



Mutual Learning Programme

Host Country Discussion Paper - Czech Republic

Czech pilot CIE: Successful project with insignificant results?

Peer Review on “Counterfactual Impact Evaluation”

Czech Republic, 6-7 October 2014

This publication is commissioned by the European Community Programme for Employment and Social Solidarity (2007-2013).

This programme is implemented by the European Commission. It was established to financially support the implementation of the objectives of the European Union in the employment, social affairs and equal opportunities area, and thereby contribute to the achievement of the EU2020 goals in these fields.

The seven-year programme targets all stakeholders who can help shape the development of appropriate and effective employment and social legislation and policies, across the EU-28, EFTA-EEA and EU candidate and pre-candidate countries.

For more information see:

<http://ec.europa.eu/progress>

For more information on the Mutual Learning programme, please refer to:

<http://ec.europa.eu/social/mlp>

European Commission:

Unit C1 European Employment Strategy and Governance, DG Employment, Social Affairs and Inclusion

EMPL-C1-UNIT@ec.europa.eu

Author: *Klara Kaliskova*

In collaboration with ICF International

The information contained in this publication does not necessarily reflect the position or opinion of the European Commission

Table of Contents

1. Executive summary	1
2. Background	2
3. Approaches to CIE	3
4. Key challenges	9
5. Success factors and transferability	11
6. References	13

1. Executive summary

The pilot Counterfactual Impact Evaluation (CIE) of active labour market policies was conducted in the Czech Republic in 2011-2013. This pilot project aimed at assessing the impact of the ESF Operational Programme Human Resources and Employment, support area 1.1, on supported companies. The evaluators assessed the impact of two types of projects from this support area - grant projects and the system call 'Vzdělávejte se!' [Educate yourself!]. The impact on employment, competitiveness, investment in training, and staff fluctuations of supported firms were assessed. Moreover, the heterogeneity of ESF impact was analysed by firms' size, region, sector of industry, and type of realized training.

The approaches to CIE in this project included instrumental variables, regression discontinuity design, and difference-in-differences with matching. Probably the most innovative part of the project was the use of instrumental variables method. The evaluators took advantage of the fact that the grant project proposals had been evaluated by external appraisal experts, who determined if the proposal were to be considered for approval. Two experts were randomly allocated to each proposal, and it turned out that the identity of experts affected the probability of approval substantially. Therefore, the identity of appraisal experts was used as an instrument in the evaluation of the impact of grant projects on supported firms. The main idea of the method was to compare the successful and unsuccessful applicants, for whom the probability of receiving the support was affected by the identity of appraisal experts assigned to them.

The second estimation method used in the pilot CIE also took advantage of a specific aspect of the appraisal process for the grant projects. There was a threshold of 65 points (an average number of points from the appraisal experts) for grant proposals to be considered for approval. Simply, the proposals below this threshold were not considered for approval, which created a nice discontinuity in the probability of approval. The regression discontinuity design was used to compare the successful applicants just above the threshold with unsuccessful applicants just below the threshold based on the assumption that these applicants should be similar in all characteristics but the actual probability of approval.

Third, the difference-in-differences method uses two differences to calculate the impact of an intervention – the first difference is between supported and unsupported firms and the second over time (before and after the intervention). It is often combined with matching to find the most similar comparison group of the unsupported firms, which should ideally differ from the supported firms only as a result of an intervention. The advantage of this method was that it was applicable not only to the grant projects, but also to the system call 'Vzdělávejte se!', for which the other two methods were not applicable.

Results of this pilot CIE suggest that the ESF projects had no impact on most of the measured outcomes of supported companies. The evaluators did find a positive impact of the grant ESF projects on employment in large and medium-size companies. However, interpreting this as an evidence for a positive impact is questionable, because this was confirmed by only one of the three estimation methods used and finding an effect on one outcome, in one group of firms, by one method might be a mere coincidence (in statistics, this is called the type I. error). However, it is understandable that the evaluators wanted to highlight some positive effect of the intervention (if only for one project type and for a selected group of firms).

The most successful part of the project was probably the promotion of the CIE methods among the Czech policy makers and local experts. Moreover, the project prepared ground for the future CIE of the ESF projects and established a high-quality

methodology that can be applied to other programme evaluations in the Czech Republic, and with various adjustments also abroad.

2. Background

The need for rigorous quantitative evaluation of active labour market policies (ALMP) is stressed, for example, in the Barca Report (Barca, 2009) or in the Guidance document on Monitoring and Evaluation of European Cohesion Policy (European Commission, 2014). The importance of the effective use of resources grows. Nevertheless, evaluations of active labour market policy interventions are still conducted primarily on a qualitative level both in the Czech Republic and in many other Member States.

Methods of CIE provide a way to assess the net effect of an intervention quantitatively by comparing the difference between outcomes that would be achieved with and without an intervention. Although some Member States have already carried out these types of evaluations, CIE still presents a substantial challenge for many countries due to the lack of appropriate data, methodological or other issues.

There were several attempts to conduct quantitative evaluations of active labour market policies in the Czech Republic. Above all, Prof. Sirovátka from the Masaryk University in Brno lead or participated in several research projects that focused on the evaluation of active labour market policies at the individual-level. This research was mostly based on individual-level data from the Czech employment office register, and several recent studies used some of the CIE methods. In particular, they used a propensity score matching technique comparing the employment status of unemployed individuals after participating in an ALMP program with matched unemployed individuals who did not participate in such a program (see e.g. Sirovátka and Kulhavý, 2007; Hora et al., 2009; Hora and Sirovátka, 2012).

However, the Czech employment office database is not ideal for CIE for several reasons. First, the only information available to researchers is whether or not (and when) an individual left the labour office register. There is no information about the type of job or wage that an individual got when he left the register. Second, the database does not provide information about the type of ALMP program (e.g. training or employment subsidies) for programs financed from the ESF, and some ESF programs are not even present in the database (Hora and Sirovátka, 2012). Third, data from the employment office register are often incomplete or missing (for details, see Hora and Sirovátka, 2012, pp. 11-12).

The first large-scale CIE of active labour market policies was conducted in the Czech Republic in 2011-2013. This pilot project of CIE aimed at assessing the impact of ESF Human Resources and Employment Operational Programme (HREOP), support area 1.1, on companies. The support area 1.1 of HREOP focuses on investment into human capital and increasing the flexibility of the workforce, their knowledge and skills. The decision to do a pilot CIE on this support area of HREOP was based on the following arguments.

- This support area consumes a large portion of the HREOP budget.
- It seemed the most suitable for the CIE evaluation, because the support was provided to companies. Since the only available individual-level data (from the employment office register) are very limited in the potential for CIE (see above), this support area was ideal for evaluation, as the assessment could have been conducted at firm's level.

- Previous assessments of this support area indicated possible inefficient use of resources.¹

The main impetus for this pilot project came from within the Czech Ministry of Labour and Social Affairs (hereinafter referred to as 'the ministry'). Dr. Kváča, who initiated this project at the ministry, participated at a summer school in counterfactual impact evaluation by Prof. Alberto Martini, where he learned to use the CIE methods in evaluation of active labour market policies. After he returned from the summer school, he promoted the use of CIE for the evaluation of the ESF projects and created a call for proposals for the pilot CIE in the Czech Republic. The call specified in a detailed way the methods of evaluation that should be used including regression discontinuity design and difference-in-differences with propensity score matching. Moreover, the call included several qualification questions that were meant to ensure that the selected applicant would be able to conduct a high-quality CIE. The Institute for Structural Policy (IREAS)² submitted the best proposal and was selected to perform the evaluation (researchers from IREAS who performed the evaluation are hereinafter referred to as 'the evaluators').

3. Approaches to CIE

Here and below, I focus on the CIE that was conducted within the Czech pilot counterfactual impact evaluation project in 2011-2013. While there were other attempts to use CIE in active labour market policy assessment, this was the only project that combined several methods of CIE and provided a comprehensive analysis of the intervention.³ This is also the only CIE conducted on firm-level data.

3.1 Subject matter for study

This pilot project of CIE focused on assessing the Operational Programme Human Resources and Employment, support area 1.1. This programme is aimed at increasing the adaptability of employees and employers. It supports investments in human resources development, the development of the professional knowledge, skills and competences of employees and employers, and creation of systems that help to increase the flexibility of the workforce. The target group of this area of support is employers and their employees. It supports the investment into further education and training of employees, implementation of modern systems of human resources management, and company adaptation programmes that support the sustainability of jobs.⁴

The pilot counterfactual impact evaluation aimed at assessing the impact of this HRE operational programme on companies. The evaluators thus narrowed the focus of the project on the intervention areas for which the applicant and non-applicant firms can be distinguished. In particular, analysis of the impact of grant projects under calls no.

¹ Focused group discussions revealed that this operational programme often supports activities that would have been conducted in the absence of support as well (Final report of the Annual operational evaluation of HREOP 2009, p.47. Internet: <http://www.esfcr.cz/file/7969/>).

² IREAS is a non-profit organization, formed as a project based platform of academic and policy experts. For details, see their webpage: <http://www.ireas.cz/en/>

³ See above a description of some other attempts to use CIE on the Czech ALMP.

⁴ Source: HREOP 2007-2013, internet: <http://www.esfcr.eu/file/8570/>

23, 35, 39 and 60, and the system call no. 34 (Vzdělávejte se! [Educate yourself!]) was carried out within the evaluation.⁵

In the grant projects, firms themselves apply for the support and the grant proposals are evaluated by external appraisal experts. Each grant proposal is evaluated by two appraisal experts,⁶ and if the average number of points by the two experts is above 65 (out of 100), the proposal is then evaluated by the evaluation committee, which decides if the proposal obtains the support or not.⁷

In the system call 'Vzdělávejte se!', the support was not allocated directly to firms, but was distributed through local employment offices. Firms prepared project proposals that were evaluated by the employment offices, but the evaluation process was much more individualistic than in the grant proposals.⁸

3.2 Methods used

The evaluation aimed at assessing the net impact of these programmes on the supported firms. Ideally, comparison of the situation of supported companies after intervention with the situation of supported companies without any support should be carried out. However, the situation of supported firms without support is hypothetical. The support from programmes was not allocated randomly across firms; the situation of supported firms cannot be thus simply compared to the situation of unsupported firms. However, methods of counterfactual impact evaluation allow using the unsupported firms as approximation of the hypothetical situation of supported firms without support under specific conditions (for details on CIE, see e.g. Angrist and Pischke, 2008, or Imbens and Wooldridge, 2009). The main approaches to CIE in this project were:

- instrumental variables;
- regression discontinuity design;
- matching with difference-in-differences.

Instrumental variables (IV) method allows estimating a net impact of a programme if there is a variable that affects who receives an intervention, but does not affect the situation of supported and unsupported firms after the intervention. This variable is called an instrument. The identity of appraisal experts was used as an instrument to evaluate the impact of grant projects (calls no. 23, 35, 39 and 60) in this pilot CIE. The identity of appraisal experts affects allocation of support across firms (some appraisal experts are strict, some are more lenient), but does not correlate with firms' outcomes.

⁵ Assessment of the calls no. 2, 33, 46, 50, and 52 was not possible, because the applicants in these calls are not firms, but institutions providing education and training, or the calls are not even focused on increasing the human capital of workers.

⁶ If the two appraisal experts disagree in their evaluation of the proposal, the third expert can be invited to provide a third evaluation. The third expert was asked to provide an evaluation for over 30% of proposals (Pilot CIE, Initial Report, p. 9).

⁷ The actual threshold for obtaining supports then differs in different calls depending on the number and quality of proposals and the total budget available.

⁸ Some employment offices followed the recommended evaluation procedure (3E method) from the Ministry of Labour and Social Affairs, but other offices approved all proposals that were factually and formally correct (Pilot CIE, Initial Report, p. 11).

This approach provides a very good way to identify a causal impact of an intervention and has been used in other policy evaluations (see e.g. Dahl, Kostol and Mogstad, 2014). However, this approach works only under the assumption that the instrument is strong, i.e. appraisal experts differ systematically in their strictness/leniency and the identity of expert thus affects the probability of receiving support. This seems to be the case because the probability distribution of the number of points from the appraising process differs across appraisal experts (different experts allocate different average number of points, but also have different variance of the number of points allocated).⁹ However, as the investigators mention, the differences in the points distribution across experts does not necessarily imply that the identity of judge significantly predicts who obtains the support. To prove that the instrument is indeed strong, a regression model of probability of obtaining the support on the identity of an expert needs to be estimated. The evaluators indeed estimate such a model, but they do not report the estimation results.¹⁰

The second assumption of the instrumental variable approach requires that the identity of experts is uncorrelated with the firms' situation after an intervention (other than through the fact that experts influence whether or not a firm receives the intervention). This assumption is ultimately untestable, but should be supported by a reliable argument. In this CIE project, the evaluators argue that this assumption is reasonable because the appraisal experts are randomly allocated to proposals and do not come into contact with the appraised firms and therefore cannot influence their situation.

The second method used in this project was the **regression discontinuity design** (RDD). This method is applicable to the assessment of interventions that have a threshold for obtaining the support (for an application of RDD on program evaluation, see e.g. Blundell et al., 2004). The method is based on a comparison of supported firms that were slightly above the threshold with unsupported firms that were slightly below the threshold. The intuition behind this method is as follows – the firms that were slightly below threshold should be similar in observable and unobservable characteristics to the firms slightly above, and the allocation of support around the threshold is close to random. Therefore, we can use the situation of unsupported firms that were slightly below the threshold as an approximation of the 'unsupported' situation of supported firms that were slightly above the threshold. However, this method provides only local results informing us about the impact of support for those firms that are close to the threshold.

In the pilot CIE, this method was applied to the grant projects, where the grant proposals with more than 65 points were considered for approval by the evaluation committee, while proposals with less than 65 should not be considered. In reality, the threshold was not strictly followed by the evaluation committee.¹¹ The evaluators thus set the threshold of 65 points for calls no. 35 and 39, while for calls no. 23 and 60, the thresholds were set up for 73.4 and 60 points, respectively. The regression discontinuity estimation then requires a choice of the estimator for the relationship between the firm's outcome and the number of points. The evaluators chose to use

⁹ Pilot CIE, Progress Report No. 1, Appendix 1, p. 6.

¹⁰ The evaluators report only the effect of expert identity on the probability of obtaining the support for selected 9 experts who have the largest effect (Pilot CIE, Progress Report No. 1, Appendix 1, p. 15).

¹¹ Some proposals with less than 65 points were also approved. This was the case for proposals that were evaluated by three appraisal experts, where two of them gave high evaluation and the third one gave very low evaluation, which resulted in an average number of points below 65 (Pilot CIE, Progress Report No. 1, Appendix 2, p. 26).

nonparametric methods (Nadaraya-Watson estimator and local linear model), because the parametric methods proved to be much less robust (Pilot CIE, Final Report, Appendix 1, p. 23).¹²

The third method used in this project was the **matching with difference-in-differences** (DiD). This method uses two differences to calculate the impact of an intervention – the first difference is between supported and unsupported firms and the second over time (before and after the intervention). It is based on the assumption that the development of the two groups (supported and unsupported) over time would have been the same in the absence of an intervention. This assumption is untestable and might be difficult to satisfy for widely defined comparison groups (unsupported firms that serve as a comparison for supported firms). Therefore, the difference-in-differences is often combined with matching¹³ to find the most similar comparison group of unsupported firms to the supported firms, which should ideally differ only as a result of an intervention (for an application of matching in program evaluation, see e.g. Lechner, 2002). However, the disadvantage of the matching with difference-in-differences is that the internal validity (satisfaction of the assumptions) is probably the most questionable among the described methods.

The evaluators used the matching with difference-in-differences based on a propensity score, which estimates the probability of receiving the support based on observable characteristics and then matches the supported firms with unsupported firms that have the closest probability of receiving the support. The impact of the ESF support by the matching with difference-in-differences was evaluated by comparing the situation of successful applicants with unsuccessful applicants and also of successful applicants with non-applicants.

The advantage of this method was that it was applicable not only to the grant projects, but also to the system call 'Vzdělávejte se!', for which the other two methods were not. However, for this system call, only a comparison of successful applicants with non-applicants is possible, because this call basically did not have any unsuccessful applicants.

3.3 Results

The evaluators assessed the impact of the ESF support on the following firm-level indicators:

- employment (number of persons employed in the companies);
- competitiveness (profit and sales);
- investment in training;
- staff fluctuations.

Moreover, they evaluated how the impact differed by the size of firm (micro, small, medium, and large), regions (NUTS II), sectors of industry, forms of support (grant and system projects), and types of realized training.

Overall, the results are largely insignificant, in some cases even contradictory. The impact of system project 'Vzdělávejte se!' was largely insignificant for all methods, all

¹² Nonparametric methods are methods of estimation that do not rely on data belonging to any particular probability distribution. In that sense, nonparametric methods are much less restrictive with respect to the underlying assumptions.

¹³ Matching is a technique that assigns to each supported firm one or more unsupported firms (that have the most similar characteristics) so that the difference in outcomes of these "matched" supported and unsupported firms provides an estimate of the impact of the support.

outcome variables, and in all types of companies. This did not come as a big surprise, because this system project distributed relatively small amounts of money to firms. The evaluators also mentioned that the insignificant effects might be driven by the fact that the only method applicable for this system project was matching with difference-in-differences.

A summary of the main results for the grant projects follows.

- **Employment:** The evaluators found statistically significant impact of the grant projects on employment only in medium and large companies by the method of instrumental variables.¹⁴ The other two methods did not confirm the result, and no significant impact was found for the micro or small companies. There was no significant impact on employment when firms were divided by regions or sectors of industry.
- **Profit:** The impact on profit is insignificant for micro, small, and medium-size companies and provides contradictory results for large companies.¹⁵
- **Sales, investment in training, and staff fluctuations:** There was no significant impact of the ESF support on sales investment in training, or staff fluctuations (no matter what method was used or whether firms were divided by size, region, or sector).

Although the results are largely insignificant, the evaluators highlight the positive impact of the grant projects on employment in large and medium-size companies found by the IV method. Overall, the evaluators analysed the impact on 5 different outcome variables using 4 different methods (DiD were conducted twice with two different comparison groups) in 19 types of companies (4 types by size, 7 regions, and 8 sectors). The evaluators thus ran about 380 different specifications, out of which these two significant specifications are interpreted as an evidence for the presence of an impact. When the presence of an effect is tested in so many specifications, detection of a few significant effects can be a mere consequence of type I. error.¹⁶ Moreover, the effect was not confirmed by the other two methods and I would be thus cautious in interpreting this as an evidence for positive effect.

Further, the evaluators claim that there was a positive impact of grant projects on profit in large companies although the results by different methods are contradictory (two methods suggest negative effect, two suggest positive effect). The authors deal with this contradiction by stating that the instrumental variable method is the most

¹⁴ The results by the instrumental variable method suggest that the grant projects increase the employment on average by 15 jobs per company in large companies and by 8 jobs per company in medium-size companies. This implies 3 713 jobs created by the grant projects in total (Pilot CIE, Final Report, p. 11).

¹⁵ DiD method suggests significantly positive impact when supported and unsupported firms are compared and negative insignificant results when supported and non-applicants are compared. The regression discontinuity method suggests significantly negative impact, while the instrumental variable method implies positive and significant results (Pilot CIE, Final Report, p. 17).

¹⁶ In statistics, type I. error is an incorrect rejection of a true null hypothesis. When the coefficient is significant at 5%, which is the most often used threshold for significance, the probability of type I. error is 5% - in 5 out of 100 estimations, the estimator incorrectly rejects the null hypothesis of insignificant effect concluding that the effect is significant although it is not. The evaluators themselves mention the problem of the type I. error several times in the project reports, but do not consider its presence in this case.

reliable and it implies a positive impact (together with the difference-in-differences), so they prefer its results over the results of the regression discontinuity, which implies negative impact (Pilot CIE, Final Report, p. 15). This is a very questionable interpretation of results.

Lastly, the impact of ESF support by the type of realized training among supported firms was investigated. The estimation results suggest an above-average positive impact of training in computer skills, foreign languages, and technical skills. However, the interpretation of these findings crucially depends on the presence of a positive effect of training overall (otherwise, the above-average impact might still be a zero impact), which is questionable given the previous results.

The pilot CIE project also investigated the mechanism, through which the ESF support affects firms' situation. The evaluation team identified three possible mechanisms¹⁷ and shortly discussed which are more or less likely, but did not go into a deeper analysis.

3.4 Impact on policy planning

The results of the pilot CIE project had very little impact on policy planning of future ESF support so far. The main reasons are the lack of clear results, absence of data about firms' situation after the end of intervention, and absence of an overall picture about the mechanism through which the intervention might work. Most of the results show insignificant effects of the ESF programmes, as was mentioned in the previous section. Although the evaluators claim that there was a significantly positive impact on employment in medium and large firms, and a positive impact on profit in large companies, different methods provide different results (in sign and in magnitude). Therefore, it is difficult to base policy recommendations on one method (ignoring the results of the other) and interpret the presence of an effect only for a selected group of firms. Should the presence of positive impact on large firms result in a recommendation to target the support to large firms in the future? The evaluators do not answer such questions or otherwise discuss the policy implications of their results. The lack of policy recommendations and unclear interpretation of results was also negatively perceived by the contracting authority (the Ministry of Labour and Social Affairs).

Another reason why the policy implications are not discussed within the project is the absence of recent data that would cover the situation of firms after the end of the ESF support. The main dataset used in the project came from the Czech Statistical Office (CZSO). The data from CZSO provide time-consistent information about firm characteristics, but are available with a substantial time lag of almost two years. Therefore, the most recent data used in the project came from 2011. Given the fact that most of the ESF projects were still ongoing in 2011, the analysis says nothing about the medium or long-term effects of the ESF support. The evaluators thus recommend conducting a follow-up analysis on the data from 2012 and 2013 to see if the ESF support has any impact on supported companies after the end of the support.

Another problem with the policy implications of this pilot CIE is the absence of a broader picture. First, people from the ministry expressed a concern that the evaluators did not spend enough time with a qualitative research, and thus did not sufficiently understand the nature and types of the ESF support and the process of selection of successful applicants. This might be crucial for the appropriate choice of

¹⁷ The three mechanisms for the positive impact on employment are the following: increase in the employment due to funding wages in companies, due to employment of implementation teams, and due to increasing competitiveness through higher skills and knowledge of employees.

methods and regression specifications. Second, the ministry believes that if the results are to be applicable in actual policy planning, the quantitative evaluation should be accompanied by a qualitative one. Therefore, the ministry is currently preparing a call for a complementary qualitative analysis that would provide a more specific picture about the mechanism behind the effects found in the CIE evaluation.

Lastly, the methods of CIE are often criticized as having little external validity, in other words, the applicability of results on other ESF programmes is limited. The evaluation unit at the ministry thus believes that we need to complement the pilot evaluation with qualitative analyses and with other quantitative analysis to get a clearer and broader picture of the actual impact of ESF programmes.

However, the aim of this evaluation was not only the evaluation itself, but also an increase in the awareness about the possibilities of CIE methods for the evaluation of public policies. Given the pilot nature of the project, it was very important to ensure its wide publicity and a possibility to use experience from this project in the future evaluations. There were 6 workshops during which the CIE methods were introduced, chosen methods and data issues were discussed, and results were presented to a broad audience consisting mainly of representatives from the ministry and local experts. This CIE project was also presented at various occasions abroad. This part of the project was thus perceived as successful by all stakeholders.

Therefore, the biggest impact of the project on policy planning is probably not on the planning of the future ESF projects, but rather on the planning of evaluation policies. This was confirmed by the ministry, which intends to continue using CIE in the evaluation of ESF support. In 2013, the ministry announced a call for a new project that should use CIE methods to evaluate the ESF projects that promote social entrepreneurship (operational programme HRE).¹⁸ This project is solved by the same team of evaluators as the pilot CIE (IREAS), but the successful completion of the project currently depends on a delivery of individual-level anonymised data from the Czech Social Security Administration (CSSA).

Moreover, the evaluators received additional grant from the Technological Agency of the Czech Republic (TACR) to create a methodological cookbook, which can be used to evaluate ESF projects in the programming period 2014-2020.¹⁹ This should substantially help the successful completion of any future evaluations. The ministry also plans to conduct a follow-up CIE project at firm-level on the most recent data from CZSO (as was suggested by the evaluators in the final report), and the evaluation unit at the ministry hired two new employees for this purpose. However, there is a concern if the ministry employees will be indeed able to conduct a high-quality CIE.

4. Key challenges

The biggest challenge of the pilot CIE project was probably the availability of appropriate data. Initially, the data about firms should have been collected from the business register, because all Czech firms are obliged to report their annual financial statements to the business register by law. However, it turned out that in reality this obligation is fulfilled only by a small proportion of firms. Further, the evaluators

¹⁸ More information on the internet: <http://www.esfcr.cz/zakazky/evaluace-podpory-socialniho-a-inkluzivniho-podnikani-v-op>

¹⁹ More information on the internet: http://www.mpsv.cz/files/clanky/17051/Methodika_CIE_MPSV_131015.pdf

planned to use data from the database Albertina,²⁰ they even presented some initial results based on this dataset, but it turned out that the quality of data is very questionable. At the end, IREAS decided to use data from the Registry of economic subjects collected by the CZSO. This dataset includes time-consistent information about employment, profit and sales of firms. This dataset was combined with data from the monitoring system Monit7+, which contains information about successful and unsuccessful applicants of ESF projects and the identity of appraisal experts.²¹ The CZSO dataset had to be further complemented by survey data to get the information about other outcome variables (investment in training and staff fluctuations).²² The main issue then appeared to be the fact that the CZSO data are available with a substantial lag (of almost two years). Therefore, the evaluators were not able to assess the impact of the ESF project on the firms after the end of the support, but only in the period when the support was still ongoing.

The availability of appropriate datasets is also the biggest issue for the possible future CIE projects in the Czech Republic. The firm-level data from the CZSO can be used, but are still not ideal for programme evaluations. However, the evaluations of most of the ESF projects should be conducted at the individual level, as there are quite few programming areas that focus directly on firms. The availability of individual-level data is much more serious problem. Ideally, the individual-level data from the employment offices register should be matched with the database of the Czech Social Security Administration (CSSA). The Ministry of Labour and Social Affairs have an access to the employment offices register, but have difficulties in gaining access to the CSSA data. As was mentioned above, the ministry is currently negotiating access to this data, but the result of this negotiation is uncertain.

Among challenges in developing the CIE in the Czech Republic was also a lack of previous experience with this methodology. The evaluators from IREAS have a good statistical background and some previous experience with application of econometric methods, but this was the first project of this type in the Czech Republic and they had to solve many issues that arose during the project. In particular, the application of CIE methods on a specific programme always has its specific issues. For example, the fact that the evaluations of the ESF grant projects were conducted in more than one round complicates the use of RDD method. The fact that timing of grant projects was unevenly spread around the whole programming period might complicate the use of the IV method if the appraisal experts change their strictness/leniency over time.²³ Another issue was the fact that firms could have applied for more than one grant project or might have applied for other support from different operational programmes.

Another challenge of the CIE is the interpretation of results in a situation when more than one estimation method is used. This is particularly problematic if different methods provide different results, as was the case in this pilot CIE project. The

²⁰ Albertina is a commercial database of Czech and Slovak firms collected by Bisnode, for details, see: <http://www.bisnode.cz/produkt/albertina/>

²¹ However, these datasets must have been anonymised before the researchers from IREAS could have used them. Therefore, the data from CZSO and from Monit7+ were first matched and anonymised by an employee from the ministry and only then handed out to the evaluators.

²² The evaluators used data from the CZSO survey 'Continuing Vocational Training Survey' for the years 2005 and 2011, and from the survey that took place in 2012 and was conducted by the evaluators themselves.

²³ IREAS tested for the difference in individual experts' behaviour in the first and second part of the programming period, and the differences were mostly insignificant, but they found some differences for the call 35 (Pilot CIE, Progress Report No. 1, Appendix 1, p. 35).

evaluators argue that the IV method has the most credible identification mechanism (CIE, Final Report, p. 8), and thus give priority to the IV results. However, the evaluators do not provide the results of the diagnostic tests of the IV method and do not discuss the external validity of the method.²⁴ Moreover, ignoring the results of the other methods might raise many questions and undermine the trustworthiness of the whole project.

Another related challenge in using CIE methods is the ability to present results in an understandable way to a broad audience of policy makers and general public and at the same time document the results in a transparent way for the academic audience. It seems that the evaluators largely succeeded in the first aspect, but somehow struggled with the second aspect. There was a concern from the ministry and also from some local experts if the estimation was conducted in a correct way, because the results are not very well documented. Even the technical appendices in the final report show only the point estimates and the corresponding p-values. Authors do not report a type of regression specification, a list of other explanatory variables and their coefficients, results of the first stage for the IV approach, or results of diagnostic tests of the methods. It is not even clear if the estimated effects for different types of firms come from separate regressions or from one regression with interaction terms.

5. Success factors and transferability

Policy makers are often concerned about the results of CIE, as these might show that the policy did not have the intended effect. Therefore, one of the main challenges for the CIE evaluation is to persuade policy makers that the evaluations can be beneficial and learning how to spend money efficiently is important. In case of the Czech pilot CIE, this was possible thanks to Dr. Kváča, who promoted the CIE at the ministry and persuaded his supervisors to agree to conduct a pilot CIE project.

The overall success of the CIE projects is also largely influenced by the choice of the evaluators. The call for proposals for this project stated the requirements for the CIE evaluation in a very detailed way including the expected methods to be used and the qualification questions that were meant to test the econometric background of the applicants. The detailed way in which the call had been written was very important for the selection of the best applicant for this project. Therefore, a success factor was again a presence of a ministry employee, who was sufficiently familiar with the methods to be able to prepare such a call and select a high-quality applicant.

The main contribution of this project mentioned by all stakeholders was the mere fact that the project was conducted and the awareness about the methods of CIE was increased. As was mentioned above, although there were some previous attempts to use the CIE methods in the Czech Republic, this project represents the first comprehensive CIE evaluation. To ensure the success of this pilot project, it was very important to manage the publicity of the project and increase the awareness about CIE methods among local experts and policy makers. This part of the project was largely successful, the evaluators in cooperation with the ministry organised several workshops, and shared their experience with the project on various occasions. The progress of the project was well documented at the IREAS webpage, and all materials that were created within the project (including all progress reports and presentations from the workshops) can be easily found on the internet. Further, the evaluators

²⁴ The IV method estimates the effect only for a group of supported firms for which the identity of appraisal experts affected the success of their application.

published two articles in local journals and one article in an international journal sharing their experience with this pilot CIE.

From the methodological point of view, the biggest contribution of the pilot CIE project was the idea to use identity of appraisal experts as an instrumental variable. The idea came from Dr. Kváča, who initiated the project at the ministry. Although similar instrumental variables have been used before in programme evaluations (e.g. in the academic article Dahl, Kostol, and Mogstad, 2014, which is forthcoming in one of the top economic journals), the idea to use this IV in the pilot CIE came naturally from the discussions about the appraisal process.

Another important success factor for the CIE evaluations is an access to appropriate data. To conduct an individual-level evaluation of the ESF projects, the data from the CSSA register would be necessary. However, the ministry still does not have access to this data, the evaluation thus needed to be conducted at firm level. Therefore, the ministry chose the programme area to be evaluated in the pilot CIE with this restriction in mind. However, if further CIE evaluations are to be conducted, individual-level data from the CSSA might be necessary. The evaluation at the firm level is possible only for a limited number of the ESF projects and it does not allow assessing the impact of support on many important outcomes, such as the unemployment probability of supported subjects or their wages. Therefore, the successful completion of most of the CIE projects depends on availability of individual-level datasets that allow tracking supported and unsupported subjects over time. Nevertheless, this pilot ICE project proved that the firm-level data can be used to CIE evaluations as well.

This pilot CIE aimed not only at the evaluation itself, but also at the identification of other ESF projects that could be evaluated using the CIE in the future. This (together with the project publicity) was a very important aspect of the project that helped to prepare ground for future evaluations. Moreover, the evaluators created a methodology for future application of CIE methods on the ESF programmes in 2014-2020 within a follow-up TACR project, which should help the transferability of their experience on future evaluations.

6. References

Angrist, J. D., Pischke, J. F., *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2008.

Barca, F., An agenda for a reformed Cohesion Policy: A place-based approach to meeting European Union challenges and expectations, 2009. Internet: http://www.europarl.europa.eu/meetdocs/2009_2014/documents/regi/dv/barca_report_barca_report_en.pdf

Blundell, R., Dias, M. C., Meghir, C., Reenen, J., Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association*, Vol. 2, No. 4, 2004, pp. 569-606.

Dahl, G. B., Kostol, A. R., Mogstad, M., Family welfare cultures. *The Quarterly Journal of Economics*, Vol. 129, No. 4, 2014, forthcoming.

European Commission, Guidance document on Monitoring and Evaluation of European Cohesion Policy, 2014. Internet:

file:///C:/Users/KI%C3%A1ra%20Ka%C3%AD%C5%A1kov%C3%A1/Downloads/ESF%20monitoring%20and%20evaluation%20guidance.pdf

Hora, O., Sirovátka, T., Tomešová Bartáková, H., Vyhlídal, J., *Hodnocení programů aktivní politiky zaměstnanosti realizovaných v roce 2007 se zaměřením na rekvalifikace (struktura, cílenost, krátkodobé a střednědobé efekty na opuštění evidence)*. [Assessment of programmes of active labour market policies realized in 2007 with aiming at retraining (structure, targeting, short- and long- term effects on leaving the employment offices register).]. Praha: VÚPSV, v.v.i., 2009. ISBN 978-80-7416-045-5.

Hora, O., Sirovátka, T., *Srovnání efektů aktivní politiky zaměstnanosti v České republice v období růstu (2007) a během první fáze krize (2009)* [Comparison of active employment policy effects in the Czech Republic in the period of growth (2007) and during the first phase of the crisis (2009)]. Praha: VÚPSV, v.v.i., 2012. ISBN 978-80-7416-106-3.

Imbens, G. W., Wooldridge, J. M., Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, Vol. 47, No. 1, 2009, pp. 5-86.

Kulhavý, V., Sirovátka, T., *Programy aktivní politiky zaměstnanosti v ČR v roce 2005: Hodnocení efektů na nezaměstnanost v roce 2006*. [Programmes of active employment policy. Evaluation of effects on unemployment in 2006.]. Praha: VÚPSV, v.v.i. - výzkumné centrum Brno, 2007. ISBN 978-80-87007-66-2.

Lechner, M., Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *Review of Economics and Statistics*, Vol. 84, No. 2, 2002, pp. 205-220.

Pilot CIE, Initial Report. Internet: <https://forum.esfcr.cz/node/24/pracovni-skupina-pro-evaluace-esf/library/>

Pilot CIE, Progress Report No. 1. Internet: <https://forum.esfcr.cz/node/24/pracovni-skupina-pro-evaluace-esf/library/>

Pilot CIE, Progress Report No. 2. Internet: <https://forum.esfcr.cz/node/24/pracovni-skupina-pro-evaluace-esf/library/>

Pilot CIE, Final Report. Internet: <https://forum.esfcr.cz/node/24/pracovni-skupina-pro-evaluace-esf/library/>