EFFECT OF NON-PRICE CRITERIA ON TENDER OUTCOMES

by

Oleh Hromiak

A thesis submitted in partial fulfillment of
the requirements for the degree of

MA in Economic Analysis

Kyiv School of Economics

2020

Thesis Supervisor: _____ Professor Elena Besedina

Approved by _____
Head of the KSE Defense Committee, Professor [Type surname, name]

_____

_____

_____

Date _____

Kyiv School of Economics

Abstract

EFFECT OF NON-PRICE CRITERIA ON TENDER OUTCOMES

by Oleh Hromiak

Thesis Supervisor: Professor Elena Besedina

This paper investigates the relationship between non-price criteria and tender outcomes. Motivation of this paper is to make tender outcomes more economically advantageous. Data from Prozorro is used to test that non-price criteria can affect tender outcomes such as probability of tender failure and number of bidders. We estimate Logit and Poisson models for payment delay and location criteria. The results show that increasing weight of payment delay criterion negatively affects tender outcomes. Also, the average marginal effect of non-price criteria on probability of tender failure is computed with the help of propensity score matching. According to results, the non-price criteria increase probability of tender failure.

TABLE OF CONTENTS

LIST OF TABLES

Number *Page*

ACKNOWLEDGMENTS

I would like to thank my thesis supervisor Elena Besedina for valuable comments, suggestions and detailed reviews. This helped a lot during the research process.

I would also like to express gratitude to all KSE professors and researchers for support and discussions during research workshops and beyond them.

Wish all the best for university staff and everyone who makes KSE possible.

Chapter 1

INTRODUCTION

This thesis discusses one of tender features, non-price criteria, which are not commonly used in Ukraine. The aim of our research is to study how these additional criteria affect tender outcomes. This is important because public procurement sector is quite large in our country.

There are two major ways to select the winner of the tender. We can select the seller with the lowest price or use a formula with different weights in addition to price criteria and then maximize it. Each additional weight is a non-price criterion.

Non-price criteria allow to select the proposal which has not only the lowest price but also provides additional benefits for the organizer. Examples of criteria are a payment type, delay, guarantee, the distance, a taxpayer type or a region. Some firms tend to add tax region criterion, which adds about 10% to the proposal of the bidder which pays VAT in the same region. Usually non-price criteria weights are less than 30% in the winner selection formula.

In this research we use data from Prozorro – the most popular e-procurement system in Ukraine. Open data from this system allows us to study the relationship between the presence of different non-price criteria and tender outcomes, such as the number of participating firms and the probability of tender failure. Most of the tender attributes are obtained through public API, but bids are scrapped from the Prozorro. The platform provides bids API for business accounts only. We limit our analysis to one homogenous good and exclude all tenders which contain bundles. This is due to data quality and low number of tenders with non-price

criteria. Total number of tenders, with or without criteria, is more than 3 million as of September 2019. According to our estimate there are around 15 thousand tenders with non-price criteria on Prozorro overall, which is less than 1% tenders with non-price criteria on the platform.

Another reason which motivates our research is that the government is one of the major buyers in the Prozorro system. State-owned firms and institutions are often required to use public procurement system to buy goods in the market as it is defined by the law.

Absolute majority of state-owned firms use only minimum price criteria in the procurement which sometimes leads to undesired outcomes. While minimum acceptable quality should be enforced by tender documentation, overall there may be a huge difference in quality and costs associated with the same goods produced by different firms. This implies that buyers may not receive the best deal in terms of overall costs associated with ownership and usage.

More and more state-owned firms and institutions can use non-price criteria for their tenders because of the recent regulations. Unfortunately, lack of information on this topic may result in wrong weights or unbalanced formulas in tenders. Our research tries to fill this gap and provide enough information for decision-makers, so that they can get the best possible deal.

We think that this research applies to the same extent not only to the state-owned but to private firms too. Rational firm are trying to minimize their costs and get the best value available at the market.

It is expected that non-price criteria hamper competition, by limiting sellers and only allowing some of them to participate in tenders. It looks like discrimination

and it should come at some price. It would be good to estimate which price approximately should be paid for each criterion used.

In this paper we use Poisson model to estimate the relationship between non-price criteria weights and total number of firms participating in the ender, and Logit model for the probability of tender failure. In the model, we account for price, tender characteristics, region and seasonality. We confirm the relationship for payment delay criterion but for location criterion it is not confirmed.

We understand that our model may be biased because of unbalanced sample: the tenders which use only minimum price criteria take more than 90% of the sample. That's why use propensity score matching algorithm to obtain results based on balanced sample. Our treatment is defined as presence of non-price criteria. We estimate average treatment effect on tender failure. For matching the nearest neighbor algorithm is used. It finds similar tenders in two groups: treatment and control, in another words, with and without non-price criteria. Then we estimate average effect of treatment and find that there is statistically significant effect of non-price criteria on probability of tender failure.

This thesis is organized as follows: after this introduction comes chapter 2 with literature review. We describe main findings in papers which studied non-price criteria in different countries. We shortly discuss main findings and methodology limitations.

Chapter 3 is divided into two sections: data and methodology. In the first section we describe in details data issues and algorithms we used to eliminate them. In the second section regression models and hypothesis are provided.

Results are in chapter 4 and conclusions follow them in chapter 5. You can find references at the end of this paper.

Chapter 2

LITERATURE REVIEW

I divide my literature review section into several parts: first part is dedicated to theoretical papers on procurement; second part concentrates on papers with simulations and empirical evidence. The third part discusses studies specifically focusing on Prozorro.

2.1. Theory

One of the first economists who investigated this topic was Stilger (2011). The main objective of the paper is to compare different formulas that can be used for tenders. Study contains more than twenty formulas. It is purely theoretical; no real procurement data is involved.

This paper is interesting for us because it provides a good collection of formulas with non-price criteria and several possible theoretical tests that may be used to differentiate them. Some of the formulas mentioned there are used in European Union public procurement systems. That is why this study may be useful to understand empirical papers investigating effects of non-price criteria on tender outcomes.

Researcher states three main goals for buyer: minimum costs, which means lower price for the same good, efficiency, the most appropriate use of available resources, and effectiveness. But there are more things to consider: protection against abnormal prices, shape of indifference curves, ranking paradox, Pareto optimality.

Ranking paradox occurs when the entire order and the winner are changed after removal of one not the best bid. Result is said not to be Pareto optimal in case there is a cheaper bid with a better quality. If the result is not Pareto optimal, then, probably, formula is ineffective.

This study is quite extensive, it also includes a comparison between relative and absolute formulas, zero scores test, symmetry of the scores. Symmetry is defined as follows: if two bidders submit the same bids, in terms of quality, price and other options (if applicable), they should be ranked at the same place or have the same score.

Results include likelihood of ranking paradox, overvaluing mistakes and shape of the indifference curve. Overall the author recommends buyers to choose different formulas according to their needs, do not complicate things, use decision support system, prohibit more than 1 bid per supplier and attract more sellers.

This study lacks empirical verification, because all conclusions are based on theoretical results. Public procurement is also about data; therefore, such tests are required before advising strategies to some tender entities.

Also, some tests are questionable. Zero score test does not imply anything about the formula, because in procurement we always evaluate relative results in formula, not absolute. This means, it does not matter whether formula returns 0 for all zeros as input or no. We can do a thought experiment to prove this. Take any formula which passes zero score test and add (-1) term to it. New formula does not pass the test, but it is the same formula with the same ranking and all other attributes. Tests related to protection against extremely high prices do not answer the question of what the desired quality is.

Overall the research is a good theoretical background for this thesis, but we can use it only as a reference to our empirical findings.

2.2. Simulations and empirical studies

Stilger, Siderius, Raaij (2012) is an empirical study. Authors take more than 20 formulas using data from one of the procurement services providers.

This research shows its unique way to analyze non-price criteria. It investigates not only existing tenders, but simulations based on real data. It is a little bit different from our goal because we are not looking for the best formula with non-price criteria. We are trying to understand whether we need minimum price formula or weighted criteria. The value of this research is in application of real-world data to analysis.

Data contains around 350 tenders and 1600 bidders related to them. Time span is 10 years. It is the first empirical research related to the topic.

Outcomes for each formula have been simulated and compared. It has been found some formulas cannot be tested correctly because additional information or weights are required.

Main conclusions are related to protection against high prices and decreasing the probability of the ranking paradox. There are empirical tipping points which show a kind of "safe" ranges for price weights. For example, if the tipping point is 20% and price weight is less than 20%, then the formula protects against very high quality and enormous price associated with it. Also, authors estimate shapes of indifference curves and probability of rank paradox.

However, paper does not provide enough insights on methodology and work done. It does not contain enough descriptive statistics of the data and it looks like non-

public dataset is used. This implies the fact that results cannot be trusted and verified.

Another reason which arises on simulations is unknown valuations. To determine the outcome of the auction you need rules and valuations of the players. We have information about rules, but valuations are not the case for empirical data. That is why one cannot just change tender conditions and precisely estimate outcomes in a new environment. Such an estimation requires assumptions about players. Unfortunately, we can have two substantially different simulations on the same conditions, but different assumptions.

Lewis and Bajari (2012) look at public procurement for road building in California. Main non-price criterion there is specified as the number of days to finish the road. Bid is constructed as a combination of the labor costs, material costs and construction duration.

The research is tightly linked to social welfare. Authors estimate how different tender parameters affect society. Social welfare depends on aggregated delays for people using the road, their time value and traffic intensity.

From the point of view of the government it makes sense to select the fastest construction period subject to budget constraint. Another approach is minimizing costs subject to maximum days construct.

Research results show that non-price criteria not only reduce the number of days needed for road construction / repair, but even more, they do not reduce construction quality. On average, such contracts reduce construction time by 35%. Social welfare gains are shown to be bigger than budget losses from this kind of policy. It implies policy is reasonable and applicable in the region and makes taxpayers better off.

What is more, authors do not observe any statistically significant relationship between number of participants and presence of non-price criteria in tenders. It seems that participation depends on tender size and road distance rather than on non-price criteria.

Lewis and Bajari (2011) in their paper focus on penalties for road construction tenders in Minnesota. Also, they added productivity shocks to the estimation.

One of the main results presented in the paper: if there is a relationship between tender penalties and traffic delay cost, then overall result is better.

Results also imply that late contract completion statistically depends on time penalty. It helps government representatives to make better policy decisions so that roads are built optimally - cheaper and faster.

Both papers mentioned above are interesting for us in terms of statistical analysis linked to non-price criteria. The social welfare cannot be estimated in our case, but our research applies similar approach to answer the questions whether we need non-price criteria and how they affect tender outcomes.

We do not add shocks, labor and other types of costs in this thesis at the moment. This option may be considered as a possible improvement to our model based on the abovementioned paper.

2.3. Evidence linked to Prozorro

There are a lot of articles written by the Centre of Excellence in Procurement researchers about Prozorro. For example, Stepaniuk (2017) studies how overestimated reserve price affects the number of tender participants and auction results such as bids and competition in different rounds.

Research was conducted for selected goods only, in particular natural gas, one popular paper type and chicken eggs.

Dependent variables are the lowest price per item in zero round, last round price divided by the first-round price and number of tender participants. Control variables include seasonality, tender type, number of bidders and quantity.

It is shown that too high reservation price does not induce any positive effects for the buyer: average zero-round price increases, average number of participants does not change and average dynamic phase savings either decrease or stay the same.

Another interesting research related to public procurement and health system is Chmel (2018), on procurement of medicine. Author assesses the level of savings after introducing an electronic procurement platform in 2015.

Using OLS regressions author estimates the effect on savings (tender price market price ratio) of several tender characteristics: type of tender, number of bidders (with square term), organizer's and winner's experience, quantity, whether tender is below or above legal threshold of 200 thousand UAH. Also, monthly and oblast dummies were included.

It was found that usage of Prozorro increased savings up to 6% which is quite a good result. Author concludes that electronic auctions should be made obligatory for more drugs.

Overall, existing papers provide some basic information about non-price criteria, but only a few related empirical papers were found. This fact confirms our paper is relevant in the field and adds value to existing discussion.

Chapter 3

METHODOLOGY AND DATA

3.1. Data source and construction of the dataset for analysis

We use data on tenders from Prozorro public API, combined with public firm registry data and web scraping from open sources.

First, we downloaded all tenders from Prozorro public procurement system. It contains information about 3 mln tenders as of September 2019. Each tender contains many attributes such as time, quantity, total price, type, status, buyer info, and non-price criteria.

Then, we dropped all tenders with bundles. Unfortunately, data for bundles cannot be used in the analysis since Prozorro does not have information on the goods prices are calculated as unit values by dividing total amount of the tender by number of units procured. If tender contains bundles this effectively means we cannot extract any prices and sometimes quantities for bundled goods. Both prices and quantities are critical for our research, that is why we drop abovementioned tenders.

Next step is to select good for our analysis. We used two criteria. First criterion is that good should be homogenous, i.e. have the same quality within all tenders. Second criterion: tenders for this good should have a significant part of all non-price criteria tenders placed in the procurement platform.

After analyzing the data, we have found that fuel satisfies both criteria. That is why we selected fuel A-92 for our analysis.

Quantity field is affected by our good properties. Some buyers specify quantity not in liters but in metric tons. We use common fuel specification to determine its average density. Then fuel density is used to convert metric tons to liters according to this formula:

$$quantity_{ltr} = quantity_{tne} * \frac{1000}{density} \quad (1)$$

where $quantity_{ltr}$ is in liters, $quantity_{tne}$ is in metric tons, density is average fuel density from the specification. This action is required to calculate comparable expected price per liter and to obtain correct comparable quantities for all observations.

There are several text fields which are useful for our research. These fields should be converted to the computer-readable format which can be used in automatic processing.

For example, "region" field of the tender organizer (buyer). It contains partially parsed information from the unified state firm's registry (EDR). Unfortunately, these fields may contain arbitrary text, either in Russian or Ukrainian language. To resolve this issue, we use the following algorithm:

1. Manually prepare a list of regions in the country.

2. Manually prepare mapping of regions to more common entities such as west, east, north, south, center.

3. For each observation we remove all unnecessary words such as 'region' or 'city'.

4. For each observation replace all Russian word endings with Ukrainian ones.

5. For each observation find the corresponding region name from the prepared list using Levenshtein distance algorithm with threshold=2. This means we tolerate up to 2 wrong (different) letters in the region name. With the help of the algorithm, we can find slightly changed region names e.g. ones with typos or very similar Russian titles.

6. Map each region name to common entities using mapping from 2.

To obtain "firm type" field we match tender organizers with the unified state firm's registry. The registry is provided by the government in open data format. We use unique firm id to match our dataset with the registry. We fill "firm type" for state firms only.

For our research we need to know two important tender attributes: whether it is public tender and whether it is above threshold.

Prozorro public API provides enough information to get above mentioned tender characteristics.

We assume then tender is open if procurement "MethodType" field isn't specified as 'reporting' or 'negotiation'.

'Above threshold' attribute is a little bit more complicated. There are three types of 'above threshold' tenders in Prozorro depending on country and tender specifics: for Ukraine, for European Union and for defense-related tenders.

We treat all three types as one and flag our tender if one of them is found. We do not observe any evidence that treating the above-mentioned types as different may distort our analysis in any way.

One of our important dependent variables is tender failure. It is a logical value which takes true if tender is failed.

By failed tender we mean one of two possible outcomes:

- tender was unsuccessful which happens in case of lack of bidders (sellers)
- tender was cancelled due to an appeal or another technical or juridical issue

In this research we do not distinguish these two situations in any way.

As mentioned above for goods prices we use unit values obtained by dividing tender value (expected or actual) by quantity.

Since our focus of research is tenders with non-price criteria, we construct a number of variables based on the information about them.

First, we group all non-price criteria in the dataset into groups, one per criterion. Before processing the criteria, we also drop outliers according to the following conditions:

- very small groups are dropped (less than 10 tenders contain the criterion)
- non-price criteria which have zero weights are dropped

This results in three types of non-price criteria in our dataset:

1. Payment delay. It comes with a number of days after shipment. Buyer is obliged to pay all bills within this period.

2. Location. Firms located in the specific region or area are preferred to others.

3. Shipment delay. It comes with a number of days after tender evaluation. Seller is obliged to deliver all items within this period.

For each criterion listed above we record weights for the criteria stated in the tender documentation.

In addition, we remove tenders with abnormal quantities and prices. The threshold is determined by the last percentiles.

We also dropped all observations with wrong or empty regions specified.

All observations with empty or zero prices or quantities were also removed.

3.2. Data summary

Our dataset has 7,889 observations. The summary statistics is in the table 1 and table 2 below.

Table 1. Summary statistics for continuous variables

Variable	Min	Max	Mean	St. dev.
payment delay, weight in evaluation, %	0	30	3.99*	3.46*
location, weight in evaluation, %	0	30	3.71*	9.78*
quantity, liters	4	1500000	4819	26840
expected price per liter, UAH	15.18	35	24.9	3.33

* non-price criteria tenders only

Table 2. Summary statistics for categorical variables

Variable	% of True values
Type of tender (=1 if open)	60%
Tender failure (=1 if failed)	19%
Type of tender price (=1 if above threshold)	3.9%
Tender organizer type	local government: 1.9% government institute: 3.3% government organization: 24% government: 12% communal: 19% communal organization: 21%
Region	center: 30% north: 21% west: 22% east: 16% south: 11%
Year	2016: 24% 2017: 38% 2018: 29% 2019: 9%

We see that typical payment delay criterion weight in evaluation is less than 10%. Quantity range is quite high but expected price per item lies in small interval from 15 to 35 UAH per liter with average of 25.

Most of the tenders are open and organized by state-owned firms. Absolute majority of expected prices for open tenders are below the threshold. Distribution across regions is close to the uniform.

3.3. Econometric model

In this thesis we study the effect of non-price criteria on tender outcomes. First, we estimate the logit model to assess how non-price criteria affect probability of failure. We expect that probability of tender failure is affected by non-price criteria. We also control for item price, quantity, tender type, firm type, region and seasonality:

$$\Pr(failed) = f(expected\ price\ per\ item, organizer\ type, type\ of\ tender, non-price\ criteria, region\ and\ time) \quad (2)$$

Our main expectation is that imposing additional non-price criteria results in increased probability of tender failure. In the table below we list expected signs of other independent variables used in the model.

Table 3. Expected signs in the Logit model

Independent variable	Expected sign
expected price per item	decrease, '-'
open tender dummy	increase, '+'
above threshold dummy	increase, '+'
government firm	increase, '+'
payment delay and location	increase, '+'

The bigger the expected price per item is, the more bidders may come to the tender. This is related to the fact that bidder's decision to participate in the tender is affected by expected payoff which is based on valuation and expected price. Increasing expected price allows more bidders to get positive payoff. We assume that there is a negative relationship between expected price per item and probability of tender failure.

Open tender procedure should increase probability of failure, and this sign has two reasons. Firstly, private firms may not publish them on Prozorro. Secondly, the open tender is more likely to be disputed relatively to closed tender.

If the tender value is above threshold defined in the law, then it should be considered as a relatively large tender in terms of quantity. Not all firms in the market are big enough to provide the desired quantity of the good to the buyer. That is why we expect that this tender characteristic leads to more failures and therefore we state expected sign as positive.

Selling goods to state-owned firms is a risk form the point of view of bidder. It may require additional documents and usually this kind of tenders are paid after delivering the goods. This kind of tenders is more likely to be disputed which may also result in tender failure. We include positive sign for government firm dummy to indicate above-mentioned problems.

Non-price criteria impose additional restrictions. This may distract firms from participating in the tender because of the discrimination. It is naturally to expect that each additional percent to weight of any non-price criterion increase probability of tender failure.

Next, we evaluate the effect of non-price criteria on the competition level, defined as the number of firms participating in the tender. We believe that heavy

weights on non-price criteria may significantly discriminate participants of the tender. It is expected that less firms are willing to participate in the procurement process if additional restrictions are imposed. Extra criteria may eliminate relative advantages of some firms so that their optimal behavior is to skip the tender.

We also control for item price, quantity, tender type, firm type, region and seasonality:

$$Bidders = f(expected\ price\ per\ item, \\ organizer\ type, type\ of\ tender, \\ non-price\ criteria, region\ and\ time) \quad (3)$$

We use Poisson approach to estimate the model.

Our main expectation is that imposing additional non-price criteria results in less firms participating in the tender.In the table below we list expected signs of all independent variables used in the model.

Table 4. Expected signs in the Poisson model

Independent variable	Expected sign
expected price per item	increase, '+'
open tender dummy	increase, '+'
above threshold dummy	increase, '-'
government firm	decrease, '-'
payment delay and location	decrease, '-'

We expect that higher prices attract more firms. Similar to the previous model, bigger expected price per item results in higher expected payoff which should encourage firms to participate in the tender.

Open tender procedure should also increase number of participants. It is easy to observe public tenders on Prozorro than on different information sources which may not be known to the firm. Some negotiation procedures may not include any notice for potential participants.

Expected value of the tender which is above threshold is expected to decrease number of firms participating in the tender. The reason for this is that is large quantity of good may not be provided by small or even medium size firms. This may prevent some firms from participating in the tender because they are not able to meet technical specification of the tender.

Tenders organized by state-owned companies should discourage firms from participating because of additional costs associated with them. This may include anti-corruption laws, bigger probability of dispute and public attention.

We expect that non-price criteria limit the competition as not many firms may be able to satisfy them. Hence some firm may expect their low bid score because of specific non-price criterion.

We have evidence that our sample may be unbalanced because of low fraction of tenders with non-price criteria. For this reason, we use propensity score matching procedure. Hereafter we define treatment as presence of at least one non-price criteria in the tender specification, other tenders will represent a control group.

First, we estimate propensity score for each tender. It is based on expected price per item, firm type, region and season variables. This metric allows us to calculate

the distance between any tenders. Propensity score matching allows finding similar tenders in terms of observable characteristics in two groups: control and treatment.

Second, we do balance checks. We check whether propensity score looks similar for control and treatment groups. In case it does not look similar, then we should stop our analysis and redefine our variables used to estimate the propensity score. If the balance condition is satisfied, we can go to the next step.

Third, we use nearest neighbor algorithm. It uses propensity score to find the closest tenders in two groups. We use tenders with non-price criteria as a base sample. For each tender in the treatment group we find one or more tenders without non-price criteria. Then we select one tender from the control group which is the most similar to the tender from the treatment group in terms of propensity score. The process continues until all tenders from the treatment group are matched.

After nearest neighbor matching, we verify whether the balancing property of our dataset is satisfied. For each dependent variable we verify whether the control and the treatment group are not significantly different. Then we conduct the analysis of the effect of the 'treatment' on the tender outcome.

Analysis is based on the statistical test between two groups. We determine the average treatment effect, i.e. whether for two groups we have a statistical difference in the probability of tender being a failure. If the average treatment effect is positive and significant then we can say that there is a statistically significant relationship between presence of non-price criteria in the tender and probability of tender failure.

We expect propensity score matching results to be consistent with other models. Expected sign of average treatment effect on probability of tender failure is positive. In other words, presence of non-price criteria should increase probability of tender failure.

C h a p t e r 4

ESTIMATION RESULTS

4.1. Probability of tender failure

Logit regression has been estimated using R. The dependent variable is the probability of tender failure. Regression output is listed in the table 5, the average marginal effects are in the table 6. Regional and seasonal effects are omitted in the output.

Table 5. Logit estimation results

Variable	Coefficient
location	-0.010
	(0.018)
pay_delay	0.084**
	(0.041)
is_open	3.373***
	(0.150)
above_threshold	0.957***
	(0.132)
firm_type_localgov	-0.594***
	(0.230)
firm_type_govinst	-0.684***
	(0.224)
firm_type_govorg	-0.312***
	(0.108)
firm_type_gov	-0.423***
	(0.120)

TABLE 5 – Continued

Variable	Coefficient
firm_type_comun	-0.705***
	(0.112)
firm_type_comunorg	-0.753***
	(0.111)
Constant	-1.481***
	(0.458)
Observations	7,889
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6. Average marginal effects for Logit model

factor	AME	SE	z	p	lower	upper
location	-0.001	0.002	-0.554	0.579	-0.006	0.003
pay_delay	0.011	0.005	2.069	0.039	0.001	0.021
is_open	0.289	0.074	39.072	0.000	0.274	0.303
above_threshold	0.140	0.021	6.723	0.000	0.099	0.180
exp_price_per_item	-0.011	0.003	-4.092	0.000	-0.016	-0.006
firm_type_localgov	-0.081	0.029	-2.803	0.005	-0.138	-0.025
firm_type_govinst	-0.045	0.016	-2.871	0.004	-0.075	-0.014
firm_type_govorg	-0.092	0.027	-3.376	0.000	-0.146	-0.039
firm_type_gov	-0.060	0.017	-3.537	0.000	-0.093	-0.027
firm_type_comunorg	-0.100	0.015	-6.666	0.000	-0.130	-0.071
firm_type_comun	-0.095	0.015	-6.217	0.000	-0.124	-0.065

Here we see that our hypotheses were mostly confirmed by the data and model. Estimation results show expected signs for control variables. All positive signs in this model should be interpreted as increased probability of tender failure,

while all negative - vice versa. Coefficients are displayed as logarithms of odd ratios.

Our main variables of interest are location and payment delay. Unfortunately, we do not observe statistically significant relationship between probability of tender failure and weight of location non-price criterion in this model. However, this result may be caused by the low number of tenders with this criterion in the dataset.

For payment delay non-price criterion, we observe statistically significant positive relationship. This means that increasing weight of this criterion in the tender is associated with significant increase in the probability of tender failure. This is what we expected.

We see that magnitude of the effect is the biggest for payment delay non-price criterion weight and open tender dummy. Probability of failure for open tenders is 18% higher probability of tender failure. Adding one more percent to payment delay criterion results in 1.09 odd ratio increase of probability of tender failure. Average marginal effect for payment delay weight is 1 percentage point increase of probability of tender failure. We can conclude that these two variables are driving probability of tender failure much more than others.

According to our model expected price decreases probability of tender failure, but magnitude is too small to affect the probability significantly in real world.

Open tender procedure leads to higher probability of tender failure**.** This result looks correct because open tenders tend to have more disputes and private firms may not publish some closed tenders.

Our estimates show that some companies, especially communal one, have better tender outcome. This may be related to law restrictions or different procurement procedures for some firm types.

We should interpret our results with caution as this regression may suffer from lack of tenders with non-price criteria. Extending this analysis to more goods can potentially help; however, many goods are not homogenous hence in this case one has to control for quality which can complicate the analysis.

4.2. Number of firms participating in the tender

Table 7. Poisson estimation results

Variable	Coefficient
location	-0.015*
	(0.008)
pay_delay	-0.085***
	(0.027)
exp_price_per_item	0.043***
	(0.008)
is_open	20.465
	(276.710)
above_threshold	0.418***
	(0.048)
firm_type_localgov	0.096
	(0.098)
firm_type_govinst	0.239**
	(0.095)
firm_type_govorg	0.099*
	(0.052)
firm_type_gov	0.078
	(0.056)

TABLE 7 – Continued

Variable	Coefficient
firm_type_comun	0.239***
	(0.050)
firm_type_comunorg	0.309***
	(0.049)
Constant	-21.571
	(276.710)
Observations	7,889
Note:	*p<0.1; **p<0.05; ***p<0.01

We got what we expected for all control variables.

Coefficients on the non-price criteria variables show two things. First, the effect of location on the level of competition is not statistically significant in the model. It means the relationship between location criterion and number of firms participating is not confirmed. Second, presence of payment delay criterion statistically decreases the number of firms participating, that is what we predicted in our hypotheses. Adding one more percent to weight of payment delay criterion decreases number of bidders participating by 8%.

Similar to the previous regression, we see that main driving forces for total number of firms in the model are non-price criterion and expected price.

Open tender dummy is not significant here, but above threshold dummy shows that large tenders seem to attract more firms.

Higher expected per unit price also attracts more firms, but the magnitude of the coefficient is small. Practically it means that there is no effect of expected price per item on number of sellers participating unless price is abnormal.

4.3. Probability of tender failure for balanced sample

In this subsection I present results for the propensity score matching based on observable characteristics of the tenders. Table 8 presents balancing statistics for matched sample generated by nearest neighbor algorithm. We observe that matching algorithm worked well as the means for two groups are similar and it is possible to conduct the analysis based on this sample.

Table 8. Propensity score matching balance statistics

	Means Treated	Means Control	Mean Diff
distance	0.14	0.14	0.00
exp_price_per_item	27.31	27.14	0.17
firm_type_	0.82	0.80	0.02
firm_type_localgov	0.00	0.00	0.00
firm_type_govinst	0.04	0.04	-0.01
firm_type_govorg	0.00	0.00	0.00
firm_type_gov	0.04	0.06	-0.02
firm_type_comunorg	0.09	0.09	0.00
firm_type_comun	0.02	0.01	0.01
region_south	0.13	0.14	-0.01
region_north	0.17	0.17	0.01
region_eat	0.15	0.17	-0.02
region_center	0.20	0.15	0.04
quarter2	0.19	0.17	0.03
quarter3	0.27	0.28	-0.01
quarter4	0.34	0.39	-0.06
year2017	0.18	0.18	0.00
year2018	0.63	0.59	0.03
year2019	0.04	0.04	0.01

Our next step is the estimation of the average treatment effect. We conduct the t-test for probability of failure based on two groups: control and treatment group. Result of the t-test shows statistically significant difference between them. P-value is very small, around zero.

According, to our t-test lower and upper confidence interval boundaries for treatment effect are 23% and 42% respectively. It means that adding any non-price criteria to the tender increases probability of tender failure. Sign of the result matches our expectations. On average, tenders with non-price criteria are 32% more likely to be unsuccessful.

Chapter 5

CONCLUSIONS

In this research we estimated a statistical relationship between non-price criteria and tender outcomes. We used two different types of dependent variables: presence of non-price criteria and formula weights for different criteria.

Applying propensity score matching procedure we find a statistically significant difference between control and treatment groups in terms of probability of tender failure. Whenever buyers include non-price criteria into their tender specification, it increases probability of tender failure.

As for specific non-price criteria, we find a negative effect of non-price criteria weights on tender outcomes for payment delay, i.e. increase in the likelihood of failure and decreased competition. This has an important implication for policy makers which may restrict usage of such criterion in tenders. Decision makers might be interested in evaluating benefits and costs before adding this criterion to the tender. We recommend policy makers to avoid payment delay non-price criterion unless a benefit from using it offsets higher probability of tender failure and a smaller number of sellers.

Another policy option is to estimate and use the optimal payment delay weight in the tender. This optimal weight may be based on buyer's objectives, benefits coming from later payment and all the costs like decreased number of firms.

On the contrary, another non-price criterion, location, does not seem to have statistically significant effect. Our model shows there are no disadvantages of

using location non-price criterion. It makes sense to apply this criterion in case it provides any benefits for buyer. Still, this model could be updated to include more data, control variables and different region selection. This may lead to opposite results in the future.

Some types of government firms in Ukraine do better than others in terms of tender outcomes. This relationship should be investigated deeper and extended to other similar countries. It could provide valuable insights about public procurement and lead to better budget allocation.

Results of this research can be applied for private firms too. In Prozorro, they act not only as sellers, but as buyers too. Usually firms have an incentive to get the best proposal in the market. They can optimize their seller selection criteria to be better off in terms of tender outcomes. Sellers could use this research to limit number of applications and therefore reduce their costs associated with tender applications.

Government could use results obtained in this paper to develop better policies and procurement rules. One field of improvement is to limit abnormal non-price criteria usage. Another field is related to the budget process. We know that it is the common reason of payment delay criterion usage. That is why it makes sense to decrease average payment time so that firms have less incentives to include this criterion in their tenders. In turn, it should lead to better tender outcomes and social benefits associated with procurement improvements.

One of possible model improvements is to combine tenders with the data from other registries. Currently, we use only one public registry, but there are much more datasets that provide information about firms. This way the more precise results may be obtained.

Game theory may be used to investigate the research question from the different point of view. For empirical results, it requires dataset with seller valuations which is hard to collect.

WORKS CITED

Chmel, Oleksandra 2018. Prozorro e-auctions: savings on public procurement of medicines in Ukraine, KSE Master thesis

Lewis, Gregory and Patrick Bajari 2011. Incentives and adaptation: evidence from procurement in Minnesota.

Lewis, Gregory and Patrick Bajari 2012. Procurement contracting with time incentives: theory and evidence

Stepaniuk, Oleksa 2017. Reserve price above the market level does not attract more participants to public procurement auctions. Evidence from Ukraine

Stilger, Stan. 2011. Formulas for Choosing the Most Economically Advantageous Tender - a Comparative Study.

Stilger, Stan, Jan Siderius and Erik M. van Raaij. 2012. A Comparative Study of Formulas for Choosing the Economically Most Advantageous Tender.