UKRAINIAN CATHOLIC UNIVERSITY

FACULTY OF APPLIED SCIENCES

BUSINESS ANALYTICS & COMPUTER SCIENCE PROGRAMMES

Success predictors

Econometrics final interim report

Authors:
Oleksandra SYMKO
Vitalii SHARAIEV

30 April 2024

Contents

1 Introduction

During the pandemic and amidst a full-scale war, a pressing concern has emerged regarding the education and prospects of Ukrainian graduates. As a result of Covid-19, most educational institutions shifted to online learning, using new and innovative methods to teach that had previously seemed impossible.

Furthermore, the war prompted nearly 2.4 million individuals to relocate to other countries, from which they have not returned. It is also vital to recognize that many of these individuals were teenagers, who are pivotal to shaping the nation's future and enhancing the educational system in Ukraine.

Against the backdrop of all these events, due to the impossibility of adhering to the old rules of the ZNO, the Ministry of Education changed its policy and introduced NMT, where pupils can take the required subjects in one day. The main reason for this change is to ensure people's safety and make the process as efficient as possible, given the realities of life in Ukraine.

In this context, the Ukrainian language, being the most critical and compulsory subject, was closely monitored. The Ukrainian Center for Education Quality notes that in 2022 more than 63% of applicants received less than 160 points on a scale of 100-200 in each of the test blocks. This statistic underscores the need to assess how the performance of graduates has changed, taking into account the policy change and other factors, thus our team was tasked with this investigation.

2 Data description and analysis

2.1 Dataset Overview

The data used was presented on the website of the Ukrainian Center for Education Quality: https://zno.testportal.com.ua/opendata. The main study years affected by the above challenges are the years of ZNO 2020 and 2021, as well as the years of NMT 2022 and 2023, which we combined into one dataset. This combination resulted in a pooled cross-sectional dataset.

In the section of data analysis, it became necessary to exclude several columns that were not pertinent to the research. These included the district and city of the participant's residence, the name and type of the participant's educational institution, the district and city of the institution's location, the governing authority of the educational institution, the district and city of the temporary examination center, the scaled test scores from various blocks, and the dates of the tests.

Additionally, due to the unique nature of each examination, Geography was omitted as a subject from the study. This was because graduates were not given the option to take this subject in the years 2022 and 2023.

To enhance our analysis, we also introduced a dummy variable named 'Urban' to signify whether a graduate is from an urban or rural area, and a variable 'TestType' to identify which type of test the participant took in a given year. GitHub.

Consequently, the processed data comprised variables such as:

- **Birth** - year of birth of the participant.
- **SexType** - =1 if male.
- **SchoolRegion** - school location of the participant.

- **Urban** - =1 if from city.
- **TestType** - ZNO2020, ZNO2021, NMT2022, NMT2023.

The following columns include scores for different subjects, each ranging from 100 to 200 points. Those who received a zero as they failed to satisfy the passing standard, however, were eliminated. None values were replaced with the mean score observed in each column to simplify future analysis and maintain the accuracy of the results.

- **Ukr** - Ukrainian language scores.
- **Hist** - History scores.
- **Math** - Math scores.
- **Phys** - Physics scores.
- **Chem** - Chemistry scores.
- **Bio** - Biology scores.
- **Eng** - English language scores.
- **Fra** - French language scores.
- **Deu** - German language scores.
- **Spa** - Spanish language scores.
- **TestStatus** - enrolled, did not appear, annulled, did not pass the threshold.

2.2 Exploratory Data Analysis

In this section, our objective is to include a visual representation of the score distributions. Additionally, we will present an overview of the annual participant numbers in testing and analyze their final status.

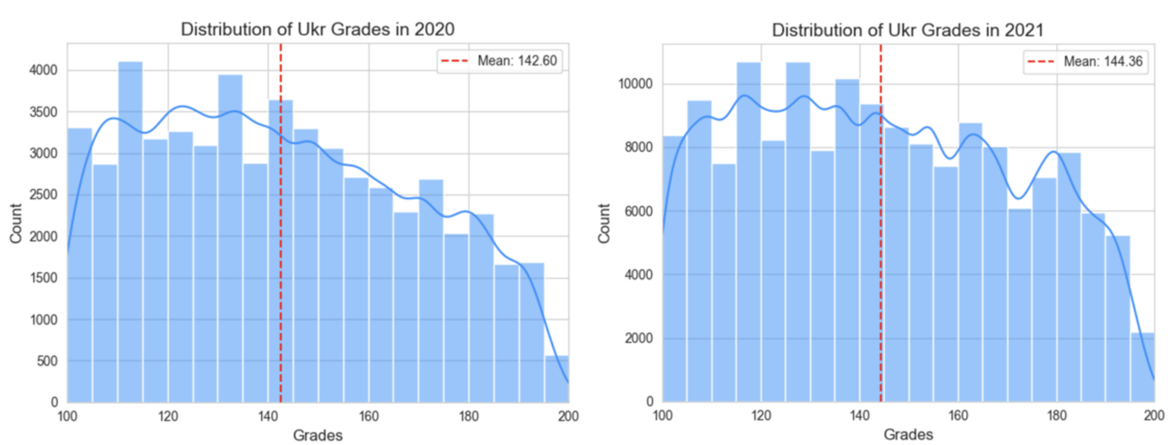

Figure 1: Distribution of scores in the Ukrainian language in 2020 and 2021.

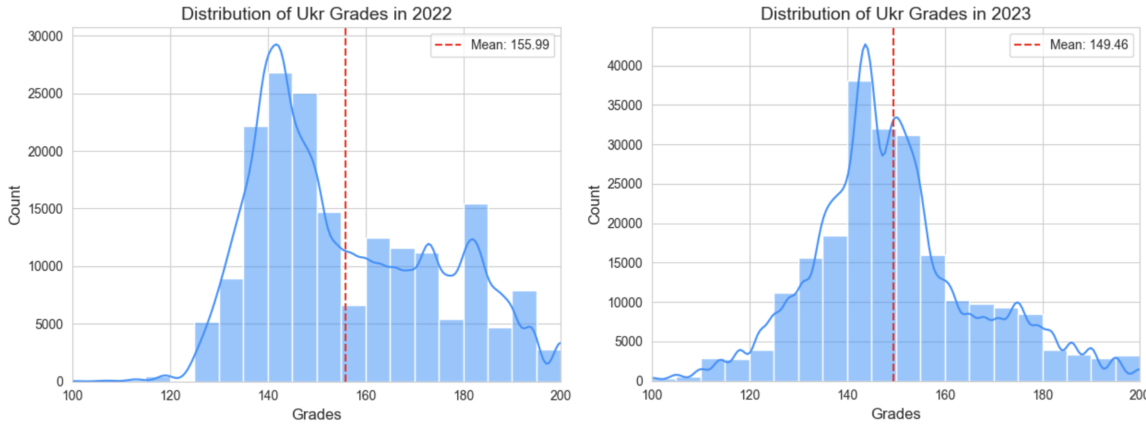

Figure 2: Distribution of scores in the Ukrainian language in 2022 and 2023.

The next table provides a detailed breakdown of participant engagement by year, detailing the total number of individuals registered for testing. It outlines those who were successfully enrolled, the absentees, the candidates whose tests were annulled, and those who did not meet the required passing threshold.

Ukr	Enlisted	Did not appear	Annulled	Did not pass the threshold	Σ
ZNO2020	251,929	95,071	29	22,525	369,554
ZNO2021	204,768	28,070	68	17,480	250,386
NMT2022	214,074	19,997	33	None	234,104
NMT2023	267,871	20,716	91	257	288,935

Table 1: Performance Status

The year 2022 is distinct from the others, as it was a pivotal year in the education field of Ukraine. The test was held on one day, so being present for the test was the same as taking the Ukrainian language exam. Also, our data is not as comprehensive, so we can't determine the number of participants who did not manage to pass the threshold.

3 Methodology Explanation

3.1 Methods Overview

In our study, we aim to explore and understand the factors influencing graduation rates in the Ukrainian language and to identify deviations in grades. Our main task is to study how the performance of graduates has changed after the introduction of NMT, which marks a significant policy change due to situation in Ukraine.

We will employ two main approaches. The first approach is just a linear regression model with interaction variables. It will help us to explore whether the NMT impacts various explanatory variables compared to ZNO. In particular, this model will only show the relationship between the introduction of NMT in Ukraine and other explanatory variables. However, it does not capture the actual deviations attributable to the policy change. Therefore, our second approach will utilize the Difference-in-Differences method, which is more suited to identifying the direct impacts of the policy shift.

3.2 DID

Given the sequential nature of the events ZNO2020, ZNO2021, NMT2022, NMT2023 and the fact that each participant only took part in the exam once, it was necessary to synthetically duplicate our data to create comparable groups across different years for DID.

To accomplish this, we need to create four groups. The first and second groups are our initial data, so they are unchanged. The third and fourth groups were synthetically generated. Firstly, we calculated the needed sample size using implemented function. In simple terms, it figures out how many people we need to include in our study to be confident that our results accurately represent the larger population, taking into account factors like the population size, desired precision (margin of error), and level of confidence. In ZNO2020, where the initial number of participants was 43,955, we obtained a random sample of 381 participants. Similarly, in ZNO2021, with an initial count of 158,755 participants, we derived a random sample of 384 individuals.

It's worth noting that these initial participant counts may not align precisely with the figures in the initial table of all participants (Performance Status) due to the data processing. Additionally, participants who passed ZNO in russian language were excluded. The initial table provides clear, unaltered data to portray the overall picture.

So, 4 main groups:

1. **Treatment group before the treatment**: ZNO participants with a dummy variable indicating 'after' set to 0 and 'treatment' set to 1.

2. **Treatment group after the treatment**: NMT participants with a dummy variable indicating 'after' set to 1 and 'treatment' set to 1.

3. **Control group before the treatment**: ZNO participants with a dummy variable indicating 'after' set to 0 and 'treatment' set to 0.

4. **Control group after the treatment**: ZNO participants with a dummy variable indicating 'after' set to 1 and 'treatment' set to 0.

By maintaining continuity with the ZNO during the years of NMT, our data is now appropriately prepared for the application of the DID method.

3.3 Choice of variables for the models

To evaluate the influence of the NMT policy on Ukrainian language graduation rates and its deviation, we carefully selected and analyzed variables for our models. Some of the same variables were used in both cases for the same purpose:

- **Birth** - year of birth of the participant.

- **Urban** - =1 if from an urban or rural area.

- **SexType** - =1 if male.

- **Math** - instead of including scores from all subjects, we chose math as a proxy for unmeasured intellectual abilities, hypothesizing that mathematical proficiency might reflect broader skills relevant to the dependent variable. This choice is intended to avoid omitting important explanatory variables, while also simplifying the model.

- **Central, Eastern, Southern, Western** - from the 'SchoolRegion' variable, four dummy columns were created by grouping all the regions according to their geographical location in Ukraine. This approach prevents the model from becoming overly complex with 25 individual columns for each specific area. The Southern region is used as the base group.

For the OLS model with interaction variables, we also included: **NMT** - =1 if NMT, 0 if ZNO, **SexType*NMT**, **Central*NMT**, **Eastern*NMT**, and **Western*NMT**. For DID, we included **Treatment*After**, this is the main variable we want to estimate, it enables us to measure the specific impact of the NMT policy on test outcomes during the years after its implementation. It will show the exact deviation of grades.

4 Results

4.1 Assumptions check

Here we will present our findings from verifying the assumptions of our models to ensure accurate interpretation of the coefficients. All the assumptions of the OLS model apply equally to DID. But in addition, it requires a parallel trend check.

In our two models, the multicollinearity checks yielded varying results, but both were acceptable. In the OLS model, the highest correlation between two independent variables was between 0.75 and 0.77. Although these values are relatively high, they are still below the threshold commonly associated with multicollinearity issues, indicating that our model is acceptable. In DID, we observed even less concern regarding multicollinearity, the highest correlation among the variables was only 0.65.

- **Error normality** - we applied the Jarque-Bera test to the residuals of our models, which indicated a non-normal distributions with p-values of almost zero. However, given the test's sensitivity to large sample sizes, and considering the substantial size of our dataset, even minor deviations from normality can lead to significant results. To further assess this, we visually inspected the residuals through Q-Q plots, which suggest that the residuals largely adhere to a normal distribution in practice. Thus, our models remain potentially reliable, particularly as the deviations are minor and primarily occur only in the tails of the distributions.

- **Heteroscedasticity** - for checking this assumption we employed both the Goldfeld-Quandt test and the Breusch-Pagan test, which yielded differing results. For both models, the Goldfeld-Quandt test showed p-values of 0.9, leading us to conclude that homoscedasticity is maintained. We give greater weight to this result because, with our large dataset, a test like Breusch-Pagan might detect even minor deviations from homoscedasticity that are not practically significant. Also, visually, there appear to be no significant indicators of heteroscedasticity.

- **Zero conditional mean (zero correlation with independent variables)** - this assumption was tested separately for each of our two models by regressing the residuals from the initial model against its independent variables. In both cases, all coefficients were very close to zero, suggesting that there is no significant relationship between the residuals and the independent variables in either model.

- **Serial correlation** - here we used the Durbin-Watson statistic to assess autocorrelation within the regression models' outputs for both models. For the OLS model, the

Durbin-Watson statistic was 1.9535, while for the Difference-in-Differences model, it was 1.9547. Both results, being close to 2, suggest that there is no significant autocorrelation.

- **Parallel trend for DID** - we did not directly test the parallel trends assumption in our analysis because it is evident that the trends are parallel due to the synthetic continuation of the exam that we created. By duplicating data to simulate subsequent years, we intentionally structured our dataset to maintain consistency in trends across different periods, effectively ensuring the assumption holds.

4.2 Coefficients interpretation

Results of OLS model with interaction variables and DID model.

In the context of our project, both the linear regression model with interaction variables and the Difference-in-Differences method serve distinct but complementary purposes.

The linear regression model is particularly useful for estimating how the introduction of NMT interacts with other variables in comparison to the previous ZNO system. This model helps to explore the relationships and potential impacts of NMT on various factors but does not directly measure the policy change's effect on graduation rates or grades.

The Difference-in-Differences method is used to directly assess the deviation in grades before and after introduction of NMT. This approach allows us to isolate and measure the specific impact of the NMT policy by comparing changes over time between groups that were and were not exposed to the policy change.

Both approaches showed similar results for the same factors we looked at. Because of this, we are mainly going to interpret resulting coefficients in DID model, because our main task is to understand how grades changed. But we will also briefly interpret coefficients of the interaction variables in the OLS model to give a bit more insight into how they might affect things.

Furthermore, it's important to mention that different interaction variables can also be estimated using the Difference-in-Differences model. However, in our effort to find the best way to analyze them, we have decided to discuss the interpretation of interaction effects separately from the specific examination of how grades changed due to the policy shift.

So, using the table we see that constant term is equal to -716.1958, knowing that grades fall within the range of 100 to 200, the interpretation of the intercept might not be meaningful. Also, 'Treatment' and 'After' variables are not significant itself, but the interaction variable is. **Treatment:After** represents the estimated decline in Ukrainian grades specifically attributable to the policy change after its implementation, relative to the control group. It suggests that, after the NMT policy was introduced, the Ukrainian language test scores for the treatment group (those affected by the NMT policy) decreased by an average of 2.9 points compared to what would have been expected based on the trends in the control group (those who continued under the ZNO policy). There might be a number of reasons for the decline in grades during that period, such as the beginning of a war and the issues it caused for the entire country. Students may have found it more difficult to concentrate on their academics and do well in class as a result. Additionally, NMT in 2023 was more challenging than the one in 2022, which may have added to the complexity for pupils.

The coefficient for Math, which is 0.5, suggests that there is a positive association between mathematical ability and performance in Ukrainian language. It indicates that pupils with stronger mathematical skills and logical reasoning abilities may achieve higher grades. For instance, if someone's math grade increases by one point, their Ukrainian language grade should increase by approximately half a point, on average. Therefore, learning math can be a good contribution to your future scores.

Birth coefficient suggests that for each additional year of birth, the average grade increases by approximately 0.4 points. Urban, which is 0.9, indicates that residing in an urban area is associated with an increase of approximately one point in the average grade compared to those who live in rural area. In terms of gender, on average, males tend to achieve grades that are nearly 9 points lower than females.

Graduates from the Eastern, Western, and Central regions outperform their Southern counterparts in Ukrainian language scores by 0.2, 2.6, and 1.8 points on average, respectively. These differences may be caused by the language issue in Ukraine. In cities like Kharkiv, Dnipro, and Donetsk, people predominantly speak Russian, while in cities like Lviv, Ternopil, Khmelnytskyi, Chernivtsi, and Uzhhorod, the majority of the population speaks Ukrainian.

The linguistic gap between these regions may impact pupils' performance in subjects like Ukrainian. Depending on which language they are more comfortable with, which language they use with friends and family, and which language they use in their daily life, pupils' grades may vary.

Interaction variables from the OLS model:

SexType*NMT: male graduates experience an increase in average of 4.6 points in their grades after the providing NMT (2022 and 2023), compared to the ZNO (2020 and 2021), after controlling for other variables. This fact may be related to the outbreak of war in 2022. A significant number of males may have decided to move abroad during this period. Therefore, the sample of men decreased and only those who were more confident in their knowledge and prospects for the future in Ukraine remained.

Central*NMT, Western*NMT, Eastern*NMT: these interaction terms reflect the changes in Ukrainian language test scores post the NMT policy's implementation relative to the Southern region baseline and ZNO scores. Since Eastern*NMT is insignificant in our model, we will not interpret it. So, Central region graduates see a 0.7 point reduction and Western region graduates a 1.3 point reduction in their scores. The observed drop in scores may be an indication of a number of underlying issues, such as the negative impact of wartime interruptions on educational standards, that are affecting academic performance.

5 Conclusions

As a result of our project we estimated the trend of Ukrainian language performance over the years, considering factors such as the policy change and other influences. The Ukrainian Ministry of Education is now able to make insightful conclusions. They may specifically use our findings to determine where educational practices need to be improved and to evaluate how effective the policy change was. Also, they can implement targeted interventions, such as specialized teacher training programs or increased funding for schools, to assist regions or demographic groups experiencing declines in performance.

6 Appendix

	Dependent variable: Ukr
Intercept	−727.6169***
	(38.297)
NMT	−4.0111***
	(0.116)
Birth	0.4029***
	(0.019)
Math	0.5466***
	(0.001)
Urban	0.9353***
	(0.043)
SexType	−11.9119***
	(0.064)
SexType_NMT	4.5638***
	(0.078)
Central	2.1917***
	(0.098)
Central_NMT	−0.6731***
	(0.122)
Eastern	0.3069***
	(0.109)
Eastern_NMT	−0.2415*
	(0.138)
Western	3.4545***
	(0.101)
Western_NMT	−1.2654***
	(0.126)
Observations	597599
R-squared	0.453
Adjusted R-squared	0.453
Residual Std. Error	12 (df = 597586)
F Statistic	$4.132e + 04$

Table 2: Regression results with interaction variables $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

	Dependent variable: Ukr
Intercept	-716.1958^{***}
	(38.312)
Treatment	-0.4749^{*}
	(0.519)
After	0.5004^{*}
	(0.733)
Treatment:After	-2.9649^{***}
	(0.735)
Math	0.5469^{***}
	(0.001)
SexType	-8.9041^{***}
	(0.037)
Birth	0.3969^{***}
	(0.019)
Urban	0.9204^{***}
	(0.043)
Eastern	0.1756^{***}
	(0.067)
Western	2.6444^{***}
	(0.061)
Central	1.7912^{*}
	(0.059)
Observations	599129
R-squared	0.450
Adjusted R-squared	0.450
Residual Std. Error	10 (df = 599118)
F Statistic	$4.904e+04$

Table 3: Difference-in-differences results
$*p < 0.1$; $**p < 0.05$; $***p < 0.01$.