

DOI: [10.52950/ES.2024.13.1.006](https://doi.org/10.52950/ES.2024.13.1.006)

# "ACTION AND REACTION" - A NEW PUBLIC SECTOR EFFECTIVENESS ASSESSMENT METHODOLOGY

VILEM CAP, MAREK FEURICH, MARTIN LUKAVEC

## Abstract:

The main contribution of the paper is the development of a new way of assessing the efficiency of public administration using statistical methods - multivariate regression (identification of exogenous factors) and residual analysis (identification of regression residuals). The methodology presented is intuitive, practical and, if appropriately set up, globally applicable for the evaluation of a wide range of processes taking place in the public sector on the principle of "action and reaction". Putting the methodology into practice can go some way to satisfying the "taxpayer hunger" for information about whether the money they pay in taxes is being used wisely. It makes it possible to assess efficiency in a particular institution or to compare similar processes in several institutions over a given period of time. The methodology has been developed on the basis of audit findings (prepared by the supreme audit institution for the public sector in a democratic state) and their implications. The developed methodology brings a new perspective to the traditional claim that public sector effectiveness is virtually impossible to measure and that factors for objective evaluation of public sector managers cannot be easily established.

## Keywords:

Performance audit, Supreme audit institution, Evaluation, Public administration, Efficiency, Public finance

**JEL Classification:** H83, H50, C80

## Authors:

VILEM CAP, Prague University of Economics and Business, Czech Republic, Email:

[Vilem.cap@seznam.cz](mailto:Vilem.cap@seznam.cz)

MAREK FEURICH, Prague University of Economics and Business, Czech Republic, Email:

[feurichmarek@gmail.com](mailto:feurichmarek@gmail.com)

MARTIN LUKAVEC, Prague University of Economics and Business, Czech Republic, Email:

[lukavec@gmail.com](mailto:lukavec@gmail.com)

## Citation:

VILEM CAP, MAREK FEURICH, MARTIN LUKAVEC (2024). "Action and reaction" - a new public sector effectiveness assessment methodology. *International Journal of Economic Sciences*, Vol. XIII(1), pp. 106-146., [10.52950/ES.2024.13.1.006](https://doi.org/10.52950/ES.2024.13.1.006)

ORCID - Vilem CAP: <https://orcid.org/0000-0003-4848-2355>

ORCID - Marek FEURICH: <https://orcid.org/0000-0001-6264-9962>

ORCID - Martin LUKAVEC: <https://orcid.org/0000-0003-4307-0781>

Funding:

This work is financed by the Internal Grant of the University of Economics in Prague with the title "Proposal of a methodology for evaluating the impact of the SAO's audit conclusions on the economy and efficiency of the use of public funds within the system of public control and proposal of opportunities to improve the SAO's position towards territorial self-governing units" [grant number IG507020].

Data availability statement: All initial data are publicly available on the Czech SAO website.

Disclosure statement: No potential conflict of interest was reported by the author(s).

## INTRODUCTION

The main purpose of this paper is to show that the public administration and the execution of public policy can be evaluated using statistical methods. The main problem with public policy, how well it functions and delivers its objectives, is the difficulty to quantify the results (Vedung, 2017). Public policy by its nature often aims to deliver unpriced social benefits to a country's residents, such as public goods and others (Cornes and Sandler, 1996). As such using money as the main measure is often not a possibility. Plus, while for evaluating of economic or business objectives, we can use well-established and widely used measures such as economic growth, inflation, profits and many others, this becomes impossible for public policy. The evaluation of the success or failure of many public policy initiatives or interventions needs to be approached in a programme-specific manner. This approach often requires in-depth qualitative analysis. However, in our research we do not seek to evaluate the success of individual programmes in detail, but to identify cases that deviate from the norm, which can then be then targeted by stakeholders. In other words, to identify cases that are underperforming or overperforming.

In this paper, we set to illustrate that you can evaluate effectiveness of public agencies using a point system, to which is applied fairly simple and widely used statistical methods, concretely multivariate regression and residual analysis. The purpose of the regression is mainly to control for various factors and to exclude them from the analysis. What remains is the unexplained portion of the performance, which can then be used to identify problematic cases and underperforming institutions.

As a case study on which we decided to illustrate this methodology is on purpose difficult: the effectiveness of the audits by the Czech Republic Supreme Audit Office (hereinafter referred to as the "SAO") on public institutions, that is whether the state administration took some steps after problems were identified and whether these problems were later rectified. The SAO is chosen because the Lima Declaration set new principles that public sector audits should include audits of all parts of the public administration and subsequently inform the government and the public about their results (INTOSAI, 2019). The traditional activities carried out by SAIs have been legal and financial audits (García, 2005), but later this activity was expanded to include aspects related to the New Public Management, which led to changes towards higher measurement of inputs and outputs (Pollitt, et al., 1999). As the nature of these problems differs, public administration deals with a changing load of work and various budgetary and other constraints over time, we used the error/violation types, as well as time, as controls in a regression. The unexplained residuals of the regression then point to audits, which, after being controlled for the above-mentioned factors, didn't result in an expected reaction of the public administration, can be evaluated as cases of under or over-performance and should be a subject of further examination (as, in this case, can be considered suspicious).

Moreover, the academic literature finds no consensus on a generally accepted methodology for assessing the impacts of audit conclusions (see e.g., Power, 1999; Lonsdale, 2000, Van Looke and Put, 2011; Reichborn-Kjennerud and Johnsen, 2015). According to many authors (see e.g., Pollitt et al., 1999; Morin, 2001, 2014; Kells, 2011) the main obstacle is the lack of

objective data and the difficulty in identifying changes that have been made based on a specific audit conclusion. Furthermore, as Persson and Nasstrom (2016) state, factors affecting the impacts of audit conclusions are not clearly defined. Developing a comprehensive methodology for evaluating the impact of the Supreme Audit Office audit conclusions and identifying the factors that influence these impacts can thus make a significant contribution to addressing this issue. Therefore, in addition to identifying problematic cases that are not given sufficient attention by the state administration, this paper seeks to answer the following questions that have arisen from the above:

*Which indicators can be used to objectively assess the impact of the inspection/audit/evaluation findings (of a specific public sector institutions)?*

*What are the impacts of the findings of the inspection/audit/evaluation (of a specific public sector institution)?*

*What factors influence the impact of the findings of the inspection/audit/evaluation (of a specific public sector institution)?*

We would like to emphasize that this methodology (with a modification of the scoring system) has the potential to assess the effectiveness of virtually any process in the public sector. In order to maintain the objectivity of the research, it should be noted that the methodology is particularly useful in situations where a specific authority (a private company, a higher body in the public sector hierarchy in a given country, or another public sector body that can carry out audits) is carrying out some form of inspection/audit/evaluation of the processes (and the results achieved by them) of a specific public sector organization. Specific ex-ante/ex-post objectives, based on the scope of the public sector organization's activities or the objectives it should have achieved in the past, can be monitored and evaluated. For each analysis, it is necessary to determine in advance: 1) whether one or more of the institutions to be studied will be involved; 2) what objectives we want to assess; 3) what influences we expect to have/had on the achievement of the objectives mentioned in the previous point.

The paper is structured as follows. The next presents a literature review which deals with the problematic of public administration evaluation, with focus on audit conclusions and how to measure their impact. Following section focuses on the methodology of empirical research, where the unique approach of this paper is closely described. Then is presented the analysis of impact indicators, regression analysis of the factors influencing these impacts and results. Subsequent section discusses the empirical results with contemporary scientific knowledge presented in a literature review, and the last section deals with conclusions and outlines possible limitations and recommendations for future research.

## **LITERATURE REVIEW**

### ***Evaluation of public administration and the execution of public policy***

Public administration and public policy are two key elements of a modern democratic state that are essential to ensure the quality and efficiency of the public sector. This is a highly debated topic and the academic literature points to many different approaches to evaluating public administration. Peters (1998) argues that performance measurement should be based on

objective indicators that will allow comparisons of performance between different government organizations and provide relevant information for resource allocation decisions. In this sense, he suggests that common standards and indicators should be established to enable comparisons of public administration performance at a horizontal level. Lynch and Day (1996) state that performance measurement should be based on realistic expectations of what government organizations can and cannot control. It would also be ideal for performance measurement systems to be tailored to the organization. Even Bouckaert and Halligan (2008) emphasize the importance of standardized indicators and methods in comparing public administration performance but point to problems in evaluation that may be caused, for example, by different structures of different government organizations or different levels of public administration development in different sectors. However, implementing data standardization practices is not an easy task and public sector organizations face a number of challenges. These challenges include a lack of management and organizational support, policy, standards and political issues, human resource capacity and a lack of governance (Ventura et al., 2023). In addition, the transition from existing accounting systems to standardized systems can be complex and time-consuming, requiring the adaptation of computer systems and numerous procedures (Hassan et al., 2020). The process of data standardization itself can be difficult and slow, with coordination problems and delayed benefits hampering progress (Berner and Judge, 2019). In addition, the quality of data standards and the costly process of developing and maintaining them are ongoing challenges (McAuliffe, 2019). Overcoming these challenges requires addressing issues such as resource allocation, stakeholder engagement and the development of effective governance mechanisms (Hongwei et al., 2016).

Moreover, it is unrealistic for state and local government units to have their own employees with the analytical background needed to measure their individual performance (McGowan and Polster, 1985). Similarly, it is not realistic for academics to create individualized assessment systems for each individual institution. The following chapters, therefore, describe the theoretical basis for the development of a general framework for public administration evaluation, which is, however, based on an attempt to evaluate the performance of a specific institution (i.e., the SAO) whose sphere of influence is sufficiently broad to encompass virtually all parts of the public administration of a given state.

### ***Role of Supreme Audit Offices***

SAOs play an important role in national democratic systems; their role is to provide information that highlights both good and inefficient public structures (Lonsdale, 2000). This places the SAO in a new position as a public accountant and evaluator, assessing whether public administration is operating economically, efficiently, effectively and transparently (Pollitt, et al., 1999). This new type of evaluation activity, which pays more attention to the system as a whole, was subsequently termed 'performance auditing' (PA) (Barzelay, 1997). It is worth noting, however, that the very way in which SAs operate has also moved towards efficiency principles (Pollitt, 2003). In the context of New Public Management, it is therefore necessary to evaluate the performance of SAs and to measure the impact of their activities (Lonsdale, 2000). Although the purpose of a PA is to contribute to the improvement of the performance of audited

entities and the general assumption is that PAs in the public sector have positive effects, Reichborn-Kjennerud (2014) argues that these assumptions are not based on empirical evidence. Morin (2001) also states that although the effectiveness of PA can be confirmed, it should not be taken for granted. Van Loocke and Put (2011), who replicated Morin's (2001) research, concluded that Morin's (2001) findings are still valid and that a performance audit may have some impact, but this impact can be very insignificant. On the other hand, one of the assumptions associated with higher impact audits that academics have been able to uncover is the type of SAO model.

### ***Possibilities for evaluating the effectiveness of the SAO***

Some academics have expressed concerns about evaluation of the impact of PA. Lonsdale (2000) states that not only is it difficult to measure the effectiveness of PA on the improvement of public service management, but the approach to measuring it is a contentious issue. The outputs of government are always difficult to evaluate, as they usually do not provide clear output units, but rather a description of its performance, from which output indicators often need to be derived (Niskanen, 1971). Power (1999) also highlights the difficulties in measuring the impact of PA on government and argues that the real impact of audits can only be measured by the perception of auditors or auditees, which brings a high level of subjectivity. Van Loocke and Put (2011) states that the impact of PA can be defined as a wide range of direct or indirect effects, required or undesirable, that PA can exert on the audited entity, which complicates the situation even more. Reichborn-Kjennerud and Johnsen (2015) further argue that measuring the impact of PA cannot be limited to the implemented changes that were proposed in the recommendations, because the audit process itself affects the audited entity both negatively and positively (according to the auditor's findings). Moreover, according to many authors (see e.g., Pollitt et al., 1999; Morin, 2001, 2014; Kells, 2011) the main obstacle is the lack of objective data on the impact of PA and the difficulty in identifying changes in state administration or in the audited organization made based on a specific conclusion of the PA.

Nevertheless, some approaches that can be used to measure the impact of a PA have emerged. It is possible to analyze, for example, the impact of PA conclusions on parliaments, media, public debate, or cost savings caused by the application of recommended measures (Pollitt et al. 1999). However, Talbot and Wiggan (2010) consider this approach to be fragmented and inadequate, and therefore present a conceptual framework (the public value scorecard and competing values framework) that addresses multiple dimensions and meets the needs of multiple stakeholders. Van Loocke and Put (2011) also recommend going beyond the standard indicators that auditors themselves usually use to assess impact, such as the percentage of recommendations implemented by the audited entity or the cash savings. Audits may also have other impacts, such as improving transparency, knowledge, or accountability of the audited entity. Some authors focus on the impact of PA at the macro level and analyze the impact of PA on fiscal policy, government efficiency, or capital productivity (Blume & Voigt, 2007; Schelker, 2008). At the micro level, the most common measure of the impact of PA are the implemented recommendations (Milgrom & Schwartz, 2008; Van Loocke & Put, 2011). Another approach by Morin (2001, 2008) and Reichborn-Kjennerud (2014), who measured the impact of PA, based on the opinions of auditees and the contribution of PA results, also

contributed to the discussion. A similar approach was chosen by Raudla et al. (2016), who argue that measuring the number of changes made is not an appropriate methodology, because change is an elusive indicator, and therefore instead focused his research on analyzing the perceived usefulness of PA which he considers to be a more accurate methodology for measuring audit impacts. However, Morin (2001, 2008) and Pollitt et al. (1999) argues that the factors influencing the impact of an audit are very difficult to determine in practice and some theories in this area are still not entirely based on empirical evidence. Therefore, the impact of an audit cannot be explained by a single theory.

Despite all the research carried out, it is still not entirely clear what impact PA conclusions have on the public sector and further empirical research is needed. There are no general conclusions explaining the impact of PA conclusions, nor are the factors that may affect these impacts clearly defined (Persson & Nasstrom, 2016). In the absence of a generally accepted methodology for measuring the impacts of PAs and what impacts to measure, the methodology developed in this paper (described in Section 3) can make an important contribution to addressing this issue and, at the same time, provide a new perspective on measuring the effectiveness of the activities of individual organizations in the public sector.

## **METHODOLOGY**

The aim of this section is to present a framework for assessing the impact of the SAO's audit findings, which can be used to identify factors that influence the impact of the SAO's audit findings and to highlight suspicious cases that should be further investigated by the public administration.

In terms of structure, the Czech SAO is closest to the Collegial model. This methodology, however, considers limiting factors arising from differences between SAOs such as internal hierarchy, approaches to drawing audit conclusions, different number and type of audited entities, the limit to the number of office staff, different ways of planning controls or a varying approach to re-audits when errors are found. Taking these aspects into account, the methodology has been designed to be as general as possible, abstracting away from the above differences and applicable to the widest possible range of public authorities.

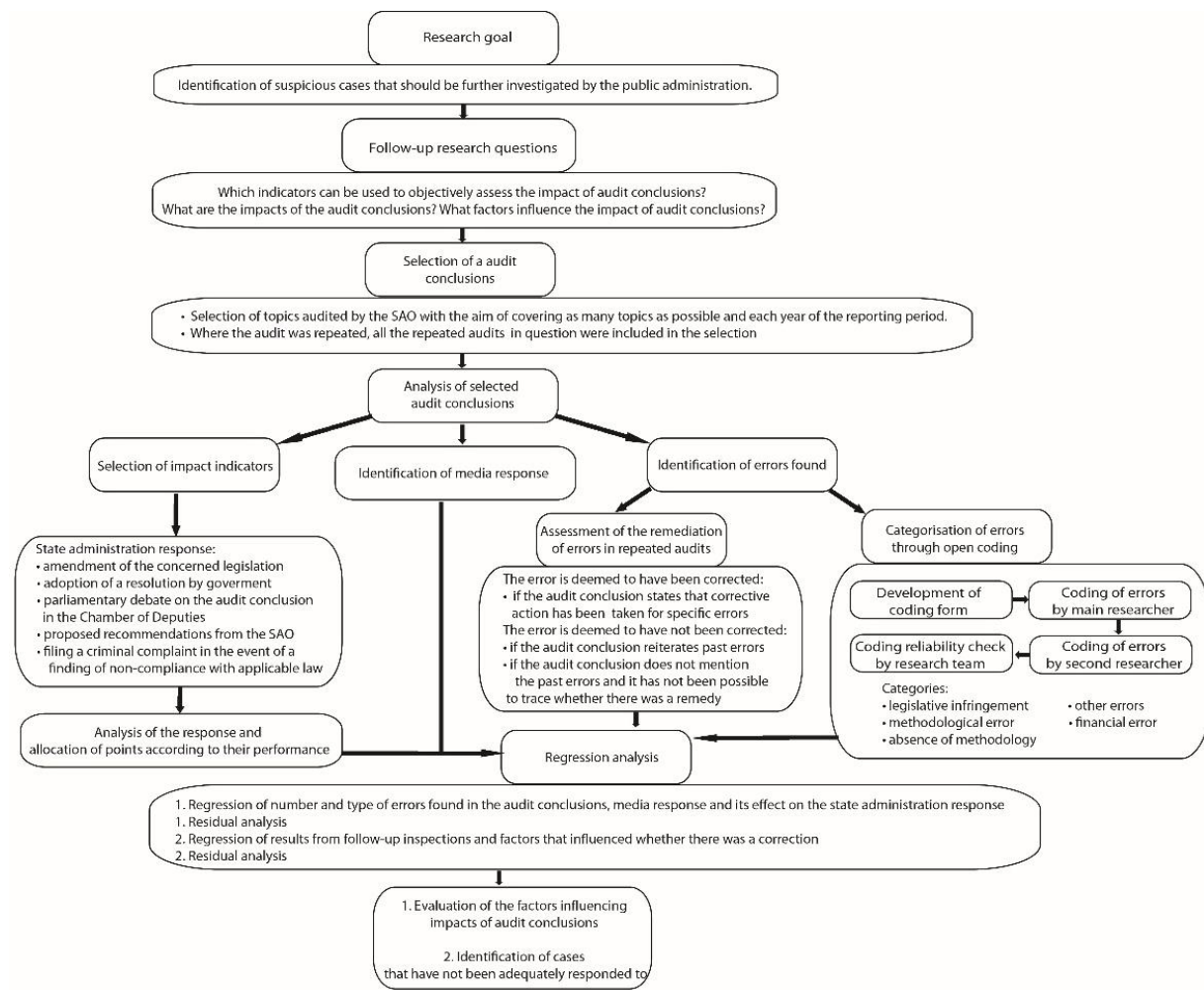
The Czech SAO publishes audit conclusions in the SAO Bulletin and on its website. The conclusions are numbered according to the year and the order of inclusion of the audit in the plan of audit activities for the relevant calendar year. Each audit conclusion has its own number, and it contains a list of inspected entities; the name of the person performing the inspection; a list of legal regulations affected in the audit conclusion, as well as information on whether the audit plan was discussed in the parliament and the government. And equally important, it has a statement of whether, based on the findings, the SAO filed a notification with the tax administrator and/or a criminal complaint. Published audits can be filtered on the SAO website by person, number, keyword, by time or by 30 different topics (housing, tourism, tax and other state revenues, transport, subsidies, etc.).

This research analyzed 101 SAO audit conclusions from the period 2000-2019. During this period, the SAO carried out 727 audits. However, the aim of this work was not to analyze as many audit conclusions as possible, but to develop a functional methodology that would allow for a larger data set to be analyzed in the future. A scheme of the methodology of this paper is shown in Figure 1.

The audits were not selected at random. The aim was to include, as far as possible, all 30 topics covered by the SAO in order to avoid distortions caused, for example, by audits with a one-sided focus. In the end, 29 topics were included. At the same time, the entire period under review (2000-2019) was covered so that each year was represented by at least one audit action. Where the audit was repeated, all of the repeated audits in question were included in the selection. Thus, of the 101 audit conclusions selected, 47 related to repeat audits. The individual audit conclusions were subsequently analyzed in detail with a focus on the identification of the **errors found** and the selection of **impact indicators**. The identified errors were further analyzed through open coding (see e.g., Corbin and Strauss, 1990) and grouped into 5 categories, based on its common characteristics. Selected indicators, identified errors, their rectifications and the categorization were double-checked by the research team. The individual categories of errors are described in Table 1. In addition to the errors found, information on whether the individual audit conclusion and the errors found were reported in the media was also included in the model as a dummy variable. Whether the audit conclusion was presented in the media was searched via Google, in the period of the month and year of the audit conclusion publication and 5 years after the audit conclusion publication. Keywords were used. The searches were carried out independently by 2 researchers and their results compared. If there was at least one mention in the media during the observation period, the value of the variable in the model is 1, otherwise 0.



**Figure 1.** Scheme of the methodological procedure.



Source: own adaptation and processing

**Table 1.** Groups of errors found.

	Type of error	Examples of errors
a	<b>Legislative infringement</b>	Non-compliance with the constitution, violation of the law, violation of government regulations, non-compliance with the ministerial decree, non-compliance with the decree approved by the municipal council, non-compliance with the regulation approved by the municipal council, etc. This also includes situations where the audited entity did not act efficiently and economically, due to bad management or poor budgetary discipline.

b	<b>Methodological error</b>	Non-compliance with the methodology for assessing grants or grant applications, breach of the internal public procurement directive (in this case it is not a breach of the law), use of outdated methodology for providing repayable financial assistance, or other non-compliance. It also includes human error due to a lack of attention to detail (sometimes 'laziness') of the audited entity in compliance with the methodological area or the existence of such a methodology that does not contain all the essential aspects that it should contain to work effectively.
c	<b>Absence of methodology</b>	The absence of standardized work procedures for individual agendas or the absence of indicators, these missing elements prevent compliance with applicable legislation or prevent the effectiveness of distribution and use of funds for audited entities. We include cases where the methodology exists, but it is set incorrectly for purposeful reasons with a probability bordering on certainty, which intentionally allows inappropriate conduct of the inspected entity.
d	<b>Other errors</b>	Errors/deficiencies that do not correspond to any of the categories a, b, c or e. These are, for example are about un-systematicity, a failure of the institution to perform a function, without a clear indication that the law was explicitly violated or the receipt of funds by the inspected entity for a certain activity, which did not provide the expected effect. In this category we also include premature terminations of activities that the inspected entities were to perform.
e	<b>Financial error</b>	Incorrect accounting of financial operations, individual errors in accounting e.g., incorrect classification of items, errors in the financial settlement of earmarked funds, errors in financial statements, use of funds contrary to the valid budget structure and a situation in which a matter was entered in the accounts which should not have appeared.

**Source:** own adaptation and processing

For repeated audits, whether previous errors had been rectified was analyzed. An error was considered to have been rectified when the audit conclusion indicated that corrective action had been taken. An error was not considered to be rectified if the audit conclusion mentioned the error again or if the audit conclusion did not mention the error and it was not possible to determine whether the rectification had been made.

**Impact indicators** selection was based on generality, objective quantifiability and occurrence in individual audit conclusions. The individual impact indicators listed below represent the government's options for addressing the audit conclusions. A total of 5 impact indicators were selected. Indicators were chosen with the assumption that the mere elicitation of a response that may lead to successful remediation in the future is an impact. The process of assessing each of them is described below. The scoring system was developed for the purpose of regression analysis.

#### *Amendment of the concerned legislation*

Each audit conclusion contains a list of legislation relevant to it. When there was a change to any law that appeared in the "front" of the audit conclusion (entry into force of the amendment), from 2 to 13 months from the date of approval of the conclusion, then 1 point was awarded to such an audit conclusion. This means, for example, that if the list of affected legal regulations contains 4 laws and a change (within 2-13 months from the date of approval of the audit conclusion) occurs in 3 of them, 0.75 points was awarded ( $3/4 = 0.75$ ). The time period from the date of approval of the conclusion of the SAO's audit conclusion and amendments to the laws in question were determined based on consultation with the members of the Czech Parliament. This indicator can be used to assume whether the result of the SAO's control activities brought about a tangible change in legal regulations. Although it cannot be proven with certainty that the change in legislation was directly caused by the SAO's audit conclusion, this indicator is an important element that should not be overlooked.

*Points in the range 0 - 1 can be obtained for this evaluation factor.*

*"0" indicates the "zero" effect of the SAO's audit on changes in legislation.*

*"1" indicates the absolute impact of the SAO's audit on changes in legislation.*

#### *Discussions in the government*

Each audit conclusion states whether the audit conclusion was discussed by the Government of the Czech Republic. The discussion at the meeting of the government is an important act from the point of view of arranging a resolution, especially in the audit conclusions concerning the central state administration bodies. In cases where the government adopted (at least once) a resolution to the audit conclusion, 1 point was awarded. Otherwise, 0 points were awarded.

#### *Discussions in the Chamber of Deputies*

The audit conclusions also contain a statement as to whether its results were discussed at least once on the premises of the Parliament of the Czech Republic. If the "front side" of the audit conclusion indicates that it was discussed in the body (or directly in) the Chamber of Deputies of the Parliament of the Czech Republic, 1 point was awarded, otherwise 0 points were awarded.

### *Recommendations in the audit conclusion*

If the audit conclusion contains at least one specific recommendation, 1 point was awarded, if it does not contain any, 0 points were awarded. This variable indicates the audit authority's own response to the misconduct found, which should not be neglected in evaluating the impact of the audit conclusion.

### *Conflict with legislation*

As part of this factor, we monitored whether an error was found in the audit conclusion, which indicates a conflict with applicable legislation. If not, it was no longer dealt with for this conclusion, and it was marked with the letter "N". If a conflict with the applicable legislation was found, we monitored whether a criminal report was filed. If the report was filled 1 point was awarded, otherwise 0 points were awarded. This variable is included in the analysis for the same reason as the previous one mentioned above in point IV.

Subsequently, the impact indicators were evaluated, and a double regression analysis was carried out on the collected sample of SAO reports. The regression analysis in both cases is based on multivariate linear regression. The first regression deals with the effect of the number and type of NKU findings (such as intentional or unintentional breaking of or non-compliance with laws, decrees, etc.) and media response on the actions which the state administration (both executive and legislative power) took afterwards. This analysis is based on all 101 observations (audit conclusions). The resulting response of the state administration is measured by the sum of the above-mentioned points. The second regression deals with follow-up inspections (usually a few years after the initial inspections) and factors that influenced whether the problems identified in the first instance reports were or weren't rectified. This analysis is based on a sample of 47 follow-up inspections by the Supreme Audit Office, which returned errors/violations of the laws, decrees, etc. identified during previous inspections. This regression analysis tries to quantify whether it is possible to predict how large a percentage of errors/violations has been rectified in the time between the original and follow-up inspections. Explanatory variables include factors such as the number of errors in individual categories, the response of the state administration, media response, the year, and the time for rectification. Both regressions were run twice, first as an exploratory regression to identify cases which deviate from residual normality - these were removed, the regression was run again, and the removed observations were returned. This limits the influence of the outliers on the residual values. Robust standard errors were used.

Furthermore, residual analysis is used, which considers the unexplained part of the above-mentioned regressions and uses it to identify controls that did not have the response that could be expected according to the regression analysis. In the first section of the analysis, the residual analysis deals with the inspections that did not elicit the response of the state administration that could be expected based on the reactions to similar errors, and in the second case, these are cases where the expected rectifications of previous errors did not occur. The core idea behind the use of the residual analysis is to eliminate legitimate factors such as type of the public institution error/omission/law-breaking, so only truly problematic cases are identified. For example, if we find that certain inspections identified an unusually

higher number of problems of the type that usually induce the state agencies to order a rectification or even change relevant laws, but this rectification has not materialized, the residual analysis identifies this situation.

## **EMPIRICAL ANALYSIS**

### ***Evaluation of impact indicators, media response and success of correction***

In 95% of cases, at least one of the laws concerned was amended. In just 4 audit conclusions, there was no subsequent change in the legislation. For the laws concerned, the 'amendment rate' was 53%. However, in the Czech legislative process, there is no obligation to state in the explanatory memorandum for a change in the law that these changes are triggered by the activities of the SAO, so this indicator should therefore be taken with caution. However, whether there was a change due to the audit conclusion cannot be excluded and therefore this variable was also included in the regression analysis. The government adopted a resolution on the audit conclusion in 100 cases. Thus, out of 101 cases, the government responded to an audit conclusion in 99% of cases. In the course of the analysis, however, it was found that the mere issuance of a resolution does not mean that the error in question will be rectified, as there is no sophisticated follow-up system (or its existence is not known to the authors of the article) to verify that corrective action has been taken. A total of 52% of the analyzed audit conclusions were discussed in the Chamber of Deputies of the Parliament of the Czech Republic. Recommendations appeared in the audit conclusions in 19 % of cases. No conflict with legislation was found in 30 cases. However, of the remaining 71 audit findings where a conflict with the applicable legislation was found, criminal reports were filed by the SAO in 4 cases, which is 5.6%. Media coverage was found for 42% of the audit findings analysed. In total, 326 errors were identified in the repeated control conclusions analysed, of which 56 errors (or 17.18%) were corrected. In 5 cases were all the errors found in the previous inspection corrected.

### ***Regression analysis***

#### ***Regression analysis I***

The methodology is motivated by the following idea: we want to identify suspicious cases based on the collected data via residual analysis. The residuals of these suspicious cases are expected to deviate from expected normal distribution. As some of these residual observations are very likely to have a large leverage over the regression line, the regression is first run with these observations included after which the observations which significantly deviate from the normal distribution of residuals are identified and removed. Then, the regression is run again, without these residuals, to map the regression results properties and confirm the regression validity. The previously removed observations are then returned to the analysis and the residual values are calculated for all observations based on the results of the second regression run.

In the regression model, the inspection date is used (to control for the influence of time as a variable), as well as the type and number of individual errors/violations of the law or decree found. The regression results are shown in Table 2. The regression model is defined as follows:

$$\ln(\text{ActionsSA} + 1) = \beta_0 + \beta_1 \ln(\text{Type}_1 + 1) + \beta_2 \ln(\text{Type}_2 + 1) + \beta_3 \ln(\text{Type}_3 + 1) + \beta_4 \ln(\text{Type}_4 + 1) + \beta_5 \ln(\text{Type}_5 + 1) + \beta_6 \text{Media} + \beta_7 \text{Date} + \varepsilon \quad (1)$$

Where:

*ActionsSA* – state administration actions - number of points according to the methodology described above

*Type1* – Number of law or regulation-related misconducts identified by the given report - violation of the law, decree, regulation, etc.

*Type 2* – Number of methodological errors identified by the given inspection - sloppiness, methodological errors

*Type 3* – Number of instances of missing, incorrectly or insufficiently defined processes

*Type 4* – Number of other errors/omissions identified by the reports

*Type 5* – Number of financial errors - errors in accounting, etc

*Media* – A dummy variable for whether the report was mentioned by internet media after publishing

*Date* – The day on which the inspection report was approved

$\beta_0$  – Intercept

$\beta_1 - \beta_7$  – slopes (effects) of regression curves

$\varepsilon$  - errors – unexplained part of the variance (residuals)

**Table 2.** Results of regression analysis I

<b>State administration actions</b>		
	<i>Ln of State administration response (prior outlier removal)</i>	<i>Ln of State administration response (after outlier removal)</i>
LnType1	0.0150 (0.3438)	0.0301 (0.0321)
LnType2	-0.0278 (0.0435)	-0.0648 * (0.0380)
LnType3	0.0559 (0.0492)	0.0826 * (0.0394)
LnType4	-0.0062 (0.0343)	0.0018 (0.0344)
LnType5	-0.0583 (0.0404)	-0.0559 (0.0377)
Media	-0.0339 (0.0525)	-0.0647 (0.0462)
Date	0.0000 * (0.0000)	0.0000 *** (0.0000)
Intercept	-0.3484 (0.6041)	-0.6021 (0.4648)
Observations	101	98
R-squared	0.1123	0.1777

Robust st. errors in parentheses

\* p&lt;0.1 \*\*p&lt;0.01 \*\*\*p&lt;0.001

**Source:** own adaptation and processing

The results of the first regression (prior to the outlier removal) are included only for information. This regression graphical post estimation is included in Figures 4 (Q-Q plot of the residuals), 5 (Added-variable plots) and 6 (Leverage plot of the residuals) of the Appendix, to identify the outliers. The results of the second regression run (after outlier removal) are also shown in Table 2. The coefficient of determination is 17.8%, meaning the included variables explain just 17.8% of the overall variance of the state administration reaction. Date at which the inspection report was approved is significant at 99.9% and Type 2 and Type 3 errors are statistically significant predictors, at 99% level of significance. The other variables are statistically insignificant.

The post estimation for multicollinearity, for the fit of the regression model and normality of residuals, is provided in the appendix (see Figures 7-9 – Q-Q plot, Added-variable plots, Leverage plot and VIF scores in Table 4). No significant multicollinearity issue has been identified in the sample. The residuals are not normally distributed, but the difference is fairly

small and an impact on the results has been eliminated by using robust standard errors to confirm the significance of individual variables.

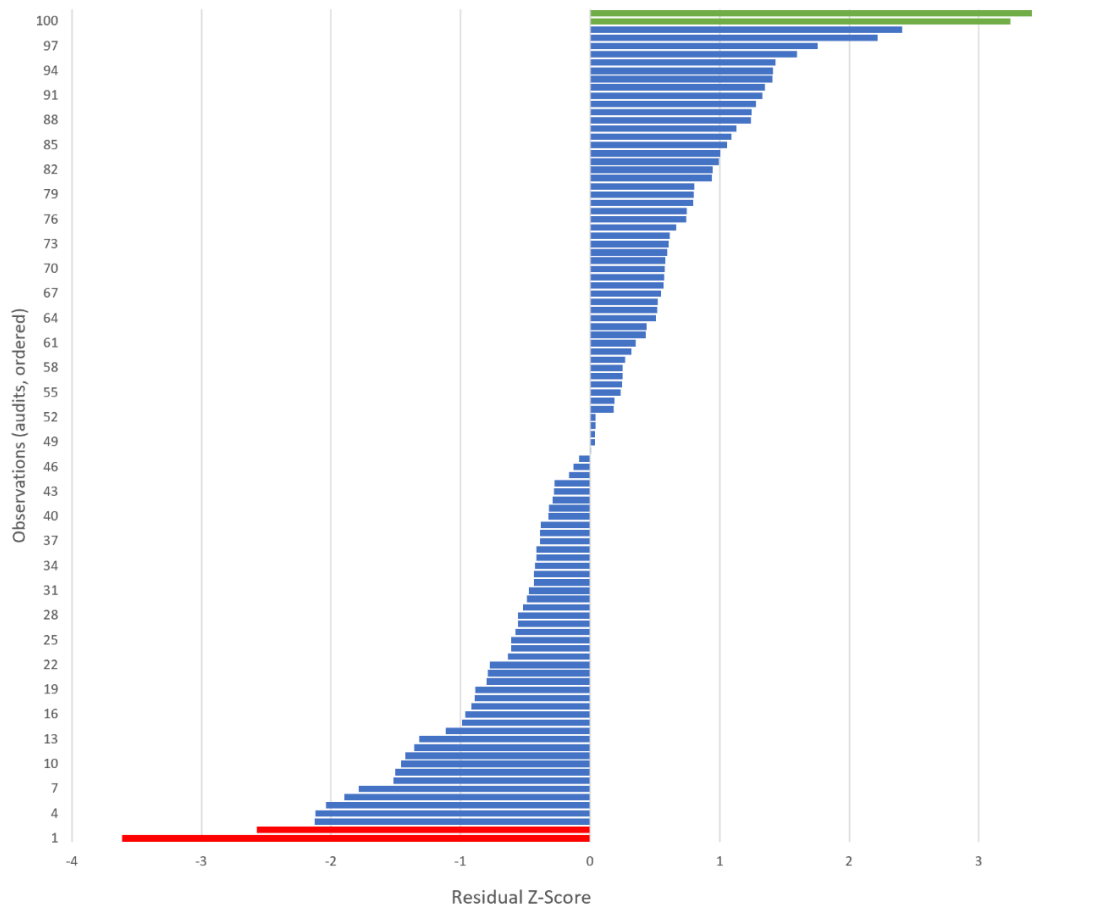
### *Regression analysis I – residual analysis*

The above-described regression analysis shows that the included variables allow us to explain only 17.8% of the variance. Each inspection and situation in each inspected institution is specific, so it is understandable that the range of the unexplained remains high. However, this unexplained part of the analysis can itself be a guide to identifying problems. Regression residuals are the differences between the observed and expected values. All variables used in the analysis above can be considered controls with influence on the response of the state administration. If the response was statistically different from what was expected and thus improbable, we can use this knowledge to identify suspicious situations, which deserve special attention on individual bases.

The improbability of these deviations from expected values can be assessed by the standard deviation of the residuals (see Figure 2). Values higher than 3 standard deviations (so-called Z-score  $>3$ ) are very improbable and therefore suspicious. It is necessary to note that the state administration did well: there was only one instance state administration of under-response that could be considered very improbable (in excess of 3 standard deviations), and two instances of over-response. As this is a draft methodology, we present a few high and low values for individual administrations that would merit further investigation. The sensitivity of the monitoring can be increased by reducing the Z-score (e.g., to a value of 2.5, as we did in the Figure 2).



**Figure 2.** Z-Scores of state administration actions (minus – less than expected, plus – more than expected, high negative residuals shown in red, high positive residuals shown in green)



**Source:** own adaptation and processing

Two high value negative residuals (responses which didn't elicit a sufficient response from the state administration given the regression models) are highlighted in red. These should be given special attention to examining what specific situations lead to low responses. Two high value positive residuals (responses which elicit a higher than expected response from the state administration given the regression models) are highlighted in green. Here, again, a special attention would be beneficial.

*Regression analysis II*

The regression results are shown in the Table 3. The regression model has been defined as:

$$Ln(Rectified + 1) = \beta_0 + \beta_1 Ln(Type_1 + 1) + \beta_2 Ln(Type_2 + 1) + \beta_3 Ln(Type_3 + 1) + \beta_4 Ln(Type_4 + 1) + \beta_5 Ln(Type_5 + 1) + \beta_6 Ln(ActionsSA + 1) + \beta_7 Date + \beta_8 DateDiff + \beta_9 Media + \varepsilon \quad (2)$$

Where:

*Rectified* – percentage errors/omissions which were rectified since the previous

*ActionsSA* – state administration actions - number of points according to the methodology described above

*Type1* – Number of law or regulation-related misconducts identified by the given report - violation of the law, decree, regulation, etc.

*Type 2* – Number of methodological errors identified by the given inspection - slopiness, methodological errors

*Type 3* – Number of instances of missing, incorrectly or insufficiently defined processes

*Type 4* – Number of other errors/omissions identified by the reports

*Type 5* – Number of financial errors - errors in accounting, etc.

*Date* – The date on which the follow-up inspection was approved

*DateDiff* – Number of days since the previous inspection

*Media* – A dummy variable for whether the report was mentioned by internet media after publishing

$\beta_0$  – Intercept

$\beta_1$ -  $\beta_8$  – slopes (effects) of regression curves

$\varepsilon$  - errors – unexplained part of the variance (residuals)

**Table 3.** Results of the regression analysis II  
**Percent rectified**

	<i>Ln of percent rectified (prior outlier removal)</i>	<i>Ln of percent rectified (after outlier removal)</i>
LnType1	-0.0321 (0.0389)	-0.0228 (0.0237)
LnType2	-0.1545 * (0.0709)	-0.1144 ** (0.0405)
LnType3	-0.0116 (0.0770)	0.0234 (0.0331)
LnType4	0.0886 (0.0717)	0.0190 (0.0336)
LnType5	-0.0334 (0.0825)	-0.0273 (0.0378)
LnSA response	0.2863 * (0.1144)	0.1999 * (0.0777)
Date	-0.0000 (0.0000)	-0.0000 ** (0.0000)
DateDiff	0.0001 * (0.0000)	0.0001 ** (0.0000)
Media	0.1135 (0.0856)	0.1141 * (0.0545)
Intercept	0.8694 (0.8340)	1.7057 ** (0.5412)
Observations	47	40
R-squared	0.2818	0.5319

Robust st. errors in parentheses

\* p<0.1 \*\*p<0.01 \*\*\*p<0.001

**Source:** own adaptation and processing

The approach in this second model is similar to the Model 1 – the regression is first run on the whole dataset and this regression is used only to identify outliers and to establish normality of the residuals (with suspicious cases, deviating from normality, removed). After the outlier eliminations, a second regression is run after which the previously removed outliers are returned to the dataset for the residual analysis. As with Model 1, the regression results for the first regression run (prior to outlier removal) in Model 2 is again only for information. The identification of outliers is graphically shown in Appendix, Figures 10 – 12 (Q-Q plot, Added-variable plots and leverage plot).

The percentage of all errors rectified, in relation to all errors/violations of the laws, decrees, etc., was chosen as the dependent variable in both runs of the regression. It was not possible to include the errors/violations by category, as this would split the analysis into 5 regressions and most results would be affected by too small a number of observations and the predictive power as well as the statistical significance of the model would be lower. The coefficient of determination is equal to 0.532. Thus, this regression model has a good predictive value, with almost 53.2% of the overall variance explained.

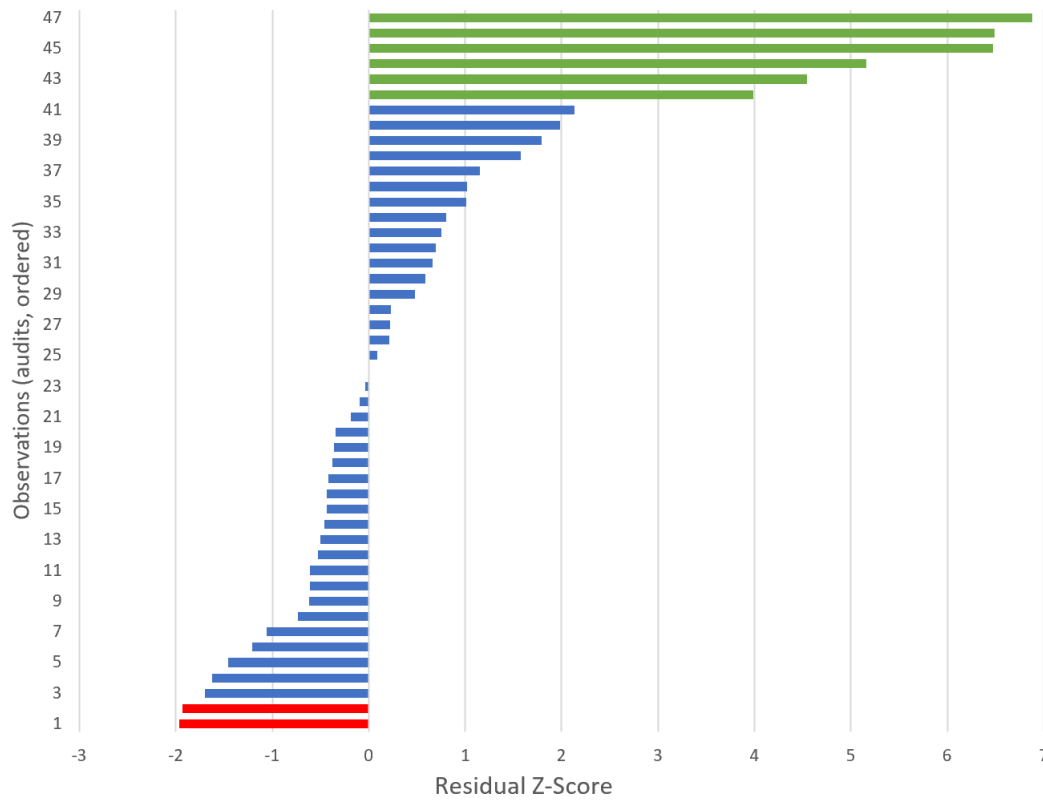
The state administration actions, that is, the degree to which the government, the legislative, etc. dealt with the inspection reports, have a positive influence on the rectification of errors/violations. This relationship is statistically significant with a probability greater than 90%. Individual types of errors/violations were used in the regression analysis in absolute numbers, i.e. as numbers of individual errors and violations in individual reports. This is motivated by the assumption that SAO reports that contained the finding of a greater number of errors/shortcomings are more likely to be rectified. Some of the results are very surprising. The observed effect is negative for the Type 2 error, that is the methodological errors. The relationship is statistically significant at 99%. The rest of the errors and their total numbers are not statistically significant. This is particularly surprising about financial misconduct, where one would expect a strong effect and a strong willingness of the state agencies to rectify past errors. Date, our time variable is negative, a finding statistically significant at 99% certainty – this seems to suggest state agencies are less and less likely to correct past errors. We also found a significant effect of time between reports - giving agencies more time to correct past errors increases the likelihood that they will do so, with a 99% statistical significance.

#### *Regression analysis II – residual analysis*

With the second model, we can also investigate unusual or unexpected cases using the residual analysis. This is an analysis of deviations of individual rectifications of previous errors compared to the expected values as predicted by the model, i.e. whether there were more or fewer rectifications that would correspond to the factors included in the model. Again, we rely on the standard deviation of the residuals to identify these unusual cases. Very extreme cases were identified for positive residuals – there are 6 cases where the residual is more than 3.5 standard deviations from the predicted value (highlighted in green, see Figure 3 below). These are cases where the agencies corrected significantly more errors than average while controlling for all factors included in our regression model. It should be said that the overall level of past error rectification is low, so it is not surprising that some agencies overperformed.

While we didn't identify any extreme cases of negative residuals (cases where state agencies corrected for less than is expected given the factors included in our regression model), in our methodology we propose that the highest negative residuals should be examined further (highlighted in red, see Figure 3 below).

**Figure 3.** Z-Scores of rectified problems (minus – less than expected, plus – more than expected)



**Source:** own adaptation and processing

The postestimation for normality of the residuals of the final modules (see Q-Q plot Figure 13 of the Appendix), for the goodness of fit (see Added-variables plots and leverage plot, Figure 14 and 15), and the multicollinearity and (see VIF scores in Appendix - Table 5). No significant multicollinearity issue has been identified in the sample. The residuals are not normally distributed, but the difference is small and an impact on the results has been eliminated by using robust standard errors to confirm the significance of individual variables.

## DISCUSSION

### ***Selection and evaluation of impact indicators, media coverage and success of correction***

The analysis of the audit conclusions, which identified individual indicators used to assess the impact of the SAO's audit conclusions, revealed several findings that influenced the final form of the methodology. First, the form in which the audit conclusions are drawn up does not have a fixed structure. This often makes it very difficult to determine exactly which of the findings described by the SAO are errors and which are merely 'cosmetic' deficiencies. Second, the audit conclusions contain recommendations on how to address the findings only in some cases, but mostly only describe the shortcomings. *Recommendations* appeared in the audit

conclusions only in 19 % of cases. The insufficient implementation of the recommendations from the PA conclusions (van Zyl, Ramkumar & de Rencio, 2009) is a problem in itself and their absence will only exacerbate the situation. Third, if an audit is followed up by another audit, it does not in all cases consider the previous misconduct. Since follow-up audits are usually not absolutely identical to the original audit, and the SAO is not obliged to verify all the results of previous audits. However, this contributes to the fact that auditees often do not have to make the effort to correct errors because corrective control is not comprehensive and tightly structured.

A clear structure of audit conclusions, as well as of the audits themselves, is an essential prerequisite for the SAO outputs to be measurable and comparable. This will make a significant contribution to obtain the objective data necessary to identify changes made as a result of audit conclusions, as insufficient data is one of the main obstacles to measuring the impact of audit conclusions (see e.g., Pollitt et al., 1999; Morin, 2001, 2014; Kells, 2011). The above findings only confirm the results of academic research (see e.g. Berner and Judge, 2019; Ventura et al., 2023), which point to the difficulties of the state administration in data standardization and the persistent shortcomings in this area. Given the difficulty of identifying the impact of individual audit conclusions, a major improvement that the SAO itself can implement is to conduct follow-ups on all cases with (significant) misconduct found and a clear standardization of procedures in such cases. In this way, the SAO itself can verify whether the previous audit investigation has led to the rectification of the deficiencies of the audited entity. A situation where the SAO's re-inspection finds that there has been no remedy provides room for making the case public and for renewed pressure on the state institutions that should have addressed the misconduct in question.

The analysis of selected impact indicators showed that it is not possible to unambiguously determine the impact of audit conclusions on the basis of *Amendment of the concerned legislation*, as there is no obligation in the Czech legislative process to disclose that the amendment of the law was caused by the activities of SAO. This confirms the insistence of many authors (see e.g., Pollitt et al., 1999; Kells, 2011; Morin, 2014) on the difficulty of identifying changes in public administration based on a specific audit conclusion. There is therefore room for increased transparency in this area as well. The analysis also shows that in the case of *Conflict with legislation* the SAO filed criminal charges in only 5.6% of possible cases, which may undoubtedly lead to less pressure to resolve the misconduct found and thus to a lower impact of audit findings. In 71 cases (70.3% of the cases analyzed) there was a violation of the law, in all of which a criminal complaint could have been filed. On the other hand, it is understandable that filing criminal charges may not always be the most appropriate way for the SAO to respond to the findings. However, the media coverage of only 42% of the inspections carried out suggests that this is probably not an issue of concern to society.

### **Regression results**

Pollitt et al. (1999), Morin (2008) or Persson and Nasstrom (2016) conclude that the factors influencing audit impact are very difficult to identify in practice. Result of the first regression model suggests that the impact of the audit conclusions (reaction of state administration) could

depend on the type of errors found. Data also suggest that over time the impact of the audit conclusions on the state administration action increase. It can be expected that with each “*Methodological errors*” the chances of a response of the state administration decrease, and for “*Instances of missing, incorrectly or insufficiently defined processes errors*”, the chances of the response increase. For all other errors, we have not identified a statistically significant influence. The results thus suggest that the state administration is reluctant or unsure how to address specific types of errors.

The results of the second regression model suggest that the extent to which the state administration addressed the audit findings had a significant impact on the rectification of errors. If the past errors were methodological (Type 2) they were more likely to be corrected. But it must be mentioned that the rectification of previous mistakes has a deteriorating tendency over time. In addition to repeated audits, another way to increase the impact of audit findings could be to create (for example) a legislative framework in which the state administration would be obliged not only to discuss and address the findings, but also to follow up on whether rectifications have actually been made. Other categories than methodological errors were not statistically significant. This is especially surprising for “*financial error*”, where a strong effort for rectification should be expected. The insignificant effect of “*financial error*” might be explained by Blankart (2017) theory, who states that most of the SAO's reports concern uneconomical behavior that took place months or even years ago. The public is no longer interested and there is no pressure on the government to make the remedy.

Our analysis has also identified outliers, which should be a subject of further examination. In the second model, we have found statistically significant outliers but only on the positive side (more than expected number of past errors have been corrected), but no statistically significant negative outliers were identified (in Model 1, one outlier of -3 standard deviations is to be expected is a sample with more than 100 observations). We, however, hope that the potential of this methodology for such identification is clear. In the residual analysis based on Model 1, there was 1 case where the audit did induce a more than 3 standard deviation lower response from the state administration than expected. In the second residual analysis, we have identified two cases just under 2 standard deviations less rectified than expected. Both of these cases can be treated as potentially suspicious and further investigated, as all factors used in the regression analysis in both models were controlled for and don't have any influence on these numbers. And the empirical rule of 3 standard deviations is just that – empirical rule.

## CONCLUSION

The aim of the paper was to develop a methodology that would show how to use statistical methods to evaluate public administration and public policy performance. The analysis carried out allowed to explore the limits and possibilities in this area of it and to identify the factors that affect these performances. A major contribution that this methodology allows is to identify outlying or suspicious cases that (do not) elicit a sufficient response from the public administration. The evaluation of the outputs of government programs is challenging due to the lack of clear output units and the need to derive indicators from performance descriptions.

Identifying and quantifying relevant output indicators that can inform decision-making and provide meaningful insights into the performance and effectiveness of government initiatives is a major challenge. For this reason, in our manuscript we have tried to use indicators that are as general as possible, but still representative, so that it will be possible to apply this methodology, in its adapted form, to other public institutions and to determine whether or not there is a reaction on the part of the entities concerned as a result of the actions of this public institution.

The case study on which this methodology is illustrated is the impact of the audit conclusions of the Supreme Audit Office of the Czech Republic. This case was chosen as there is no generally accepted methodology in academic literature for assessing the impact of audit findings. Such impacts are not defined, and neither are the factors that influence them (see Section 2). The search for and development of new methods for evaluating the impact of SAO performance can therefore contribute not only to improving their activities, but ultimately also to saving public finances and finding new models to increase public administration efficiency. Further, this article sought to answer the following questions:

*Which indicators can be used to objectively assess the impact of the inspection/audit/evaluation findings (of a specific public sector institutions)?*

*What are the impacts of the findings of the inspection/audit/evaluation (of a specific public sector institution)?*

*What factors influence the impact of the findings of the inspection/audit/evaluation (of a specific public sector institution)?*

Through the developed methodology (Section 3) and subsequent empirical analysis (Section 4) it was possible to identify objective indicators of the impact of the SAO's audit findings which are *discussion in the government, discussion in the Parliament, SAO's recommendations in audit conclusion and filing of a criminal complaint by SAO in case of conflict with legislation. Amendment of the concerned legislation* did not prove to be a fully objective indicator, as it is not possible to verify the reason for the change in the Czech legislative environment.

The analysis of the state administration's response points to shortcomings whose elimination can significantly contribute to a higher impact of the audit findings, for example, audit conclusions, contain recommendations only in 19 % of cases, but mostly are audit conclusions focused to only describe deficiencies. Regarding the factors influencing the impact of audit findings, the reaction of administration could depend on the type of errors found. It can be expected that with each "Methodological errors" the chances of a response of the state administration decrease, and for "Instances of missing, incorrectly or insufficiently defined processes errors", the chances of the response increase. For all other errors, we have not identified a statistically significant influence. Data also suggest that over time the impact of the audit conclusions on the state administration action increase. The results suggest that the extent to which the state administration dealt with the report had a significant effect on the rectification of the errors. Although the government responds to audit findings and issued a resolution in 99% of cases, the lack of a follow-up system to monitor remediation leads to uncertainty as to whether the error will be or has been rectified. This is reflected in the low



percentage of errors corrected, which is only 17.18% for the monitored dataset. The research has shown that to the improvement of the functioning of control authorities could contribute e.g. granting the right of legislative initiative to the SAO, clear and consistent records of errors found during inspections and the imposition of the SAO's obligation (for follow-up inspections) to check the rectification of errors found during past inspections.

The research has further shown that the methods of residual analysis can be used to identify outliers, possibly suspicious cases which, in this case study, don't induce a sufficient response from the state administration (regression analysis I) or which don't get rectified (regression analysis II) given the factors considered in the regression. The residuals represent what is left after controlling for these factors and what is left is not influenced by this multifactorial noise. While there very possibly are real world reasons for some of the high negative residual values, there is a good chance that negligence or ill intent also played a role. As such, this methodology allows for clearer examination of these cases individually. Moreover, the automation and implementation of this process in the control mechanisms of the state administration could significantly contribute to the efficiency of the whole system.

A possible shortcoming of this study might be the use of "black box" evaluation strategy which focuses primarily on evaluating the inputs and outputs of a system or process without a deeper examination of its internal mechanisms or functioning. We understand that this method has its limitations. That it may provide insights into general trends or correlations but may not always uncover the nuanced reasons behind decision-making processes or the effectiveness of government programs. Nevertheless, we believe that the quantitative approach we have used will allow us to identify those individual cases that require individual and qualitative attention, where it will be necessary to reveal the subtle reasons for decision-making processes or the effectiveness or ineffectiveness of government programs. In today's digital age, it is possible to automate this methodology so that stakeholders or policymakers can only benefit from the results, but a necessary prerequisite for this is data standardization.

## REFERENCES

- Barzelay, M. (1997). Central audit institutions and performance auditing: A comparative analysis of organizational strategies in the OECD. *Governance*, 10(3), 235-260.  
<https://doi.org/10.1111/0952-1895.411997041>
- Berner, R., Judge, K. (2019). The data standardization challenge. Forthcoming in *Systemic Risk in the Financial Sector: Ten Years After the Great Crash*, edited by Douglas W. Arner, Emiliios Avgouleas, Danny Busch and Steven L. Schwarcz (CIGI Press), Columbia Law and Economics Working Paper, (598). Available at SSRN: <https://ssrn.com/abstract=3323719>
- Blankart, C. B. (2017). Öffentliche Finanzen in der Demokratie: Eine Einführung in die Finanzwissenschaft, vol. 9, *Vahlens Handbücher der Wirtschafts- und Sozialwissenschaften* (München: Verlag Franz Vahlen, <https://doi.org/10.15358/9783800653485>).

- Blume, L., Voigt, S. (2007). Supreme Audit Institutions: Supremely Superfluous? A Cross Country Assessment. *ICER - International Centre for Economic Research*, ICER Working Papers. DOI: 10.2139/ssrn.965434
- Bouckaert, G., Halligan, J. (2008). Comparing Performance across Public Sectors. In: Van Dooren, W., Van de Walle, S. (eds) *Performance Information in the Public Sector. Governance and Public Management Series*. Palgrave Macmillan, London. [https://doi.org/10.1007/978-1-137-10541-7\\_6](https://doi.org/10.1007/978-1-137-10541-7_6)
- Corbin, J. M., Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, 13(1), 3-21. DOI: 10.1007/bf00988593
- Cornes, R., Sandler, T. (1996). *The theory of externalities, public goods, and club goods*. Cambridge University Press.
- García, M. (2005). Public control: a general view, in M. Crespo (ed.), *Public Expenditure Control in Europe. Coordinating Audit Functions in the European Union (3-29)*, United Kingdom: Edward Elgar.
- Hassan, H., Kamsuriah, A., Sallehudin, H. (2020). Diagnosing the Issues and Challenges in Data Integration Implementation in Public Sector. Retrieved from: [http://insightsociety.org/ojaseit/index.php/ijaseit/article/view/10271/pdf\\_1365](http://insightsociety.org/ojaseit/index.php/ijaseit/article/view/10271/pdf_1365)
- Hongwei, Z., Yang, W., L., Rosenthal, A. (2016). Data Standards Challenges for Interoperable and Quality Data. *Journal of Data and Information Quality*, DOI: 10.1145/2903723
- INTOSAI (2019). International Organization of Supreme Audit Institutions. *INTOSAI-P 1 – Declaration of Lima*. Retrieved December 15, 2020, from <https://www.issai.org/pronouncements/intosai-p-1-the-lima-declaration/>.
- Kells, S. (2011). The seven deadly sins of performance auditing: implications for monitoring public audit institutions. *Australian Accounting*, 21(4), 383-396. <https://doi.org/10.1111/j.1835-2561.2011.00150.x>
- Lonsdale, J. (2000). Developments in value-for-money audit methods: impacts and implications, *International Review of Administrative Sciences*, Vol. 66, N. 1, 73-89. DOI: 10.1177/0020852300661007
- Lynch, T. D., Day, S. E. (1996). Public sector performance measurement. *Public Administration Quarterly*, 404-419.
- McAuliffe, B. (2019). Challenges Facing Technology Standardization in the Age of Digital Transformation. Doi: 10.4018/978-1-5225-9008-8.CH003
- McGowan, R. P., Polster, T. H. (1985). Impact of productivity measurement systems on municipal performance. *Review of Policy Research*, 4(3), 532-540. <https://doi.org/10.1111/j.1541-1338.1985.tb00253.x>
- Milgrom, T., Schwartz, R. (2008). Israel's auditor as policy change agent: The case

- of water policy. *International Journal of Public Administration*, 31, 862-877.  
<https://doi.org/10.1080/01900690701696178>
- Morin, D. (2001). Influence of Value for Money Audit on Public Administrations: Looking Beyond Appearances", *Financial Accountability & Management*, vol. 17, no. 2, pp. 99-117.  
<https://doi.org/10.1111/1468-0408.00123>
- Morin, D. (2008). "Auditors general's universe revisited", *Managerial Auditing Journal*, vol. 23, no. 7, pp. 697-720. DOI: 10.1108/02686900810890652
- Morin, D. (2014). Auditors general's impact on administrations: A pan-Canadian study (2001-2011). *Managerial Auditing Journal*, 29(5), 395-426. <https://doi.org/10.1108/MAJ-10-2013-0948>
- Niskanen, W. (1971); *Bureaucracy and Representative Government*, Chicago: Aldine-Atherton.  
<https://doi.org/10.4324/9781315081878>
- Persson, E., Nasstrom, M. (2016). Performance Audits: A study of the factors affecting audit impact in Sweden. DiVA, id: diva2:944052.
- Peters, B. G. (1998). Managing horizontal government. *Public administration*, 76(2), 295-311.  
<https://doi.org/10.1111/1467-9299.00102>
- Pollitt, C., Girre, X., Lonsdale, J., Mul, R., Summa, H., Waerness, M. (1999). Performance or Compliance? Performance Audit and Public Management in Five Countries. *Oxford: Oxford University Press*. DOI:10.1093/acprof:oso/9780198296003.001.0001
- Pollitt, C. (2003). Performance audit in western Europe: trends and choices. *Critical Perspectives on Accounting*, 14, 157-170. <https://doi.org/10.1006/cpac.2002.0521>
- Power, M. (1999). The Audit Society: Rituals of Verification. *Oxford: Oxford University Press*. DOI:10.1093/acprof:oso/9780198296034.001.0001
- Raudla, R., Taro, K., Agu, C., Douglas, J.W. (2016). The Impact of Performance Audit on Public Sector Organizations: The Case of Estonia, *Public Organization Review*, 16(2), 1-17. DOI: 10.1007/s11115-015-0308-0
- Reichborn-Kjennerud, K. (2014). Performance audit and the importance of the public debate. *Evaluation*, 20, 368–385. <https://doi.org/10.1177/1356389014539869>
- Reichborn-Kjennerud, K., Johnsen, Å. (2015). Performance Audits and Supreme Audit Institutions' Impact on Public Administration: The Case of the Office of the Auditor General in Norway. *Administration and Society*. <https://doi.org/10.1177/0095399715623315>
- Schelker, M. (2008). Auditors and Corporate Governance: Evidence from the Public Sector. *Center for Research in Economics, Management and the Arts (CREMA), CREMA Working Paper Series*. 66. DOI: 10.1111/kykl.12021.
- Van Looche, E., Put, V. (2011). The impact of performance audits: A review of the existing evidence. In J. Lonsdale, P. Wilkins, & T. Ling (Eds.), *Performance auditing: Contributing to accountability in democratic government* (pp. 175- 208). Cheltenham, UK: Edward Elgar.

van Zyl, A., Ramkumar, V., de Rencio, P. (2009). Responding to the challenges of Supreme Audit Institutions: Can legislatures and civil society help? *U4Issue*, 1. Retrieved from <http://www.u4.no/>

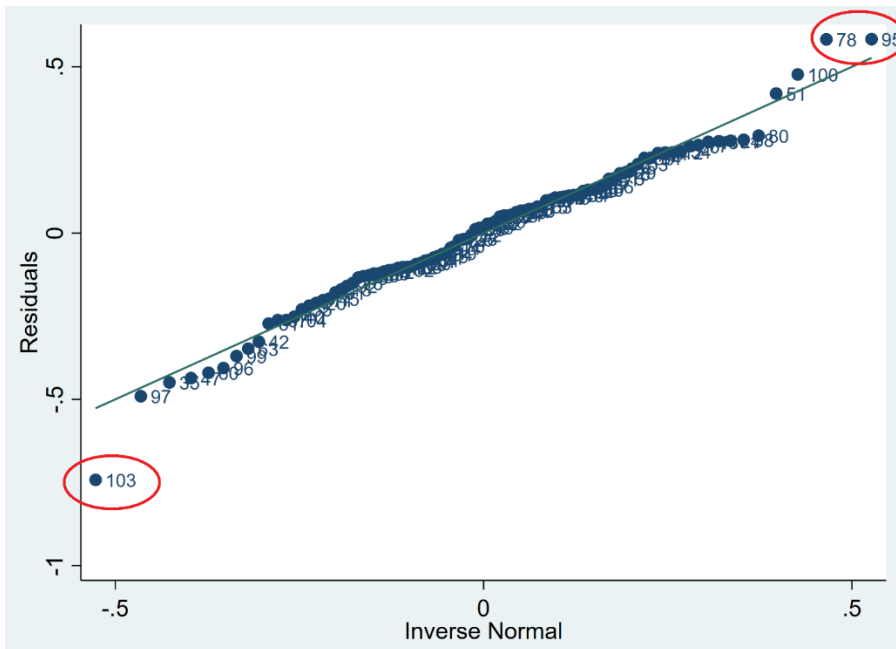
Vedung, E. (2017). *Public policy and program evaluation*. Routledge. ISBN 9780765806871

Talbot, C., Wiggan, J. (2010). The public value of the National Audit Office. *International Journal of Public Sector Management*. DOI:10.1108/09513551011012321

Ventura, R., Silva, A., Cunha, A., Oliveira, H. (2023). Accounting Standardization in the Municipality of Elvas: stages and challenges. *Dutch Journal of Finance and Management*. 5. 21104. 10.55267/djfm/13354. DOI: 10.55267/djfm/13354

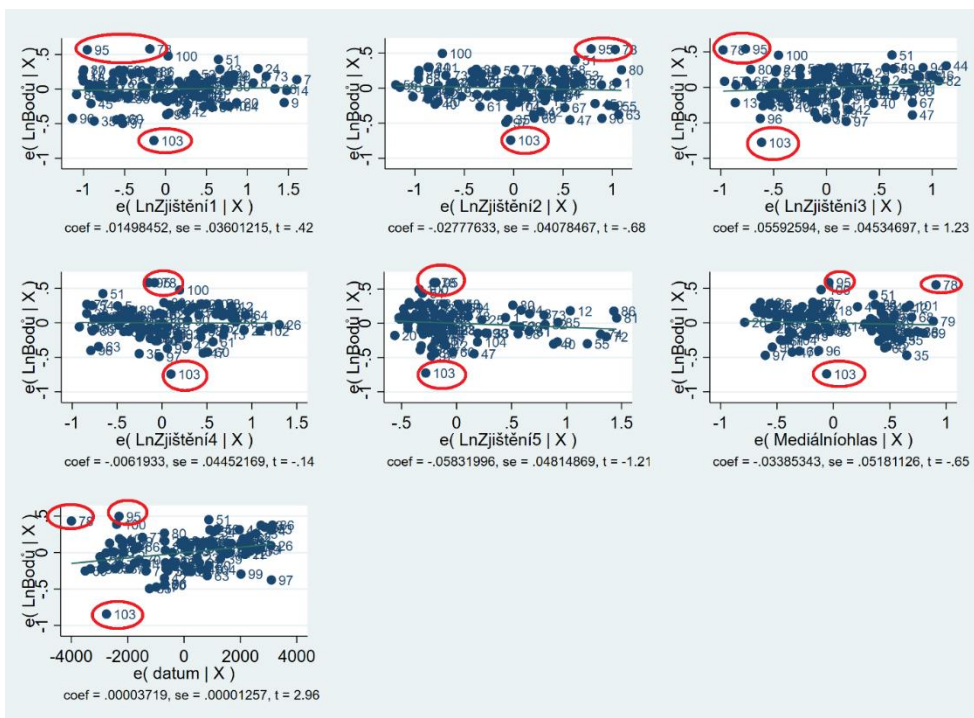
**APPENDIX**

**Figure 4 – Model 1 Q-Q plot of the residuals prior to outlier removal.**



Source: own adaptation and processing

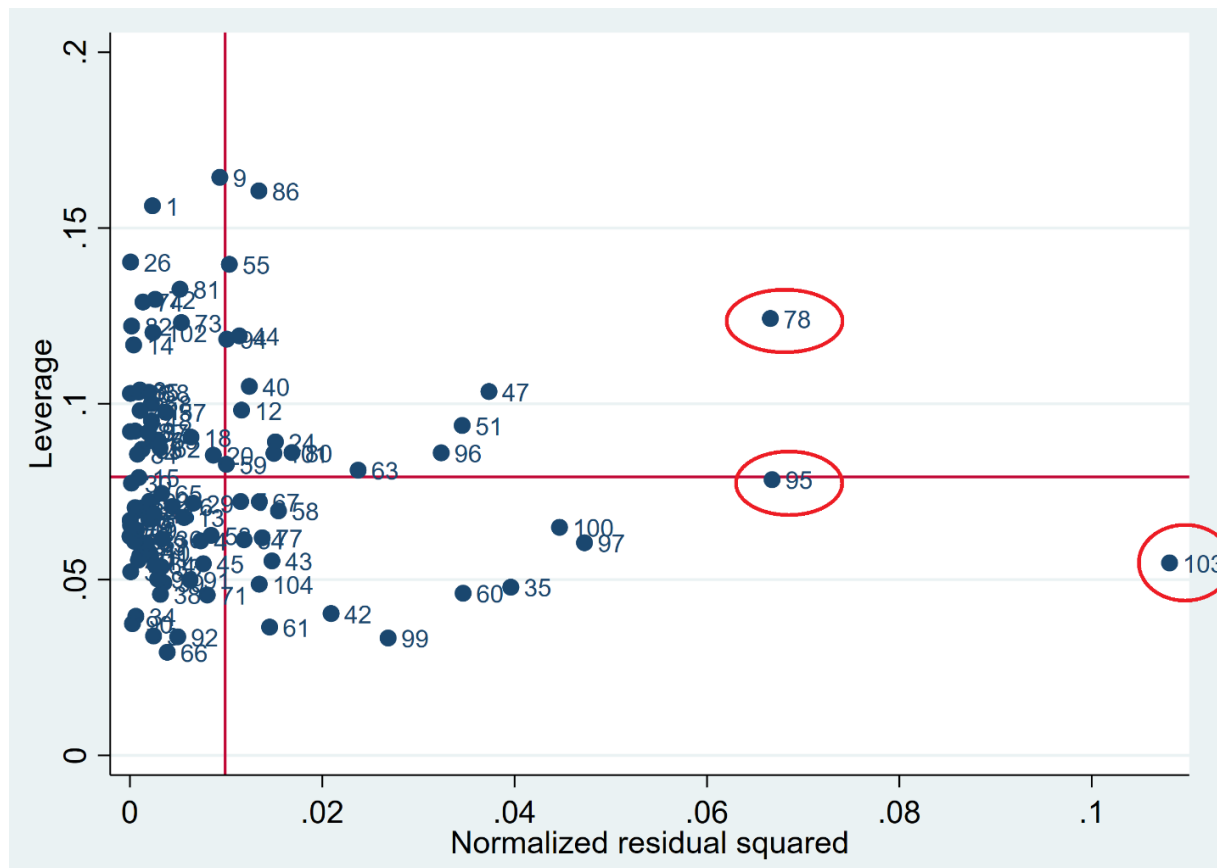
**Figure 5 – Model 1 added-variable plots prior to outlier removal.**



Source: own adaptation

and processing

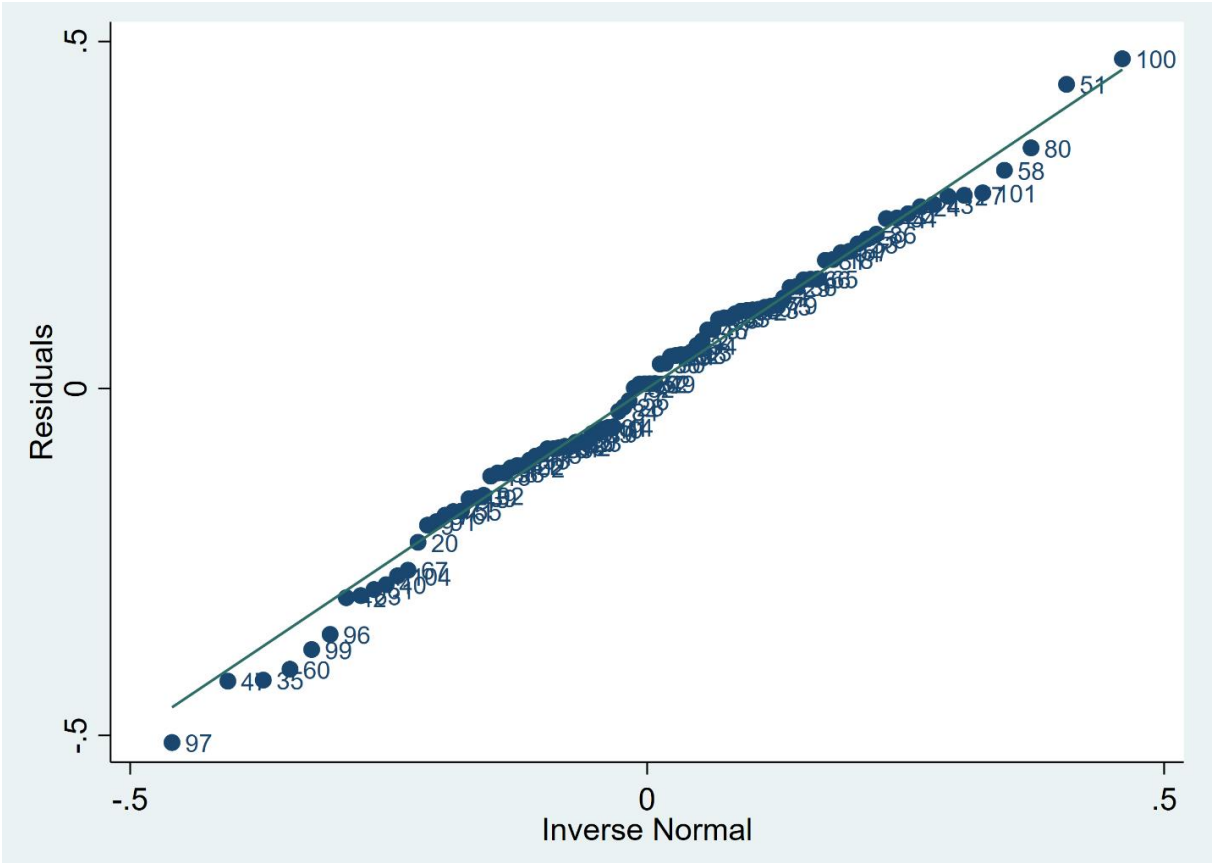
**Figure 6 – Module 1 residual leverage plot prior to outlier removal.**



Source: own adaptation and processing

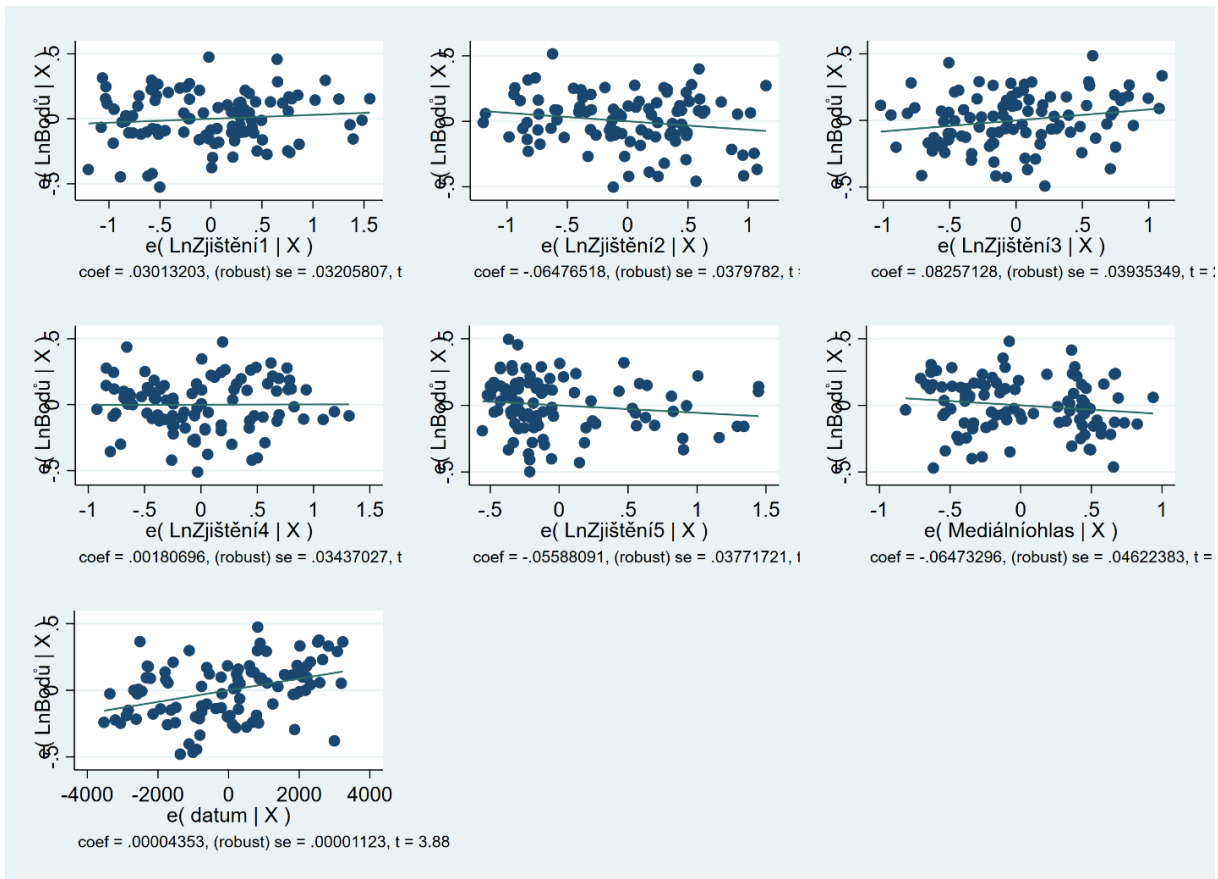
We remove observations 78, 95 and 103, as they don't follow the normal distribution of residuals, with a stark contrast of the rest of the observations, as per the Q-Q plot above. The added variable plots (also above, show the position of these removed observations in relation to individual variables. The leverage plot (Figure 6), shows the sizes or the residuals and their influence (leverage) on the position of the regression line. All three plots above warrant removal of these observations.

Figure 7 – Model 1 Q-Q plot of the residuals after outlier removal.



Source: own adaptation and processing

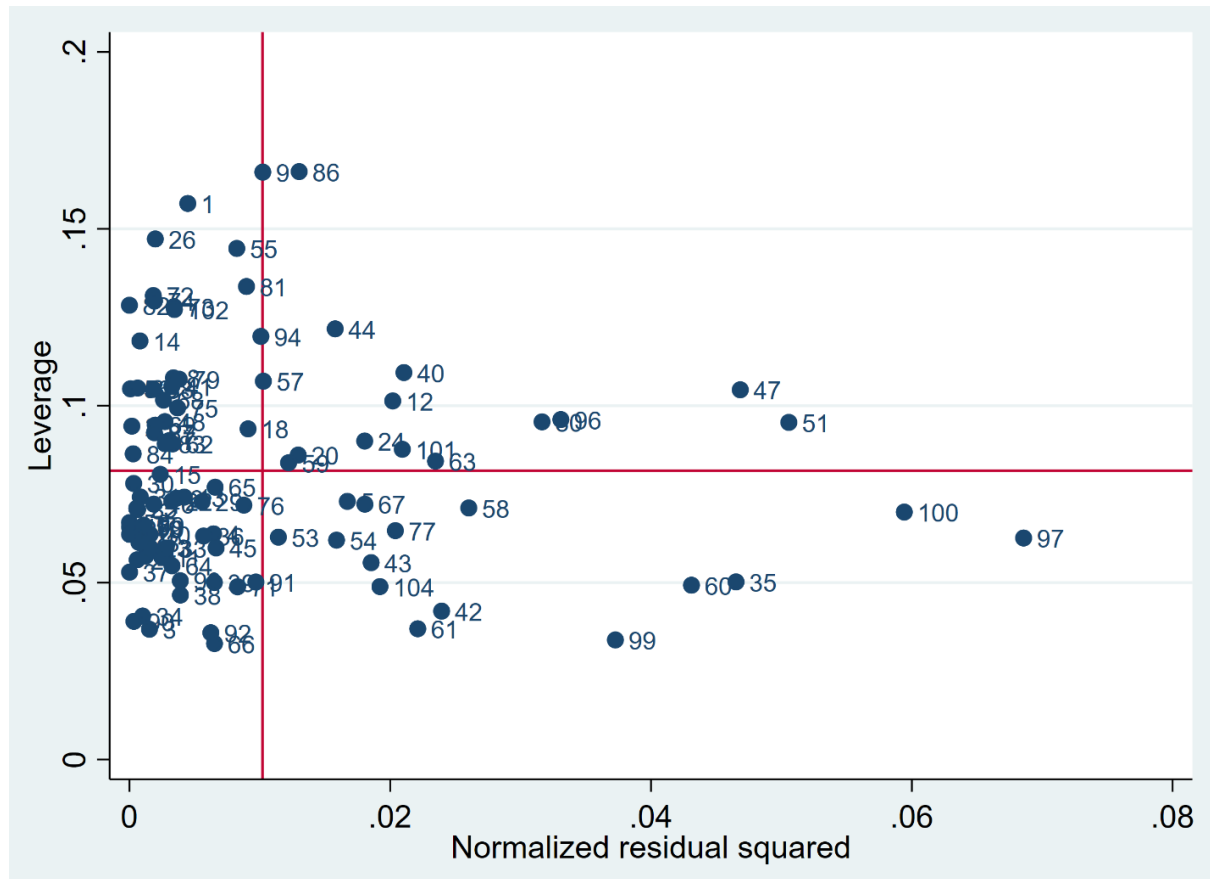
**Figure 8** – Model 1 added-variable plots after outlier removal.



Source: own adaptation and processing



**Figure 9** – Module 1 residual leverage plot after outlier removal.



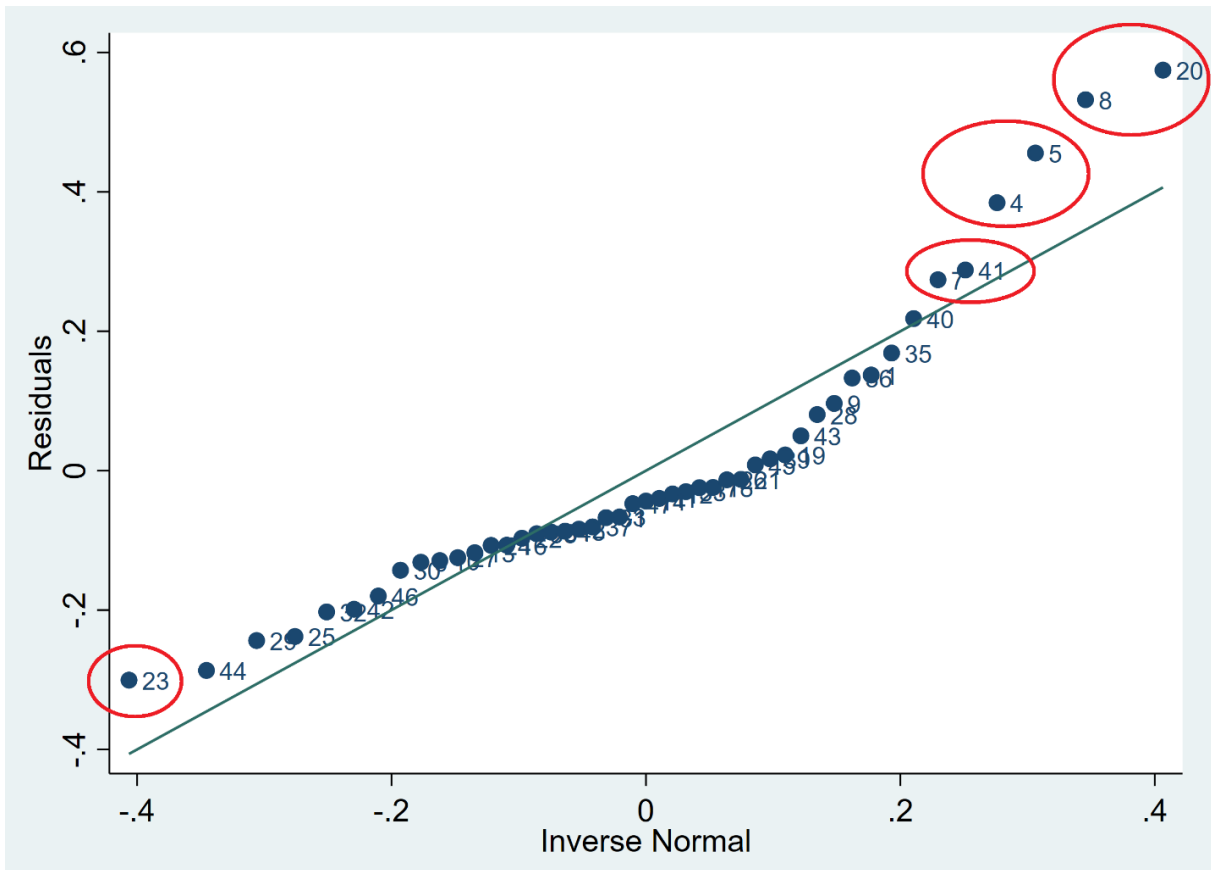
Source: own adaptation and processing

**Table 4** – Model 1 VIF scores (after outlier removal)

<i>Variable</i>	<i>VIF</i>	<i>1/VIF</i>
Intype1	1.20	0.8331
Intype2	1.19	0.8425
Intype3	1.23	0.8152
Intype4	1.28	0.7802
Intype5	1.12	0.8959
date	1.22	0.8209
media	1.22	0.8179
<b>Mean VIF</b>	<b>1.21</b>	

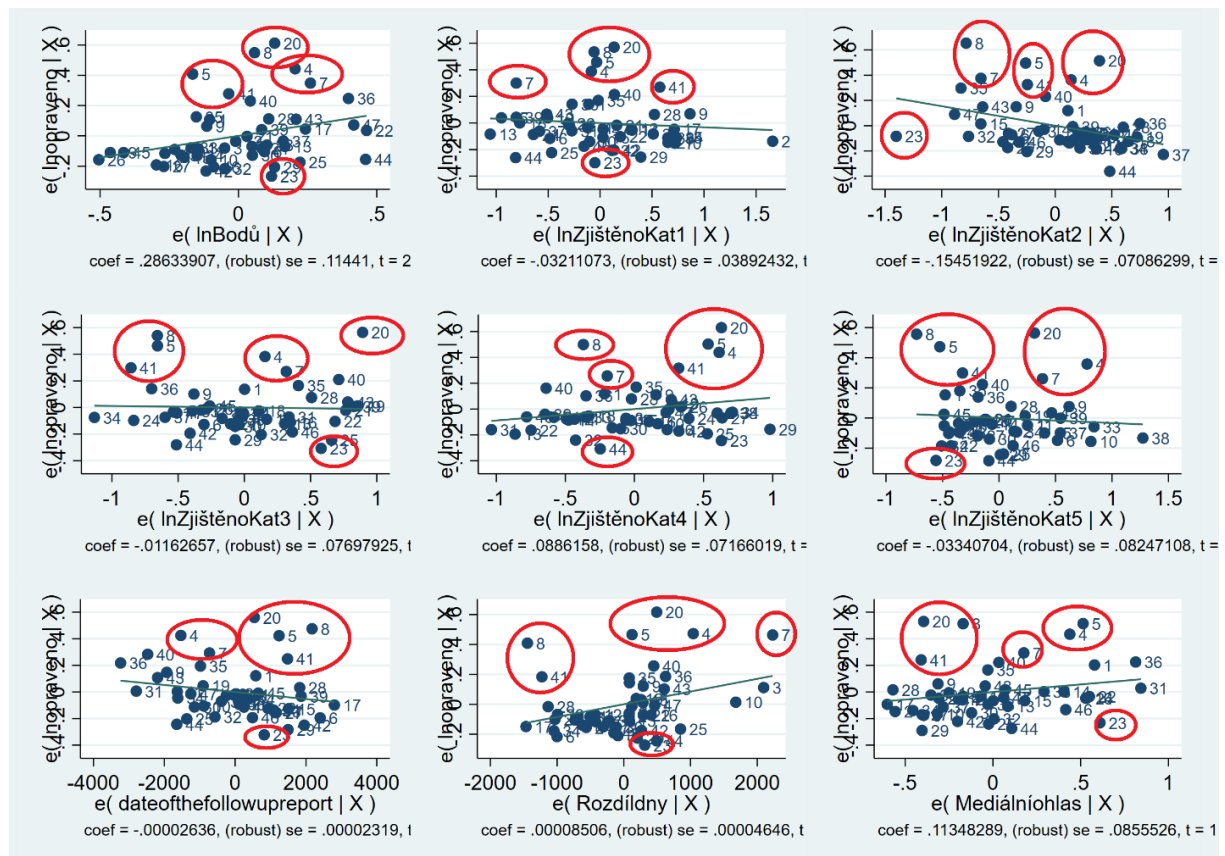
Source: own adaptation and processing

Figure 10– Model 2 Q-Q plot of the residuals prior to outlier removal.



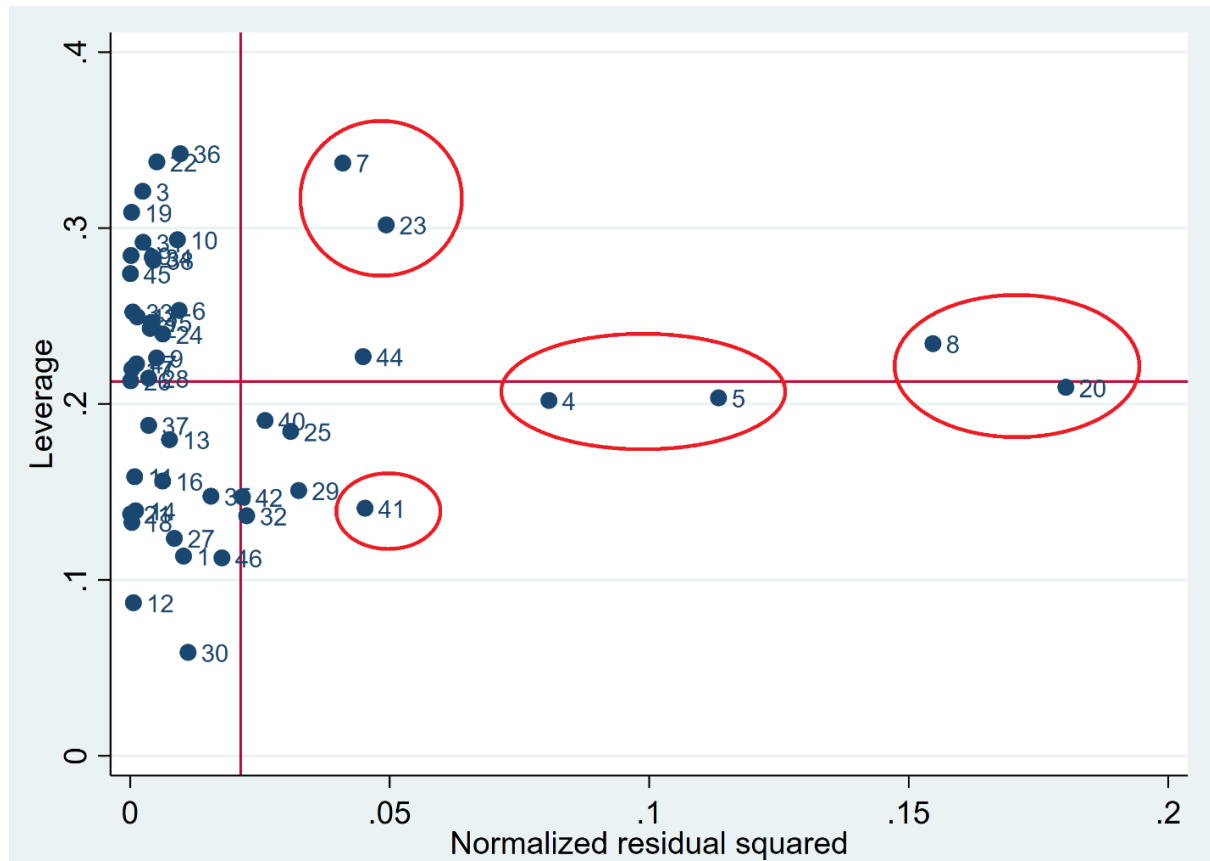
Source: own adaptation and processing

Figure 11 – Model 2 added-variable plots prior to outlier removal.



Source: own adaptation and processing

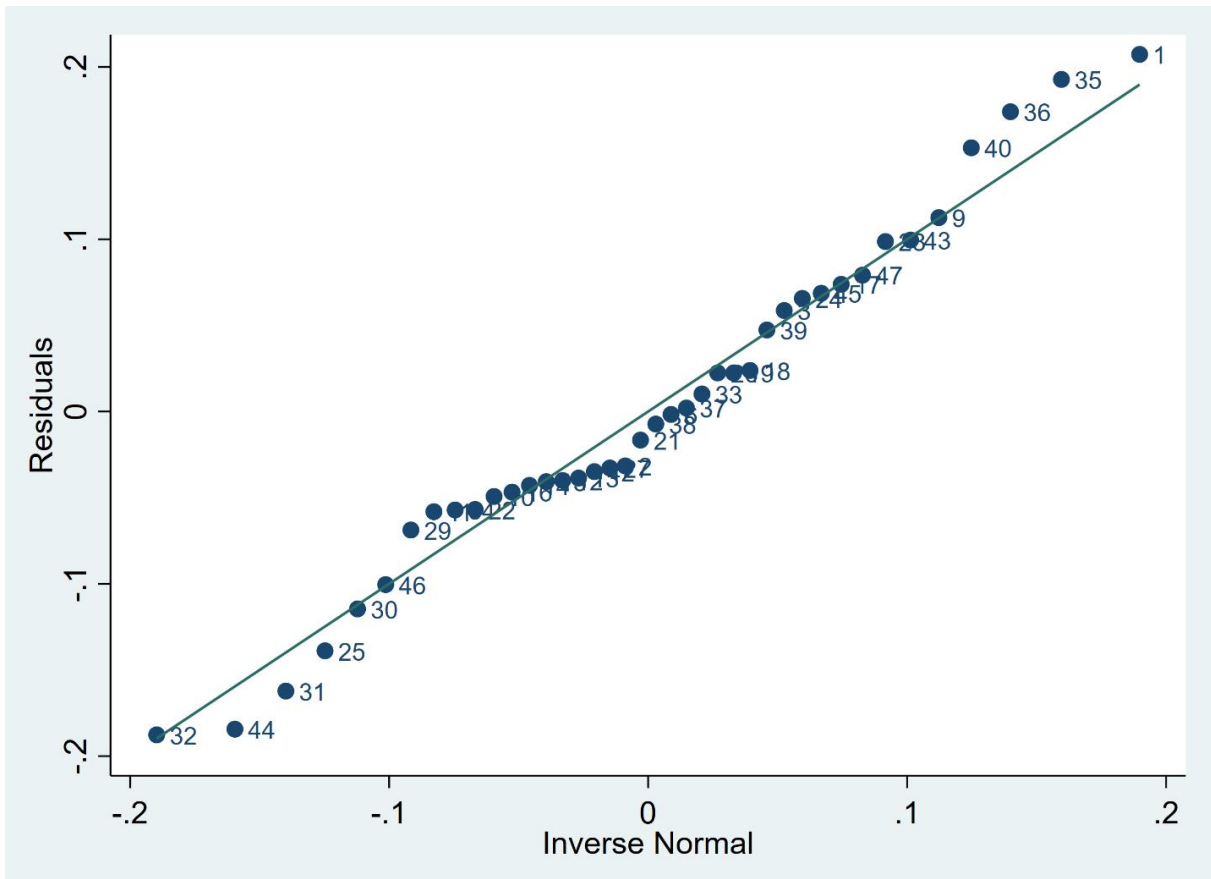
Figure 12– Module 2 residual leverage plot prior to outlier removal



Source: own adaptation and processing

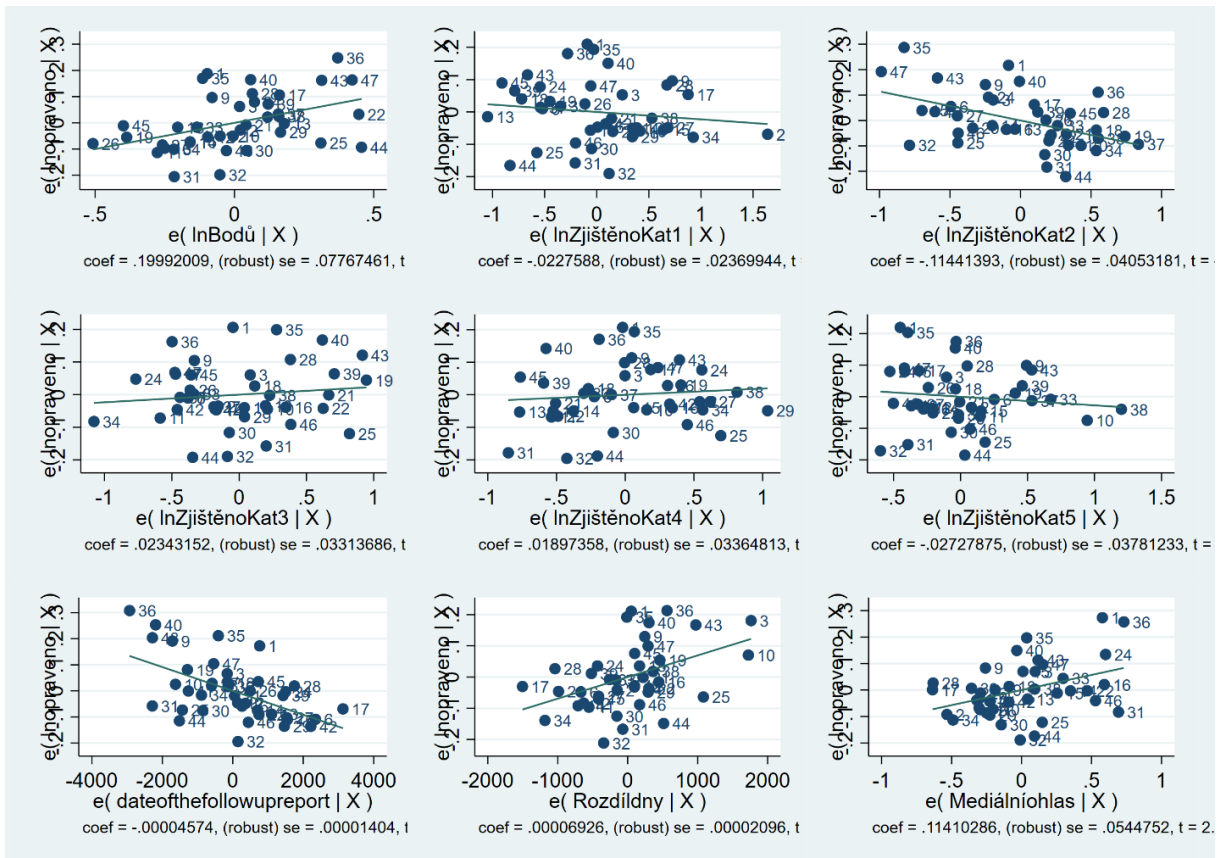
Observations 7, 8, 20, 23 are consistently being confirmed as outliers in all three diagnostic tools. The same applies, to a lesser degree, to observations 4, 5. All these observations are removed from the sample. Observation 41 was also removed, as it showed extreme deviation from the Q-Q plot and high leverage in the leverage plot, but only after the above mentioned observations were removed from the model.

Figure 13 – Model 2 Q-Q plot of the residuals after outlier removal



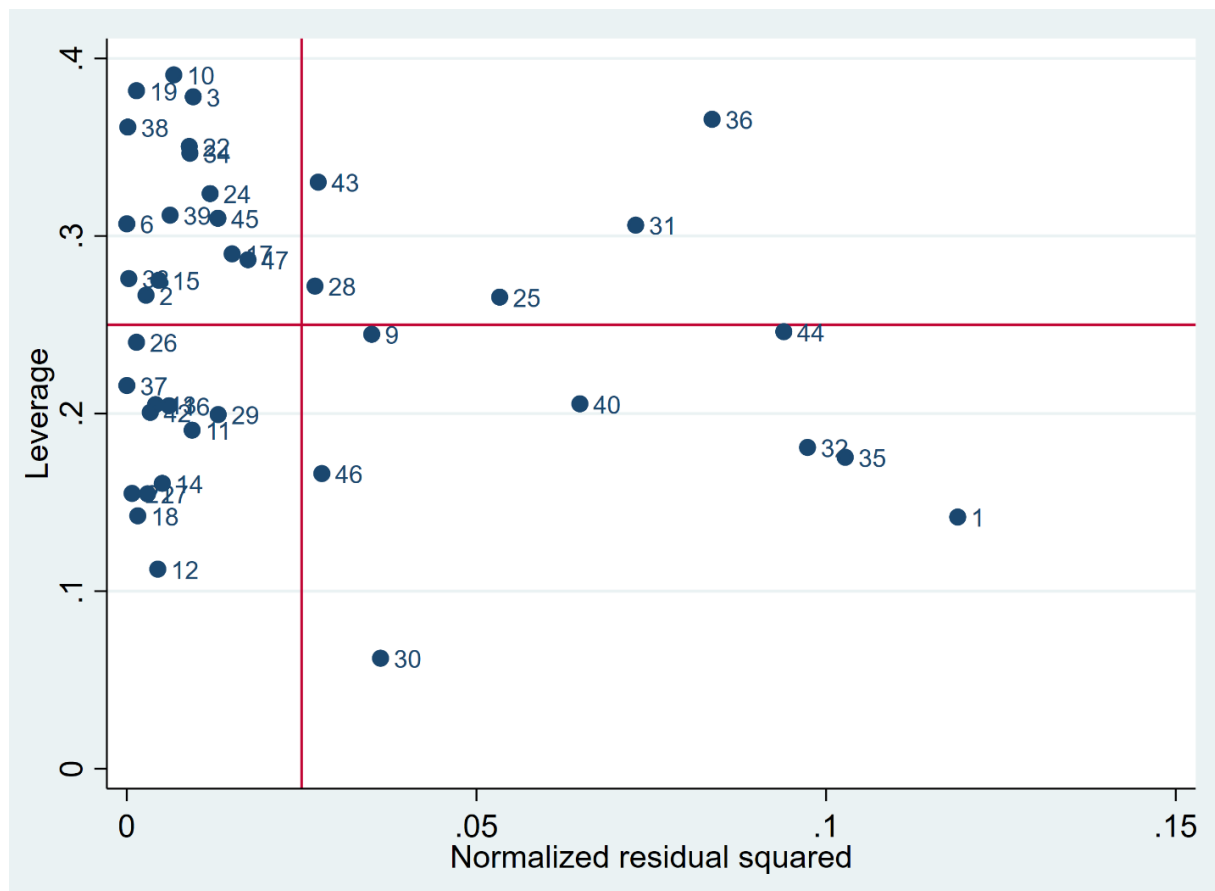
Source: own adaptation and processing

Figure 14 – Model 2 added-variable plots prior to outlier removal



Source: own adaptation and processing

Figure 15 – Module 2 residual leverage plot after outlier removal



Source: own adaptation and processing

Table 5 – VIF scores

<b>Model 2</b>		
<i>Variable</i>	<i>VIF</i>	<i>1/VIF</i>
Intype1	1.40	0.7127
Intype2	1.51	0.6628
Intype3	1.37	0.7276
Intype4	1.46	0.6869
Intype5	1.76	0.5684
date	1.73	0.5765
datediff	2.20	0.4543
Inactionsa	1.14	0.8736
media	1.54	0.6507
<b>Mean VIF</b>	<b>1.57</b>	

Source: own adaptation and processing

**Author contributions****Conceptualization, Vilem Cap and Marek Feurich.**

Ideas; formulation or evolution of overarching research goals and aims.

**Methodology, Vilem Cap, Martin Lukavec and Marek Feurich.**

Development or design of methodology; creation of models.

**Validation, Vilem Cap, Martin Lukavec and Marek Feurich.**

Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.

**Formal analysis, Martin Lukavec.**

Application of statistical, mathematical, computational, or other formal techniques to analyse or synthesize study data.

**Investigation, Vilem Cap.**

Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.

**Resources, Vilem Cap.**

Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.

**Data curation, Vilem Cap and Martin Lukavec.**

Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use.

**Writing—original draft, Vilem Cap and Marek Feurich.**

Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).

**Writing—review and editing, Vilem Cap and Marek Feurich.**

Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.

**Visualization, Martin Lukavec.**

Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.

**Supervision, Vilem Cap.**



Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.

**Project administration, Vilem Cap.**

Management and coordination responsibility for the research activity planning and execution.

**Funding acquisition, Vilem Cap.**

Acquisition of the financial support for the project leading to this publication.

**All authors have read and agreed to the published version of the manuscript.**