modelling of results based on the characteristics of firms.

Data requirements are similar to previous methods, but we also have to know the appraisal experts of individual projects (their identifiers) and how many points did appraisal expert give to the particular project.

In the case of using this method we discuss in detail the conditions of its application. Applicability may be at risk if:

1. All the appraisal experts evaluated very similarly (the instrument is too weak, standard error of estimates are large and the results unreliable);
2. or if a significant part of the appraisal experts skewed to certain types of companies (e.g., geographically or professionally).

As a third method we plan to use the propensity score matching. Among variables entering the first stage (the discrete choice model), we consider the following ones: economic sector, regions, size, employment of women, disabled or minorities, and economic outcomes before the support (both levels and growth rates). We use bayesian methods (O´Hara and Sillanpa 2009) for variable selection. We need a sample of both supported and not supported companies. The propensity score matching is also the only way applicable to addressing some questions comparing treated and non-treated firms. To do it, we need a random sample of companies, which did not ask for the OP HRE support (a control group). The PSM method will be combined with the method of conditional difference in difference, which is the standard approach, allowing isolation of the effect of observed characteristics of firms to the outcome.

Here we need similar data as above, in the case of non-applicants to construct "random" sample. Similarly as above, we need to get some data through a questionnaire survey (especially about employment).

To summarize, in all three methods described above, we first use simple models and then more difficult models. First we use linear models to explain the changes of indicator (employment, profits, or sales) using the observed characteristics, and we check deflection due to self-selection, which does each method in other way. In the next step we consider more advanced methods (mostly non-parametric), which can overcome the implicit limitations of linear relationship and thus make the results more robust. On the other hand, these methods are very demanding on the data (their amount).

Also - as mentioned in methods - we try to use those variants of methods that allow the involvement of additional regressors. It can both increase the efficiency of statistical estimates and allow us to have a better idea of which group of companies the support works and how much.

However, the heterogeneity of the impact of support can be compared from multiple perspectives. As already have been described, it is a comparison of the average impact of the support on different groups of companies. We plan also to characterize the distribution of the impact of the support (unconditional or conditional on observed characteristics). Indeed, the average support may not provide the complete picture, since some the impact of types of support or the impact on different businesses can be considerable variable.

The characteristics of this distribution can be addressed differently in different methods. In the parametric method of instrumental variables it is - under certain circumstances – possible to use quantile regression methods (see Chernozhukov and Hansen 2005 and Torgovitsky 2010 for more details).

Results and discussion

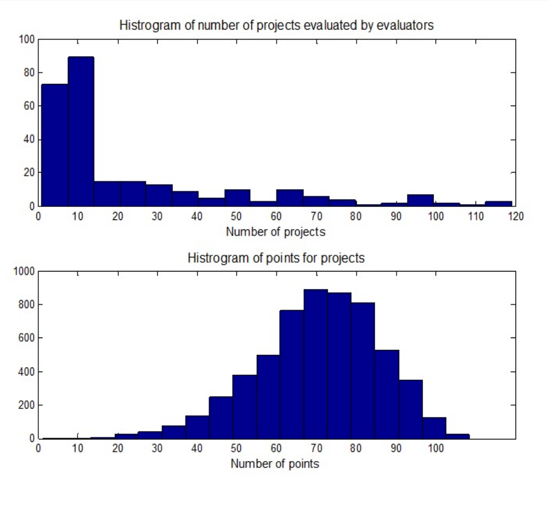
Before we proceed to the analysis, we describe the data on which a given hypothesis is examined. Especially, we inquire whether the identificators of appraisal experts can be used as a valid instrument in the instrumental variable regression.

Monit7+ provides data on applications appraised by 267 appraisal experts. In some cases, we suspected duplicate cases (two different appraisal experts having the same name and surname), or when the same name appears once with the academic title and one without it. After excluding uncertain cases, we end-up with 220 appraisal experts.

[3] See Chen 2007 or Blundell 2007 for an introduction to semi-parametric instrumental variable techniques

About 120 appraisal experts rated more than 10 projects and 100 appraisal experts rated at least 15 projects. The graph 1 shows a histogram of the number of projects that appraised individual appraisal experts and the histogram of number of points awarded by each appraisal expert.

Figure 1: Frequencies of appraisal and appraisal experts

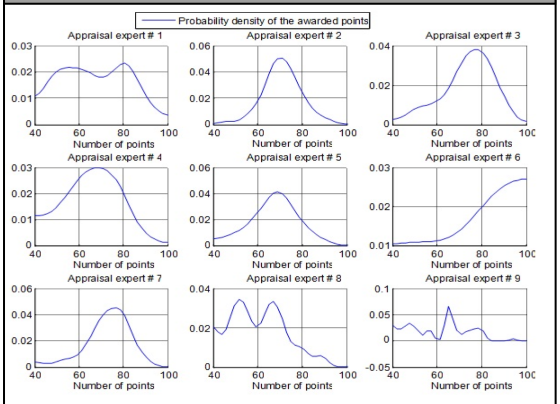


Source: Monit7+ and own calculations

In addition, we examine the individual characteristics of individual appraisal experts. Since placing the names of appraisal experts can be sensitive, we present the numerical codes of appraisal experts. The appraisal experts were ranked according to number of appraised projects, so the appraisal expert 1 is the appraisal expert, who appraised the highest number of applications. In the case that two appraisal experts appraised the same number of projects, they are randomly sorted.

Graph 2 shows the nonparametric kernel estimate of the probability density of points given by nine appraisal experts with the largest number of appraised projects. The results indicate that the probability density varies a lot among appraisal experts.

Figure 2: Probability density of the awarded points for appraisal experts



Source: Monit7+ and own calculations

Some appraisal experts gave scores symmetrically around certain values (e. g, No. 2, the appraisal expert gives scores of 70 plus / minus 10, similarly appraisal expert No. 5), others awarded points more or less evenly (e. g, appraisal expert No. 9), and some tend to give high values approaching 100 points (e. g, appraisal expert No. 6).

Furthermore, we compare the distribution of selected statistical indicators of selection for 150 appraisal experts who appraised more than 10 projects. These sample characteristics include mean, median, a robust mean (based on a deletion of 20% of the outliers), standard deviation (STD), margin and interquartile range (IQR).

Some appraisal experts gave an average of 55 points and another 85 points. Typical appraisal expert gave between 65 and 70 points. We obtain similar results, if we look at the median or trimmed mean. There are differences between appraisal experts not only in the mean (or median) of number of points, but also in the variability. There are appraisal experts who appraise projects almost all the same way (low standard deviation, low margin - close to zero) and other appraisal experts have their appraisal with high variability (range near 100 points, interquartile range is higher than 25 points, a high standard deviation).

Comparison of these characteristics indicates a significant heterogeneity of individual appraisal experts and thus the possibility to use the number of points awarded by the application by the particular appraisal expert as an instrument. However, a rigorous assessment that is not enough for two reasons.

1. Are these results really significant?
2. It is possible that differences in appraisal were caused by the situation that different appraisal experts appraised systematically different projects?

Answers to these questions were found by using the regression analysis. We construct a regression model, where the dependent (explained) variable is the number of points awarded by individual application and individual call for proposals and regressors (explanatory variables) are both persons of the appraisal experts and the observed characteristics of firms (especially considering the following dummy variables - region, company size and legal form and NACE).

There is a large number of possible regressions. We proceeded by testing a general-to-specific (i.e. first estimated the model including all possible variables, i.e. identifiers of appraisal experts and dummy variables for the region, company size and legal form). Then we gradually eliminated some variables. This approach is possible because the covariance matrix of regressors is almost diagonal (appraisal experts identifiers are mutually orthogonal, the other regressors are close to mutual orthogonality).

Interpretation of results is as follows (technical details are available upon request). Unconditional mean score is 73.64 and 99% confidence interval is narrow. The appraisal expert No. 1 gave an average of 4.36 points less than the other appraisal experts. Appraisal expert No. 6 gave an average of 6.21 points more in comparison with others. Small firms (with less than 50 employees) received an average of 2.36 points less than the unconditional average. Firms located in the Moravian region received about 2 points more than other companies. Stock companies received an average of 1.7 points less than the average. Using these regressors, we estimate the robustness due to a probit model where the probability of acceptance of the project explained above mentioned regressors[4]. Probit model is able to correctly

[4] Because that not all identifiers of appraisal experts proved significant, we use a probit model to test the selected identifiers for their sufficient predictive power.

classify 83.6% of cases, suggesting that the identifier of appraisal experts is significant predictor not only of the number of points, but also the success of the application.

The results therefore show that appraisal expert identifier satisfies a necessary condition for being used as an instrumental variable.

Then, we have applied the model of the regression discontinuity design and instrumental variable methods on turnover of firms, but with insignificant estimates. The results will be sent upon request. One of the probable reasons for insignificant results may be that the support will impact the economic outcome of firms in future and therefore we have to wait until the effect will appear in data.

Conclusion

The econometric analysis proved that the identifier of appraisal experts meets the necessary condition for being used as an instrumental variable. Therefore, this analysis will continue in the further research.

Existing estimates of impact of OP HRE, support area 1.1 do not show significant effects of aid on turnover of firms in this moment. The results are influenced by a short period of collecting data. This result is valid for all methods that were used in this phase of the solution.

According to the result of test of identifier of appraisal experts as an instrumental variable, Managing Authority of OP HRE should pay attention to the training of appraisal experts. The continuous work with the appraisal experts in the area of intervention 1.1 may ensure the harmonized approach to project appraisal. This can be done, for example, by appraising a single, artificially created, application by all appraisal experts who want to evaluate projects 1.1. Managing Authority should than provide feedback on the results for each appraisal expert. Similar activities have already begun to practice in the OP HRE in the areas of support 3.4 and 5.1.

The area of support 1.1 in the OP HRE met the basic requirement for CIE. There is enough assisted firms and the assistance is homogenous (i.e. there are trainings in just a few type of educational topics).

A combination of econometric methods (regression discontinuity design, instrumental variables and propensity score matching and difference-in-difference method) can be used in the OP HRE.

To answer the questions what is the impact, of what on whom and why, it is necessary to use both quantitative and qualitative methods. Both of them have some strengths and weaknesses.

References

BLUNDELL, R., CHEN, X., KRISTENSEN, D. (2007). “Semi-nonparametric IV estimation of shape-invariant Engel curves”. Econometrica.

BUHMANN, M., D. (2003), Radial Basis Functions: Theory and Implementations, Cambridge University Press.

CHEN, X. (2007): Large Sample Sieve Estimation of Semi-Non Parametric Models, Handbook of Econometrics, Volume 6B, chapter 76,p. 5550-5632.

CHERNOZHUKOV, V. and HANSEN, C. (2005): An IV Model of Quantile Treatment Effects. Econometrica, 73, 245-261

IMBENS, G., W., LEMIEUX, T., 2007. Regression discontinuity designs: A guide to practice. Journal of Econometrics, Elsevier 142, p. 615-635.

MINISTRY OF LABOUR AND SOCIAL AFFAIRS OF THE CZECH REPUBLIC (MoLSA) (2007): Human Resources and Employment Operational Programme 2007-2013, downloaded on 17th October 2011 at http://www.Strukturalni-fondy.cz /CMS Pages/ GetFile. aspx?guid=cc15a308 -13ee- 48ba- b678- 2664f6399fe2

MOUQUÉ, D. (2011): The best of both worlds: combining approaches, Open Days 2011, downloaded on 17th October 2011 at http://ec.europa.eu /regional_policy/ conferences/ od2011/ presentation_frame.cfm?nmenu=333

MURDOCK, J. (2006): Handling unobserved site characteristics in random utility models of recreation demand. Journal of Environmental Economics and Management, Volume 51, Issue 1, January 2006, Pages 1-25.

O'HARA, R. B.; SILLANPA, M. J. (2009): A Review of Bayesian Variable Selection Methods: What, How and Which, Bayesian Analysis 4(1), p. 85-118.

PAPAY, J. P., WILLETT, J. B.; MURNANE, R. J. (2011). Extending the regression-discontinuity approach to multiple assignment variables. Journal of Econometrics 161, p. 203-207.

TORGOVITSKY A. (2010) Identification and Estimation of Nonparametric Quantile Regressions with Endogeneity, mimeo Yale.

WHITE, H.; An introduction to the use of randomized control trials to evaluate development interventions; Working Paper Nr 9, 3ie, 2011