CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Metal Detecting Ownership and Non-Ownership Motives

Bachelor's thesis

Author: Jan Hrušák

Study program: Economics and Finance

Supervisor: doc. PhDr. Martin Gregor, Ph.D.

Year of defense: 2023

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, August 1, 2023

Jan Hrusak

Abstract

This thesis addresses the phenomenon of hobby metal detecting in the Czech Republic, aiming to determine whether individuals' wealth influences the formation of detected finds collections. The mass spread of metal detecting in the Czech Republic since the 1990s has proven the significance of studying this activity. The phenomenon of metal detecting can be classified under contest theory, where agents make costly efforts to compete for a limited resource, which in this case is archaeological finds. From the standpoint of economics, a study about resources allocation, the metal detecting hobby can be an intriguing topic for investigation. This thesis presents estimates of models based on five different datasets, each containing several thousands of observations obtained from a renowned Czech metal detecting website. The findings suggest that relatively wealthier metal detectorists are more likely to submit coins, but not artifacts. Given that coins form a relatively homogeneous group, the estimation results associated with coins might be applied to the formation of finds collections overall. Hence, the collecting of finds is likely to be negatively associated with an individual's socioeconomic status.

JEL Classification	D14, D61, D91, O33, Q01, Q32, Q34, Z13
Keywords	hobby, private ownership, motivation, motives, metal detecting, finds

Title	Metal Detecting Ownership and Non-Ownership Motives

Author's e-mail	`53600945@fsv.cuni.cz`
Supervisor's e-mail	`martin.gregor@fsv.cuni.cz`

Abstrakt

Tato práce se zabývá fenoménem hobby metal detectingu v České republice, s hlavním cílem zjistit, zda bohatství jednotlivců ovlivňuje tvorbu sbírek nálezů. Masové rozšíření metal detectingu v České republice od 90. let 20. století potvrzuje důležitost studia této aktivity. Fenomén hobby metal detectingu může být zařazen do teorie soutěží, kde agenti vynakládají úsilí a soutěží o získání limitovaného zdroje, jímž jsou v tomto případě archeologicky cenné nálezy. Z hlediska ekonomie jakožto vědy o alokaci zdrojů, může být hobby metal detecting zajímavým předmětem studia. Tato práce prezentuje odhady modelů založených na pěti různých datasetech, obsahujících několik tisíc pozorování získaných z české etablované stránky o metal detectingu. Zjištění naznačují, že relativně bohatší detektoráři mají větší pravděpodobnost odevzdání mincí, ale nikoli artefaktů. Jelikož mince jsou relativně homogenní skupinou, ve výsledku by odhady spojené s mincemi mohly být aplikovány na tvorbu sbírek nálezů obecně. Sbírání nálezů je tedy pravděpodobně negativně spojeno se socioekonomickým statutem.

Klasifikace JEL	D14, D61, D91, O33, Q01, Q32, Q34, Z13
Klíčová slova	hobby, soukromé vlastnictví, motivace, motivy, detektor kovů, nálezy

Název práce	Vlastnické vs nevlastnické motivace držitelů detektorů kovů

E-mail autora	`53600945@fsv.cuni.cz`
E-mail vedoucího práce	`martin.gregor@fsv.cuni.cz`

Acknowledgments

I want to express my sincere gratitude to doc. PhDr. Martin Gregor, Ph.D., the supervisor of my thesis, for the countless insightful comments, suggestions and willingness to offer advice whenever I sought it. It meant a great deal to me and I learned a lot from their expertise that will undoubtedly enhance my future work. I would also like to extend my thanks to my family and all those who motivated and supported me throughout my studies.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University. The AI language model (ChatGPT) was utilized for grammatical corrections, adjustment of text and tables for LaTeX formatting, and adjustment of code.

Bibliographic Record

Hrusak, Jan: *Metal Detecting Ownership and Non-Ownership Motives.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2023, pages 149. Advisor: doc. PhDr. Martin Gregor, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

LPM Linear Probability Model

OLS Ordinary Least Squares

MLE Maximum Likelihood Estimation

PEA Partial Effect at the Average

APE Average Partial Effect

ROC Receiver Operating Characteristic

AUC Area Under the Curve

LLR Log-Likelihood Ratio

LR Likelihood Ratio

WLS Weighted Least Squares

CLT Central Limit Theorem

VIF Variance Inflation Factor

TP True Positives

TN True Negatives

FP False Positives

FN False Negatives

TPR True Positive Rate

FPR False Positive Rate

PAS Portable Antiquities Scheme

HTML Hypertext Markup Language

URL Uniform Resource Locator

IQR Interquartile Range

Chapter 1

Introduction

With millions of metal detectorists worldwide, the hobby of metal detecting has become increasingly popular over the last few decades. With such a high level of popularity, it is likely that there are a number of driving forces behind this trend. It could be either a passion for history, a passion for discovering new things, or on the other hand a passion for gathering, creating collections of items and possibly a desire for individual wealth and status improvement. As a significant amount of metal detectorists' finds being a part of the antiquities segment of art (van der Lande (2021)), in general, they became one of the frequently traded assets, thanks to the properties of both, consumption and investment goods (Havlovicová (2020)). Consumption means that one can gain utility from a find, either as an item that completes a collection, simply lying on a collector's shelf or as a find displayed in a showcase. Such a find may also have investment value, as it is a scarce good with great potential to appreciate over time; additionally, it may serve as a symbol of prestige (Thompson (2016)).

Existing literature addresses the motivations for collecting art, particularly in the case of antiquities, with the aim of identifying collectors' motivations to prevent the illegal trade in artifacts and thereby the destruction of cultural heritage. On the one hand, it is assumed that cultural heritage is being destroyed primarily for financial reasons, i.e. those who destroy cultural heritage, mainly through detector prospecting, are motivated precisely by the high prices of the antiquities found and the vision of high profits from their sale (Thompson (2016)). By contrast, interestingly, the vast majority of detectorists claim that they search out of a passion for the search itself and not out of a passion for the objects they find (Maaranen (2016)).

From the above example of conflicting assumptions and opinions about the motivations for metal detecting from both sides, it might be beneficial to clarify the actual motivations of metal detectorists by analyzing their characteristics and the nature of their finds, i.e. to get a clearer picture of the driving forces behind the metal detecting hobby.

Importantly, the phenomenon of metal detecting could be generally classified in contest theory, which is an economic theory describing contests. In a contest, economic agents expend effort to obtain scarce prizes (Fu & Wu (2019)). In the case of metal detecting, these prizes would be represented by artifacts and coins. The agents would then be both archaeologists and metal detectorists, with a group of archaeologists competing with metal detectorists, while individuals from both groups compete with each other to obtain valuable finds. This situation has arisen because property rights in the Czech Republic to these finds are not well defined and well protected, as evidenced in practice (Hajšman *et al.* (2019); Komoróczy (2022)). It is important to model how decisions are made when rival agents compete for contestable resources and to what extent and for what reasons they are wasted (Ngo (2013)).

Therefore, the aim of this thesis is to find characteristics and patterns behind the metal detecting hobby in order to get a clearer picture of possible motivations for metal detecting and finds collecting, which, as a result, might potentially cause the allocation of this resource to be socially inefficient.

In 1907, Georg Simmel wrote in his book The Philosophy of Money, that value "is never a 'quality' of the objects, but a judgment upon them which remains inherent in the subject" (Simmel (2004), p.60). Using this proposition in our setting, we assume that individual metal detectorists (subjects) have different utility functions from the hobby itself (object). More specifically, each individual metal detectorist should have a specific utility function from metal detecting, shaped by different weights or valuations of the metal detector search itself on the one hand, and the individual objects found on the other. In this setting, one has different preferences for metal detecting, valuing the activity (e.g. for an environment, nature, passion for history or curiosity), which may support the overall hobby motivation for metal detecting. On the other hand, valuing metal detecting for the value of the finds found could mean that ownership or possibly investment motives form a basis for this hobby.

Moreover, we can connect the latter observations to the collecting part described earlier. If one had a greater utility from finding a valuable find than from the activity itself, one might enjoy the collecting part more than the actual discovery part. Therefore, building on this trade-off between the enjoyment of finding and the enjoyment of ownership of the finds, each individual likely has unique preferences, which we try to identify by tracking the submission rate of finds to the archaeological authority, and demographic characteristics of individual metal detectorists (see the chapters on data collection and methodology for detailed discussion). This could give us an idea of what the real driving force is, whether it is, on average, finding or the vision of ownership, i.e., we can distinguish between the enjoyment of the hobby itself - hobby motivation - and the enjoyment of collecting items either for consumption or investment motives.

For this purpose, and due to the properties of the data used, we estimate several models with five different datasets. We use mainly an Ordinary Least Squares (OLS) and the limited dependent variable, Probit and Logit, models. Those plausibly model the respective preferences and simultaneously fit the nature of the used data as well. We compare the results of all the models to check for their robustness in estimating the relationship between the change in the rate of submission of metal detector finds and the respective regional demographic characteristics. The latter contains most importantly the average wealth in the residing region, the value of an owned metal detector, and the availability of archaeological sites in the residing region of an individual.

The method of analysis used in this text has very likely never been used before in a similar setting, since, to the author's knowledge, there is no academic literature studying the ownership and non-ownership motives for metal detecting. The remaining literature that was found, written on metal detecting topic, builds on questionnaire surveys, that are not only consisting of relatively small samples with a maximum of hundreds of observations (e.g. Hardy (2017), Winkley (2016)), but might also suffer from biases caused, among others, by non-disclosure of certain information by questionnaire participants. This is supported for example by Hardy (2017), or by Komoróczy (2022), who mentions the "overrepresentation of those detectorists who have already established a connection to archaeology and accepted many of its prevailing paradigms" in such questionnaire surveys (Komoróczy (2022), p.3). Therefore, the main contribution of this study might be not only filling the gap in the knowledge of metal detecting motivations but also correcting for the plausible bias of other studies in the topic area.

The thesis is structured as follows; Chapter 2 presents the reader with an overview of the existing literature on the topic; Chapter 3 describes the metal detecting hobby characteristics - namely history, differentiation within the metal detecting community, the current conditions of the hobby in the world, and finally the current state of the metal detecting hobby in the Czech environment; Chapter 4 describes how the data were obtained, processed, and what are their key characteristics. Chapter 5 provides a reader with a description of the datasets, methods and models used; Chapter 6 comments on the results of the analysis; and finally Chapter 7 puts the results into the broader context, compares the analysis to other studies, and proposes the topics for further examination.

Chapter 2

Literature Review

The topic of metal detecting motives is generally considered significantly understudied (Thomas (2016); Hardy (2017)), leading to a debate, that is consequentially "conducted in an equally opinionated and irreconcilable manner" (Huth (2013), p.133). This lack of knowledge might seem surprising since collecting and artifacts hunting has followed humanity for centuries, and professional archaeology itself developed from an activity nowadays considered as 'artifacts hunting', in history represented mainly by 'antiquaries', 'travellers' and 'tomb riders' (Glyn (1967), p.56). This claim is further supported by Taylor (1995), that similarly stresses the transition from the eighteenth-century 'antiquaries', through the nineteenth-century 'amateurs' to the twentieth-century qualified 'professionals'.

Reeves (2015) observed that the possible cause for the lacking information about metal detecting practices and motives might be the stigmatization and oversimplification of views of metal detectorists by archaeologists. Simultaneously, Winkley (2016) mentions that for too long the approach to metal detecting has been too narrow because of agendas focusing on who owns cultural heritage objects and who can best preserve them to serve the public. Additionally, ethics and specific interpretations of professional norms may be other limiting aspects of an objective study of artifact collectors and metal-detecting communities (Pitblado (2014)).

Subsequently, the prevalence of existing studies regarding metal detecting attempts to estimate the impact on cultural heritage (Navrátil (2015)), thus illustrating the effect of metal detecting especially on archaeological knowledge;

focusing namely on the extent of artifacts looting (Hardy (2017)), connecting it to the antiquities illicit trade (e.g. van der Lande (2021); Thomas (2016)), the emotional background of metal detecting activity (Moilanen (2023)), as well as the relationship with the landscape (Winkley (2016)) or other more specific cases, such as 'dark heritage tourism' (Koskinen-Koivisto & Thomas (2017)). Any case of literature focusing on non-ownership versus ownership motives in a metal-detecting hobby has not been found.

Nevertheless, there exists literature discussing not only, but also the motivations, their development and why the specific hobbies are popular. For example, Maines (2009) describes in detail the needlework pastime, and Codignola & Mariani (2022) focuses on the art collecting hobby, which is related to the topic analysed in this study.

As far as Czech literature is concerned, it reflects the state of world literature on the subject, but it deals with it much more narrowly. For example, Krásný (2014) describes the issue of metal detecting and archaeology, among other things, estimates the then state of detector activity, proposing the counts of active detectorists to account for fifteen to thirty thousand in the Czech Republic. Navrátil (2015) further broadens this topic with examples of archaeological sites in the Czech Republic that have been the target of so-called 'artifacts looting' by detectorists. The general overview of the current state in the Czech Republic, with the opinions on the topic from both sides, the professional archaeologists and metal detectorists, can be found, for example in the Handbook of the Amateur Archaeologist by Hajšman *et al.* (2019).

To sum up, the different literature is stressing the opposite claims, on the one hand, that the value of the metal detecting hobby lies in the collecting (and financial) motives (e.g. Ofiu (2013)) and on the other that the primary motives are the experience of metal detecting, either for the relationship with the landscape and local history (Winkley (2016)) or, for the excitement from the moment of discovery of a find (Moilanen (2023)). Overall, both sides conclude, that despite relatively recent attempts to undercover the motivations for metal detecting, it has not been studied to a sufficient extent (Hardy (2017)). The overall current state of the knowledge on the metal detecting motives is relatively well summarized by Moilanen (2023), that notes that metal detector "finds are valued in several ways: chronologically, financially, and historically".

Moreover, the author adds that the valuation of artifacts "is likely influenced by different factors such as the social background and educational level of the detectorist - aspects that were not studied in detail (· · ·) Understanding how the finds are valued and how the past is experienced is important, for example, when working with the detectorists."

Finally, the prevalence of the literature concerning metal detecting comes from Great Britain and North European Countries that have their specific regulations and other characteristics regarding the metal detecting hobby. Hence, our study conducted on the dataset representing the metal detecting community in the Czech Republic might be another enriching contribution to the overall discussion, but perhaps more importantly, might form a basis for discussion about metal detecting and its motives in the Czech Republic.

Chapter 3

The Metal Detecting Hobby Characteristics

The metal detecting hobby is a complex phenomenon with varying characteristics in different countries worldwide. In order to understand the beginnings, causes, and current situation of this phenomenon, this chapter provides a general overview of the metal detecting hobby, starting with its definition and history. It also includes a description of the issues that arise, different points of view on the motivation for the hobby from the main committed parties, and a brief overview of the current situation and legislation in the Czech Republic.

3.1 What is Metal Detecting?

Metal detecting definition can be simply a targeted search for metal objects using a metal detector since metal detectors are instruments locating buried metal (Tite (1972)). Metal objects can take many forms, such as artifacts, coins, or precious metals such as gold nuggets. Each metal detectorist focuses on one or more of the respective items and selects the appropriate detector accordingly (Moltaš (2007)). Thus, some people prefer to search, for example, for coins, some for artifacts, such as various jewellery, and some for golden nuggets, among others. Furthermore, different individuals prefer different locations - some like to go in fields, some in forests, public parks, or along popular paths, whereas some can focus on the occasions when ponds or other water reservoirs are drained (Addyman (2009); Moltaš (2007)). Metal detecting on beaches is very popular as they are busy places that are frequented by many people. Observing the metal detectorist groups and respective finds on websites and social media, there is also a significant group of detectorists focusing on the search for militaria, especially World War II seems to be very popular. This observation is also supported, for example, by Moltaš (2007) who claims that militaria, that is usually military badges, are one of the easiest targets to spot when metal detecting. Last, but not least, some detectorists focus on objects buried deeply or shallowly and also on objects larger or smaller. It is easy to imagine that some might be interested in all of the above, and so they may be one day looking for small jewels on the beach, and the other day walking in the woods looking for, for example, World War I militaria.

3.2 History of Metal Detecting

The first written records mentioning treasure hunting date back to the Middle Ages in the form of permission to search for treasure in castle ruins or buried in the ground. Such a permit or contract specified what share of the value of the potential find would go to the finder and what share to the owner of the estate. Also, if anyone was caught digging for treasure without the permission of the estate owner, they were punished accordingly (Moltaš (2007)).

At that time, supposedly, a wand was used to locate treasure (Moltaš (2007)). Research in this area did not cease, and so the first working metal detectors appeared in the mid-1830s (Cornelison & Smith (2009)). However, according to Moltaš (2007), the first dated record of a device used to locate metals in the ground dates to May 1879 when a device named the Hugnes Induction Balance Metal Detector sent sound signals into the ground and picked up the response with a microphone. Two years later, Alexander Graham Bell used a personally developed metal detector to look for the bullet in the body of U.S. President James A. Garfield, who was shot by an assassin (Roberts (1999)).

The use of the metal detector is reportedly further recorded in Harvey de Montmorency's autobiography, Sword and Stirrup, which describes, among other things, the search for the then-famous pirate treasure on Cocos Island in the Pacific using a metal detector. This device was manufactured by the London Electrical Ore Finding Company and was patented in 1903 as the first of its kind in both the US and the UK. This device worked on the basis of pulse induction, using loops of cable laid on the ground as coils and a generator mounted on a nearby car as a power source (Moltaš (2007)). Furthermore, the development of metal detectors was significantly influenced by the Radio Metal

Locating Company (USA), which was the first to produce metal detectors for the market. Its products were known as ‘Radio Locators’ and had the dual coil design of today’s depth metal detectors. Another innovation came from the Goldak Company with its RadioScope detectors, which evolved into the so-called Pancake Detectors during the 1930s, which already had the now familiar form of circular search coils. It was at this time that the first treasure hunters began to use the detectors, and the first official treasure-hunting club was even founded in France, called the French Treasure Seekers Club. Its president, Robert Charroux, declared at the time, interestingly, that “The treasure seeker goes on expeditions primarily because he craves adventure” and furthermore, “The treasure seeker lives for the joy of possible discovery, for the love of the fantastic and the supreme thrill of a few moments.” In addition, in the 1930s, metal detecting was popularized by, among other things, photographs of treasure found while searching on the beaches of islands in the Caribbean Sea, that were shown in the press (Moltaš (2007)).

The first known successful hunt for a buried treasure with a detecting device was reported by James Young of the New York Times in 1927 when one American and two Englishmen found gold chains, jewels, and plates from pirate hoards in Panama (Roberts (1999)).

During World War II, significant developments in metal detector technology occurred in response to the need to locate mines and bombs. Operators in mine clearance units using electromagnetic survey tools encountered not only mines and bombs but also other artifacts and over time learned to identify them (Addyman (2009)). That brought more attention to these instruments and very soon they were used to locate buried ancient metal artifacts (Tite (1972)). Thus, the real use of metal detectors as a hobby did not emerge until after World War II, when the great demand for metal detectors during the war led to the sale of the resulting surplus of thousands of detectors in Europe and North America. The sale of these detectors at prices ranging from \$5 to \$50 created a new group of experimenters and treasure hunters (Roberts (1999)).

However, since the World War II devices were not user-friendly, for example, the weight of the German Berlin 40 metal detector was 12 kg, and since the supply of these devices soon dried up, the development of detectors continued mainly in amateur designs (e.g., Addyman (2009); Moltaš (2007)).

With the major technological innovations in metal object location systems during the 1960s, enabling, for example, the discrimination of ferrous and non-ferrous objects, came the boom of metal detecting as a hobby in the 1970s.

Detectors were marketed to the general public for use in locating lost objects or ancient artifacts, often even using lures to discover buried treasure (Cornelison & Smith (2009); Addyman (2009)). These efforts, coupled with technological improvements and the lower cost of detectors, have led to a rapid increase in the number of metal detector users (Addyman (2009)).

Hence, from the 1970s and 1980s onwards, metal detecting gradually became extremely popular, leading to the founding of several hobby magazines. Fantastic finds also attracted the media, which highlighted aspects of treasure and tales of fortune smiling on 'small people', creating a feeling of something for nothing (Addyman (2009)).

This has led to various campaigns and legal regulations of metal detecting in many countries. For example, nowadays, in France or Northern Ireland, detecting without a license is completely banned. In England and Wales, one needs the landowner's consent to search. However, in the two latter countries, there simultaneously exists a system called the Portable Antiquities Scheme (PAS) that documents finds recorded by metal detectorists (Thomas (2009)).

3.3 The Metal Detecting Dilemma

Following on from the previous chapter, the popularisation and subsequent massive expansion of metal detecting have created a problem: there has been massive destruction of the archaeological context, i.e., the context in which a particular object is found. Already for decades, until the present day, this has been an issue since artifacts are elements of the complex data on archaeological sites which, when studied together, can create a picture of past human activity. Thus, artifacts that are recorded via careful stratigraphic excavation are crucial for defining the date, nature, and former use of a site, the social and economic status of its former inhabitants, or their rituals and burial customs. However, the practice of metal detectorists does not account for the context of the objects found since holes dug by them not only remove the objects that are crucial for understanding the site but also destroy all other evidence that could allow archaeologists to put together a complex story (Addyman (2009); Kobylinsky & Szpanowski (2009)).

Therefore, contrary to Austin (2005) that claims that "Metal detecting is first and foremost a legitimate recreational hobby", there might naturally arise a question if, and to which degree, the metal detecting hobby is actually legitimate. We could see it in a way that as society consists of individuals that favour the hobby, the hobby is legitimate. However, this way, society as a whole might be heading towards the potential destruction of the most information about the past that we do not know about yet. We could compare the issue, for example, to the extraction of natural resources such as coal or oil, which are likely to be, for the most part, extracted in the future, or to climate change. In those examples, we might also consider the activities leading, for example, to the latter as legitimate; however, the result, the actual climate change, is not likely of common interest. That is, metal detecting as an activity might be legitimate; however, the results of this activity might not be, since the target of the activity (objects of the past) might be considered something similar to a non-renewable resource.

On the other hand, metal detectorists believe that their activity helps to preserve objects that would otherwise be destroyed due to agricultural activity, and therefore they are not destroying the historical record but rather helping to protect it. Deep ploughing over recent years has caused enormous damage, disturbing objects buried in the ground, and many would have been lost without metal detecting (Redesdale (2008)).

Thus, the view of archaeologists that they should be the only ones with access to the remains of the past and the view of metal detector users that they have the right and freedom to search for objects on agricultural land outside protected areas are in direct conflict (Redesdale (2008)).

3.4 Metal Detecting Motives and Motivations

According to Thomas (2009), archaeologists and metal detectorists are two very different groups of people, though both share a deep interest in the past. Moreover, both archaeology and metal detecting have in common that they are based on hard work with little reward. Thousands of digs are carried out in both archaeological preparation work as well as in metal detecting practice while finding little of historical or financial worth. Moreover, although some archaeologists might protest, they have with metal detectorists a common dream, and that is finding some long-lost treasure. Nevertheless, the hostility between the two groups stems from different points of view about who controls access to the past Redesdale (2008). Therefore, it might not be surprising that archaeologists and metal detectorists have different views on the motives behind the metal detecting hobby.

3.4.1 Archaeologists' Point Of View

According to many professional archaeologists, metal detectorists are a major threat to the exploration of the past. They are at best a big problem, at worst a group of people fostering an antiquities illicit trade for their financial gain (Thomas (2009)).

For example, Kobylinsky & Szpanowski (2009) argue that there are at least two different motivations behind the activities of metal detectorists in Poland. The first is digging archaeological sites for profit. The second one is the desire to possess ancient artifacts, to find them, and to collect them. The authors even try to explain this phenomenon by stating that targeted search on archaeological sites for financial reward stems from the "pauperization of society and the search for any activity which can bring profit."

The evidence from 1970s Britain about treasure hunting (then a popular equivalent to the term 'metal detecting') might help in understanding the views and fears of many archaeologists. One location of the former Roman town in England was allegedly the prime target for people hoping to get rich fast by using a metal detector. At the time, one American air force man allegedly built up a collection of 2,000 objects and brought it back to the US. Businessmen were stopping at the site and walking it over. This led to a dilemma in which it is important to give information on local sites to local inhabitants; on the other hand, they are the ones who make up the groups of metal detectorists and collectors (Addyman (2009)). This dilemma seems to persist until today.

Kobylinsky & Szpanowski (2009), in their article about the state of metal detecting in Poland in 2008, further compare the metal detecting phenomenon to the situation in Latin America, where archaeological sites are allegedly "plundered for subsistence." Moreover, the authors claim that after the collapse of the communist regime, the public lost the fear of the police, contributing to the destruction of archaeological sites. Moreover, they explicitly link the metal detecting hobby to collecting and antiquities trade by claiming that there are collectors creating the market by purchasing illicit antiquities. Interestingly, institutions and archaeologists themselves might create such a market for antiquities by simply buying the found items or by providing professional expertise on the items, thus contributing to raising the commercial value of illicit antiquities (Kobylinsky & Szpanowski (2009); see also Hajšman *et al.* (2019)).

Moreover, some are stressing the role of the so-called nighthawks – metal detector users working on sites illegally (Thomas (2009)). For example, Richards

& Naylor (2009) claim that there are "rich pickings" for such individuals and that it is possible for experienced nighthawks to even make a good living from selling found items.

According to Hajšman *et al.* (2019), the rest of the metal detectorists, i.e., detectorists that are not in the profit-seeking category, seem to be further divided into two major categories. The first group is enthusiasts, i.e., people who search for the sake of searching and do not care about the finds, and if they do, they evaluate them only for their historical value. People in this group especially enjoy the freedom, action, and adrenaline rush that detecting brings. The last category is those interested in military and militaria. This group of detectorists like to search battlefields, especially modern ones. Some of those are willing to submit finds; however, some form their own private collections.

3.4.2 Metal Detecting Hobbyists' Point of View

Although admitting that there exist certain groups of detectorists that are not respectful (e.g. Moltaš (2007); Austin (2005)), either simply for not covering the holes dug up or by not respecting the law and even targeting archaeological sites, the metal detecting hobbyists claim that the primary objective of the majority of people buying the metal detector is relaxation or searching on beaches, and that they are not interested in prospecting archaeological sites. Austin (2005) further supports this by stating that "metal detecting is foremost a legitimate recreational hobby" which, together with the possibility to be "pursued by all, young and old, rich or poor" is one of its prime qualities. Furthermore, metal detectorists have supposedly been trying to cooperate with archaeologists for the past decades, however, often being rejected (Austin (2005); Moltaš (2007); Hajšman *et al.* (2019)).

As may be expected, metal detectorists stress the fact that the majority of them are responsible hobbyists (Austin (2005)), that search outside of archaeological sites, with the permission of the landowner, or search, for example, for modern war militaria. According to Moltaš (2007), the majority of those detectorists that operate lawlessly are apparently aware of their behaviour, since they do not want to take part in any meetings or competitions, and when they do, they do not want to disclose their personal information. This group of detectorists, therefore, should not be substituted for the majority of metal-detecting hobbyists (Moltaš (2007)).

Moreover, there are several arguments that metal detecting is helpful – for example, as stated earlier, by removing the objects from the ploughed soil, where they would have been destroyed due to cultivation. Another example is the discovery of new sites that nobody was aware of before. One example of such a discovery by metal detectorists is the battlefield of Celts and Romans in Great Britain, which the archaeologists allegedly assumed to be in a completely different place (Moltaš (2007)). Furthermore, Spencer (2009) describes how metal detecting improved the state of the art in numismatics. Comparing the period while being a pure numismatist and the period after discovering metal detecting, the author illustrates the vast knowledge improvement of the different mints of coins, their frequency, and places of occurrence, which was in pure numismatic knowledge assumed to be in many cases very different (Spencer (2009)). Hence, there seems to be evidence that metal detecting may prove helpful in various instances. Richards & Naylor (2009) point out that many metal detectorists have much better knowledge of the locations of some sites than most archaeologists, true particularly for the sites rich in metal objects.

To conclude the complex issue of differing views on metal detecting, already in the year 1983, it was suggested that the rise of metal detecting is caused by the fact that archaeology is not appealing enough to people outside the middle classes (Thomas (2009)). Moreover, Hodder (1984) broadened this thought and claimed that campaigns against treasure hunting added to social divisions between archaeologists and the public, which was assumed to have the same views as archaeologists. This assumption was most likely wrong. Therefore, the majority of metal detector users should be regarded as part of the public that is interested in the physical past, rather than as selfish treasure hunters (Thomas (2009)). At the same time, metal detectorists should refine their view that the position of archaeologists is elitist, trying to restrain those outside academia to engage with their passion (Thomas (2009)).

3.5 Metal Detecting in the Czech Republic

The metal detecting hobby has been widespread in the Czech Republic about 20 years later than in western-European countries, such as France or Germany (Moltaš (2007)), meaning that the 1990s were the period of metal-detecting hobby upswing in the Czech Republic. The boom of metal detecting in the country continues from the end of the 1990s until today. In the meantime, it became a popular hobby as well as an important “scientific and social present-day problem" in most European countries (Komoróczy (2022)). The estimates of a constantly expanding metal detecting community in the country account for 20,000 to 30,000 people (Komoróczy (2022); Hajšman *et al.* (2019)).

The resulting massive loss of archaeological data, according to most estimates of about 100,000 artifacts per year, led to many archaeologists deciding to accept and record the finds from metal detectorists, inducing at least a partial rescue (Komoróczy (2022)). According to Komoróczy (2022), no other approach was possible due to the confrontation with thousands of detectorists prospecting the country. Hajšman *et al.* (2019) estimate that thanks to this activity, just about 20 – 30% of all the finds were 'saved'. The authors also emphasize the fact that the items found are 'a common property' – it belongs to all the inhabitants in the region, thus they are 'stolen' from all of them, and not from some 'strangers'. This brings us to the legal issues specific to metal detecting in the Czech Republic.

If we were to describe the law mentioning metal detecting directly, we would end up with nothing at all. However, Czech law indirectly accounts for metal detecting via The State Monument Act from the year 1987, including its amendments. First, according to The State Monument Act, detectorists might be conducting an archaeological excavation without permission, since among other objects, they excavate also archaeological finds (Hajšman *et al.* (2019)). On the other hand, there is no clear definition of the age of such an archaeological find, stating only that it is "an object (or a set of objects) that is leftover evidence of human life and activity from the beginning of its development to the modern age" (Komoróczy (2022)). Not only that, but also the fact that the form and methodology of archaeological excavation are not clearly defined in the law creates confusion. Furthermore, The State Monument Act declares all archaeological finds to be public property, that is, only the municipalities, regional authorities, and the state can become the owners of the discovered object. Hence, no private personas should possess archaeological finds; what is more, only those who discovered the object accidentally are eligible for the potential financial reward (Komoróczy (2022); Hajšman *et al.* (2019)). As metal detecting is a targeted search for metal items, the discovery by means of a metal detector cannot be considered a random or accidental find, thus one cannot claim a reward when discovering the find with a metal detector (Hajšman *et al.* (2019)). Hajšman *et al.* (2019) further mention one more possible offence when metal detecting, which is damaging the property of someone else, that is either the land or the find itself. Despite this rather restrictive law, there are almost no instances of prosecutions connected to metal detecting compared to the magnitude of metal detecting in the Czech Republic; supporting the stance that in practice, it is not feasible to penalize metal detecting and private holding of archaeological finds in the Czech Republic (Komoróczy (2022); Hajšman *et al.* (2019)). This paradoxical situation already persists in the Czech Republic for about 25 years, from the 1990s until today.

Chapter 4

Data

The data collecting was a crucial and quite extensive part of the analysis process. It was conducted via web-scraping, a rather experimental and not very widely used method in a similar setting. This chapter encompasses a brief overview of the source of the data, the method used to obtain them, and the data pre-processing, including the creation of key variables and their description.

4.1 Data Source

The website used to gather the data for the analysis is one of the two most significant and widely used Czech metal detecting websites (Komoróczy (2022)) with the characteristic name, 'LovecPokladu.cz'. With the weekly traffic of hundreds to thousands of detectorists, it currently contains more than 7,700 profiles of detectorists, over 200,000 uploaded artifacts, and almost 100,000 uploaded coins. Thus, the website can be assumed to be renowned and to a high degree representative of the Czech metal detecting scene and population.

The website was created in the year 2006 as a community website, designated for amateur fans of archaeology and searching with a metal detector. The name of the website, 'LovecPokladu.cz', might have an English equivalent of 'Treasure Hunter' but also 'History Hunter', stressing that it is intended for people who "love adventure and history" and at the same time "realize the importance of cooperation with professional archaeologists".[1] The website also includes a shop with metal detectors distributed through the network of physical stores in the Czech Republic and Slovakia. The website management is

[1]Source: https://www.lovecpokladu.cz/en/kontakt (Retrieved on: June 14, 2023)

composed of active metal detectorists, stressing that it is aware of its social responsibility, therefore organizing many meetings of amateurs with professional archaeologists, as well as supporting the creation of clubs and organizing a Republic Championship, recognized even abroad.[2]

There are many features on the website, such as the possibility to interact with others via chat and online forums, sections with manuals, tests, and reviews of metal detectors and equipment, or interaction within different detectorist clubs. However, the most important for our analysis are the personal profiles of metal detectorists, their uploaded artifacts, and coins. Therefore, below we describe the structure and key characteristics of the webpage parts that contain information of our interest.

4.2 Website Data Description

4.2.1 The Profile-Specific Data

We begin with the description of the data contained in the personal profiles of metal detectorists. From the registration form on the website[3], we are particularly interested in the 'City' column since it is a key variable for obtaining the demographic characteristics of individual metal detectorists. In addition, we utilize the 'Detector used' column, as it may serve as a proxy for the economic status of a detectorist.

After an individual registers on the website using the mentioned form, the registration is manually confirmed by the web administrator within the next 24 hours. After that, the account information can be viewed. An example of such an account/profile can be viewed for example here.[4] The specific variables of our interest from the profile page are (going from top to bottom of the profile respectively):

- Profile (*'profile'*) – a nickname of an individual entered upon registration
- City (*'residence'*) – the variable containing the municipality an individual resides in
- Detector (*'detector'*) – the variable containing the detector used by an individual

[2]Source: https://www.lovecpokladu.cz/en/kontakt (Retrieved on: June 14, 2023)
[3]Registration Form Available at: https://www.lovecpokladu.cz/en/registration
[4]An Example Profile Page: https://www.lovecpokladu.cz/uzivatel/kovboj78

- Experience (*'experience'*) – an individual score that reflects the successful identification of various finds that are uploaded on the website, reflecting not only knowledge but also activity
- The Number of Articles or Club Posts (*'contributions'*) – reflects the activity of an individual in the club (sharing stories, curiosities, experience) or similarly via articles on the whole website, possibly reflecting the integration on the website through active creation of new content (not necessarily related to finds)
- The Number of Total Comments (*'comments'*) – the number of comments that an individual provided under the contributions (e.g., artifacts, coins, club posts, articles) on the website (i.e., comments further specifying or explaining the history of the finds, appreciatory comments), possibly reflecting the integration on the website via supporting others to create new content
- The Number of Artifacts (*'artifacts'*) – the total number of artifacts uploaded on the website by an individual
- The Number of Coins (*'coins'*) – the total number of coins uploaded by an individual on the website
- More Profiles (*'link'*) – the variable indicating if one filled links to other personal profiles, for example, 'facebook.com', 'youtube.com', or personal website. However, this variable is not present in both the initial registration form as well as the example profile provided above. This is due to the fact that this box can be filled only additionally in the settings of an already established profile under the name 'Social Media'.

4.2.2 The Find-Specific Data

Furthermore, we are interested in the data of uploaded artifacts and coins. The coins, when uploaded, usually include more information than artifacts, as generally, they are more consistent in shape and more easily identifiable. Nevertheless, since we use the information that is common to both coins and artifacts, serving for consistency of the analysis, it is sufficient to describe the properties of uploading artifacts only. An example of an uploaded artifact to the website can be viewed here.[5]. In the case of individual finds (artifacts and coins), we are interested in the following variables:

- Location – the region where the item was found. Not present in the example provided in the form of a link, since it is necessary to select the country first in the uploading process. First after that, one can select the specific region of the Czech Republic. (An example of a constraint leading to missing data)
- Detector used – the information about which detector was used for the discovery of an object
- Submitted to – a crucial variable indicating if the find was submitted and to which archaeological institution
- Votes – the number of likes assigned to the specific find by other users – indicating how interesting the find actually is (in comparison to the number of views below)
- Viewed – how many times the find was viewed – possibly indicating if the find looks interesting
- Comments – the number of comments under the uploaded find – possibly indicating that the find is interesting, difficult to identify, or otherwise controversial
- Finder – indicating who is the finder of the item, i.e. the profile nickname
- Photographed – the date when the photography of an object was taken
- Uploaded – the date when the actual object was uploaded to the website

4.3 Data Extraction and Processing

4.3.1 Data Extraction

The data for our analysis were taken from the above-described website 'lovecpokladu.cz' using the web scraping method. Web scraping can be defined as "the construction of an agent to download, parse, and organize data from the web

[5] An Example of the Uploaded Artifact: https://www.lovecpokladu.cz/artefakty/nalez/stredoveke-pecetidlo-242087/

in an automated manner" (vanden Broucke & Baesens (2018), p.3). In other words, one can either copy and paste interesting parts of a website into a spreadsheet manually or, can leave this task to a computer program that can execute the task much faster and more accurately. This might be especially true for large amounts of data whose manual retrieval might take too long. In our case, we needed hundreds of thousands of observations of uploaded artifacts, coins, and individual profiles, each containing several variables. Thus, precise manual collection of the data might have not only been exhausting but also likely impossible to do in a reasonable amount of time.

Nevertheless, the creation of such a 'web scraper' mostly requires competencies in different fields such as web technologies, authentication strategies, regular expressions, text parsing, different encoding systems, or efficient data storage, to mention a few. Hence it might be far from being a single task to do (Iacus (2015)). It was no different in our case, requiring knowledge and understanding of Hypertext Markup Language (HTML), programming, and the corresponding libraries, packages, and modules. For the creation of the web scraper, we used the Python[6] programming language, and additional libraries for Python, most importantly BeautifulSoup, requests, RegEx, openpyxl, and pandas. The actual code used for web scraping is available in the Appendix in the form of a link to a GitHub repository.

4.3.2 Legal Issues

As web scraping is a form of information retrieval and information is a means of building intellectual capital, it is vulnerable in many ways - concerning privacy, accuracy, property, and accessibility (Mason (1986)). Therefore, it might be a good practice to check for potential legal issues regarding data retrieval first. In our case, we went through the Privacy Policy of the website making sure that the publication of the data does not violate any of the provisions. The conditions even explicitly state that the finds uploaded on the website serve for mapping purposes and that their deletion from the database is hence not possible (except obvious exceptions - items with no meaningful value, poor quality photographs). Additionally, we checked the website for the possibility of not allowing web scraping. This can be done by attaching 'robots.txt' to the website Uniform Resource Locator (URL) address and checking the result

[6]More About Python: https://www.python.org/

when opened (Mazilu (2022)). When applied to the website 'lovecpokladu.cz', there was no restriction, thus we assumed compliance with the web scraping.

4.3.3 Data Processing

The data obtained via web scraping described above had a raw form. In order to be suitable for actual analysis, they needed to be cleaned. For those purposes, mainly the Pandas library in Python was used. The data cleaning part was a lengthy process including adjustment of variables, creating dummy variables, and merging data frames. Next, we describe the key data manipulation steps taken and the definitions of the respective variables.

Creation of the Main Dependent Variable – The Submission Rate

The submission rate, i.e., the share of artifacts/coins that were submitted to the archaeological institution, will be our key dependent variable. To obtain it, we first group all the artifacts (in the artifacts-specific dataset) by the name of the profile that uploaded them. Then, we sum up the number of artifacts that do not have a missing value in the 'Submitted to' column and assign this number to the respective person. By doing so, we get the number of artifacts submitted for each profile. Now, we simply divide the number of submitted artifacts by the number of total artifacts the respective person uploaded. For the profiles that have not submitted any artifacts yet, we have the total number of artifacts as zero, thus the final submission rate is missing (since it is not possible to divide by zero). Hence, the main dependent variable of our analysis, the submission rate, is defined as follows.

Submission Rate of Artifacts:

$$artifs\_rate = \frac{\text{number of submitted artifacts}}{\text{number of all artifacts}}$$

Submission Rate of Coins:

$$coins\_rate = \frac{\text{number of submitted coins}}{\text{number of all coins}}$$

These submission rates will be critical in understanding the level of engagement of metal detectorists in submitting their finds to archaeological institutions. A higher submission rate would indicate a more hobbyist motivation for metal detecting.

Creation of Dummy Variables

The first dummy variable created is the ***link*** dummy variable. This variable replaces the old variable contained in the raw data by assigning 1 if one made an effort to go to the settings of the profile and uploaded the link to other profiles like youtube.com or facebook.com. Otherwise, this variable is 0. In other words, the *link* dummy variable takes a value of 1 if the value in the original raw 'link' column is not missing, and 0 otherwise:

$$link = \begin{cases} 1 & \text{link shared} \\ 0 & \text{link missing} \end{cases}$$

The second dummy variable is called ***residence_additional_info***. This variable was created from the column 'residence'. When someone uploaded more than one city (e.g., 'Praha, Děčín') or uploaded the city and further specified the city district (e.g., 'Praha, Ruzyně'), this variable was assigned 1, and 0 otherwise:

$$residence\_additional\_info = \begin{cases} 1 & \text{more cities or district specified} \\ 0 & \text{just single residence uploaded} \end{cases}$$

The third and most important dummy variable is the ***detector_expensive_dummy***. This variable was created using a list of metal detector names with prices above 30,000 CZK. The list was manually created from the shop of the website 'lovecpokladu.cz', the prices comparison web 'heureka.cz', and confirmatory research of other shops operating worldwide, together with the metal detecting producers' websites. Nevertheless, since in our main dataset, the 'detector' column included imprecise entries from users (specifically 3,883 different entries - see Table 4.1), it was not easy to match them to the exact names of detectors obtained via manual collection. Therefore, only the model names (without the producer's name) were used as a base list (Table 4.2), whose letters were further lowered and put together. The same method of lowering and putting together all the letters was applied to detectors in the 'detector' column of our dataset. A specific function was then created to generate a new variable assigning 1 if at least one of the adjusted model names from the list was included in the adjusted 'detector' used column, and 0 otherwise. That means, 1 was assigned whenever the word from the list appeared in the 'detector' column, regardless if in the same column was written anything else. In other words, if one declared in the 'detector' column using one of the detectors from the list, the *detector_expensive _dummy* would take the value of 1, and 0 otherwise:

$$detector\_expensive\_dummy = \begin{cases} 1 & \text{detector used being in the list} \\ 0 & \text{detector used not in the list} \end{cases}$$

Order	Value	Count
0	Equinox 800	126
1	Equinox 600	90
2	XP Deus	84
3	Vanquish 540	70
4	Minelab Equinox 800	62
...	...	...
3879	TEJON	1
3880	Zero lp II	1
3881	Rutus proxima	1
3882	Simplex +	1
3883	XP-250	1

Table 4.1: Metal Detector Data

'Expensive' Detectors		
Manticore	GTI 2500	SDC 2300
CTX 3030	Axiom MS2	ATX
GPX 5000	Axiom MS3	SSP-5100
Excalibur II	GPX 6000	UPEX ONE 2
Standard MP V2	GPZ 7000	GPX 4500
Standard MP V3	Spectra V3i	Invenio PRO

Table 4.2: List of 'Expensive' Metal Detector Models

In 2018, 30,000 CZK was approximately the average gross wage in the Czech Republic.[7] Therefore, we considered this number to be possibly significant enough to serve as a threshold for distinguishing between expensive and non-expensive metal detectors. By 2018, the website lovecpokladu.cz was already

[7]Source: https://www.czso.cz/csu/czso/cri/prumerne-mzdy-4-ctvrtleti-2018

well-established, with over 120,000 artifacts uploaded. The upload of artifacts on a larger scale seems to have begun in the year 2012, so 2018 lies approximately in the middle from 2012 until the present day. Considering the current number of slightly more than 200,000 artifacts on the website, we might imply that the intensity of uploading perhaps peaked close to the year 2018. Hence, we assume that the year 2018 is a good reference point for our analysis, mainly since the website was already well-established and profound, and included a significant number of individuals uploading their finds on a daily basis.

Economic and Demographic Data

1. Real Net Monetary Index

The first and main variable used for the analysis is the *real_net_monetary_index*. This variable, along with the *detector_expensive_dummy*, serves as one of the key explanatory variables. We obtained the *real_net_monetary_index* variable from the work of Kocourek *et al.* (2021) entitled 'Money Income and the Cost of Living of the Population: A Detailed View of the Czech Republic'. The real net monetary index used in our analysis is constructed as a nominal net monetary index adjusted to the regional price index. The resulting numbers for different regions vary from 0.77 to 1.25, depending on the purchasing power of a person with a permanent stay in the region. The number 1 represents the average of the Czech Republic, allowing for the comparison of different regions.

For example, an average citizen living in the region with the highest index (1.25) can buy about 62.34% more goods and services than a person living in the region with the lowest real net monetary index (0.77). This is calculated by $1.25/0.77 = 1.6234$. The real net monetary index thus reflects the amount of goods and services that an average person residing in a specific region can buy for their net income, representing the regional purchasing power. Compared to a simple average income, the real net monetary index provides a more precise indicator of the economic situation of citizens in different regions of the Czech Republic (Kocourek *et al.* (2021)).

It is important to note that the real net monetary index was created as a moving average of the years 2017 - 2019. Therefore, also to maintain consistency with this timeframe, we use demographic variables from the year 2018.

For the analysis to be as precise as possible, we used the smallest regional unit available, which is an 'Administrative District of Municipality with Authorised Municipal Authority' (from now on 'municipal office'). There are currently

393 of these administrative districts in the Czech Republic[8], providing greater data variability compared to other available regional units.

The respective data for the variable *real_net_monetary_index* of the 'municipal offices' along with other variables, were downloaded from the database of the Technical University of Liberec.[9]

Nevertheless, since our main dataset contained data only for the municipalities that people reside in, we needed to handle the correct matching of those individual municipalities with their corresponding 'municipal offices'. There are about 6,254 municipalities in the Czech Republic, so we conducted additional web scraping of a 'Wikipedia article' containing the list of all municipalities in the Czech Republic[10] together with regions and districts they belong to. Moreover, we needed to scrape the data from the portal 'epusa.cz'[11], which contained not only the municipalities but also the desired respective 'municipal offices', unlike Wikipedia.

After merging those two new datasets, we had the dataset containing the specific municipality, corresponding region, district, 'Administrative District of Municipality with Extended Powers' and 'Administrative District of Municipality with Authorised Municipal Authority' ('municipal office'). All those variables may be used for possible merging with any other dataset containing the desired variable (monetary index, demographic data, etc.) on any of those levels and subsequent analysis on that level. Particularly of our desire was the 'municipal office' level.

Finally, we merged our main dataset containing individual detectorists' profiles and their 'residence' with the variable describing the purchasing power of an average citizen (*real_net_monetary_index*) of the certain 'municipal office' to which the residence/municipality belongs. An example of the part of the merged dataset can be seen in Figure 4.1 (with 'municipal office' being the third column from the left named here as 'municipal_authority').

We might notice in the example of the merged data (Figure 4.1) that some of the residence entries by profile users were matched with the respective municipalities, but some were not. This was another issue, similar to the 'detector_expensive_dummy' variable creation, caused by imprecise entries of resi-

[8]Source: https://www.czso.cz/csu/rso/poverene_obecni_urady (Retrieved on: May 25, 2023)

[9]Source of the Data: https://opendata.tul.cz/search?tags=nominální příjmy

[10]Source of the Municipalities: https://cs.wikipedia.org/wiki/Seznam_obcí_v_Česku (Retrieved on: May 25, 2023)

[11]Source: https://www.epusa.cz/ (Retrieved on: May 25, 2023)

first_residence	municipality	municipal_authority	municipality_with_extended_powers	district	region	real_net_monetary_index
[]						
Vochov	Vochov	Město Touškov	Nýřany	Plzeň-sever	Plzeňský kraj	1.074743972
u kolína						
Praha	Praha	Hlavní město Praha	Hlavní město Praha	Praha	Praha	1.131666
Česká Lípa	Česká Lípa	Česká Lípa	Česká Lípa	Česká Lípa	Liberecký kraj	0.954653183
Hodonín	Hodonín	Hodonín	Hodonín	Hodonín	Jihomoravský kraj	0.923956911
Světce	Světce	Jindřichův Hradec	Jindřichův Hradec	Jindřichův Hradec	Jihočeský kraj	0.949712396
HK						
doprava						
[]						
BRNO	Brno	Brno	Brno	Brno-město	Jihomoravský kraj	0.960725944
křidluvky	Křídlůvky	Znojmo	Znojmo	Znojmo	Jihomoravský kraj	0.91621004

Figure 4.1: An Example Part of The Merged Dataset

dence by users of the website lovecpokladu.cz. To handle the right merging of the dataframes, the original imprecise entries were divided to separate columns by comma and semicolon, handling some cases where there were multiple entries. Further only the column 'first_residence' was used (as can be seen in Figure 4.1). The other resulting column was used to create a dummy variable *residence_additional_info* as described in the section on creating dummy variables. Yet, the 'first_residence' column was lowered first (to handle cases like 'BRNO' in Figure 4.1), the special Czech characters were replaced with their equivalents without punctuation (to handle the case of 'křidluvky' in Figure 4.1), and finally, all the words and letters merged together (to avoid the possibility of accidental space, for example in the case of names comprising of two words like 'Česká Lípa'). First after applying the same process on the 'municipal_authority' ('municipal_office') column, we merged the dataframes on those two adjusted columns. Hence, this was an attempt to match as many municipalities with the *real_net_monetary_index* as possible. The result can be partly seen in the example of Figure 4.1.

2. Further Demographic Data – Men Proportion, 65+ Proportion, Average Age

The other demographic variables used in our analysis were obtained from the Czech Statistical Office.[12]

The obtained variables are namely *men_proportion*, defined as the total number of men in the 'municipal_office' divided by the total number of citizens in the respective 'municipal_office'; the *65+_proportion*, defined as the total number of people older than 65 years in the 'municipal_office' divided by the total number of citizens in the respective 'municipal_office'; and finally, the

[12]Source: https://www.czso.cz/csu/czso/demograficka-rocenka-spravnich-obvodu-obci-s-poverenym-obecnim-uradem-2012-2021 (Retrieved on: May 25, 2023)

average_age, which was taken from the dataset without modification. All those variables were merged on the 'municipal_office' column to the main dataset.

Archeological Localities Rate

The final variable called *localities_rate* was more complex to create, since no data about the number of archaeological localities in the 'municipal_offices' were found. Hence, we manually collected the data from the map displaying significant archaeological localities as points (Figure 4.2).[13] We counted the number of points in each of the respective 'municipal offices' displayed on the map and wrote down the results into an MS Excel sheet, creating a new variable 'number_of_localities'. Furthermore, in order to account for the different sizes of different 'municipal offices', we used the data from the Czech Statistical Office[14] about the area of each of the 6,254 municipalities in the Czech Republic since the needed data about the area of individual 'municipal offices' were not found. After summing up the areas of municipalities respective to their 'municipal office', we got the overall area of each of the 'municipal_office'. Finally, we divided the number of archaeological localities by the area of the respective 'municipal_office' (in hectares) and obtained the variable *localities_rate*, serving as a proxy variable for the density of archaeological localities in each of the 'municipal offices'.

4.4 General Dataset Characteristics

The final dataset used for the analysis of metal detecting ownership and non-ownership motives consists of 7,728 observations of profiles of individual metal detectorists. The other two datasets contain observations of 203,846 artifacts and 94,269 coins, respectively. However, the key characteristics of the datasets are further divided into different sections, each describing the specific subset of the main datasets used for different kinds of analyses. The division into more separate analyses is mainly due to the significant number of non-matched municipalities (Datasets 1, 2, 3, and 4 in the Methodology Chapter),

[13] Source: http://isad.npu.cz/tms/arch_public/index.php?client_type=map_resize&Project=TMS_ARCH_PUBLIC&client_lang=cz_win&strange_opener=0 (Retrieved on: June 16, 2023)

[14] Source: https://www.czso.cz/csu/czso/maly-lexikon-obci-ceske-republiky-2020 (Retrieved on: May 25, 2023)

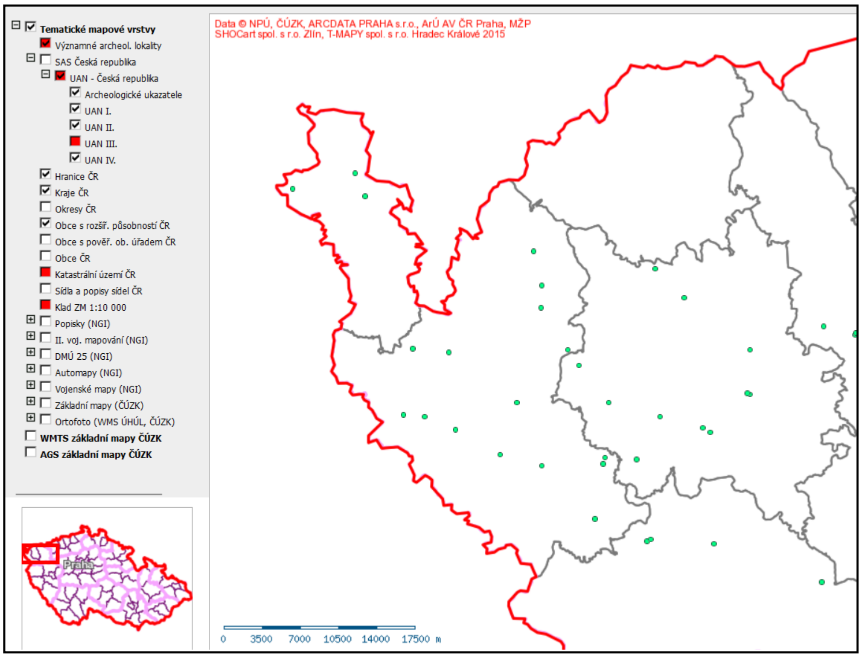

Figure 4.2: Part of the Map With Significant Archaeological Localities as Points

and other concerns such as the significance of the finds uploaded (Dataset 5 in the Methodology Chapter).

We can see the rationale for the analysis division into different parts in the following table (Table 4.3), which shows under the 'Count' column the number of non-missing values in the base (non-manipulated) dataset. First, marked by †, we can see that out of the total of 7,728 observations, 4,216 were successfully matched with the respective *municipal_office*, thus resulting in 3,512 missing observations of the *real_net_monetary_index*, which is one of our key independent variables.

Other key independent variables contain observations for the whole dataset (all 7,728 observations), such as the *detector_expensive_dummy* variable. Therefore, the deletion of all the 3,512 missing observations would result in the loss of a great amount of data. In order to prevent this loss, we fill in the value of an average of the Czech Republic for all the economic and demographic variables. That means, for example, that we fill in the number 1 instead of the missing values in the case of the *real_net_monetary_index* (thanks to the definition of this variable). All filled values can be seen in Table 4.4 below.

In the section describing the creation of the submission rate variable, we discussed the possibility that an individual registers to the website but does not upload any artifact or any coin. The counts of people that uploaded at least

Variable	Count
profile †	7728
link	7728
experience	7728
contributions	7728
comments	7728
artifacts	7728
coins	7728
residence_additional_info	7728
municipality	4230
municipal_office †	4216
real_net_monetary_index	4216
submitted_number_artifs	6677
number_artifs	6677
artifs_rate ‡	6677
submitted_number_coins	5323
number_coins	5323
coins_rate ‡	5323
average_age	4216
rate_artifs_dummy	6677
rate_coins_dummy	5323
uploaded_at_least_one_artif_or_coin_dummy	7728
men_proportion	4216
65+_proportion	4216
detector_expensive_dummy	7728
area_municipality	4216
municipality_type	7728
population_density	4216
localities_rate	4216

Table 4.3: Counts of Non-Missing Values Within the Base Dataset Columns

Variable	Filled Value
real_net_monetary_index	1
average_age	42.4734915211329
men_proportion	0.497149745062897
65+_proportion	0.197914581620951
localities_rate	0.021478453745227236
population_density	135
area_municipality	12.603110956375838

Table 4.4: Filled Missing Values

one artifact or at least one coin, respectively, can be seen in Table 4.3 marked by ‡. There are thus 6,677 observations of people that uploaded at least one artifact and 5,323 observations of people that uploaded at least one coin. This leads to further division of this dataset (with already filled missing values of economic and demographic variables).

First, we keep the observations of profiles that have at least one artifact or coin uploaded, assuming that all the individuals behind profiles with at least one uploaded find are metal detectorists; assigning the respective missing values of submission rates the value of 0. This modification, therefore, raises the count of observations for both the submission rate of artifacts - *artifs_rate*, and the submission rate of coins - *coins_rate* (Table 4.3, marked by ‡) to the overall number of 7,622. Hence, we lost 106 observations of the people with no uploaded artifact and at the same time no uploaded coin. This dataset is used for the analysis of artifacts only (Dataset 1).

Second, we delete the observations for missing coins only, leaving us with the dataframe consisting of 5,323 observations. This dataset is then used for the analysis respective to coins only (Dataset 2).

Third, we create the Datasets 3 and 4, that are used to verify the influence of filling the average values in the first two datasets.

Finally, we create the Dataset 5 that includes data on individual artifacts, serving as a final robustness check.

Chapter 5

Methodology

This chapter provides an overview of our key hypotheses, models and measures. Moreover, it describes each of the five datasets used and their respective key data characteristics along with the estimated models. Lastly, it provides a brief summary of the results and suggests potential drawbacks of our analysis.

5.1 Hypotheses

The purpose of our analysis is to test the following hypotheses:

1. **Main Hypothesis**: With greater individual wealth, the submission rate of finds will be higher.

   This main hypothesis is further divided into two separate hypotheses that are tested in all of the models simultaneously:

   (a) **H1**: The submission rate of finds is increasing in the real net monetary index of an individual.

   The idea behind testing this hypothesis is using the variable *real_net_monetary_index* as a proxy for the wealth of an individual. The key drawback of this independent variable is, however, that it is approximated by the municipal office an individual resides in. In fact, it is likely not the case that each individual has an average real net income in the given area; however, we assume that on average, it is a good approximation since we have a relatively large sample. This hypothesis might be interesting since it could hint if relatively richer individuals tend to submit artifacts and/or coins more. We are interested in this hypothesis since it might reveal that the preferences for submitting the finds (thus non-creating of the collection) differ among different socio-economic groups.

(b) **H2**: The submission rate of finds is higher for individuals who own more expensive metal detectors.

The idea for testing this hypothesis is the same as for the previous explanatory variable. This time we use the variable *detector_expensive_dummy* as a proxy for the socioeconomic status of an individual. This binary variable takes a value of 1 if the value of a given detector is above a threshold of 30,000 CZK. Nevertheless, first, it is likely that the value of the detector an individual possesses is not only a matter of income or wealth but also how enthusiastic one is with respect to the metal detecting hobby. We might look at the correlation matrices corresponding to the respective samples we use and check for correlations of the *detector_expensive_dummy* variable with some adept variables reflecting enthusiasm. The first adept might be *experience*, reflecting the amount of help to others with identifying their finds. The correlation with this variable is, however, almost none, and always negative. The second variable potentially expressing enthusiasm is an active creation of content on the website represented by *contributions* variable, with the correlation with the detector dummy ranging from 0.13 to 0.22 for the different datasets. However, we are mainly interested if the *detector_expensive_dummy* variable is positively related to the *real_net_monetary_index* variable. In all of the correlation matrices, it is very mildly positively correlated, except for Dataset 5, for which the correlation between the two variables is bigger (0.13). By running simple Linear Probability Model (LPM) and Probit models with the mentioned variables as independent variables and the *detector_expensive_dummy* as a binary dependent variable, all the variables are significant at the 5% significance level. And indeed, the *real_net_monetary_index* variable has a large positive coefficient with respect to the other two variables. We cannot rely on such a simplistic approach to regression much; however, it might reflect the overall usefulness of an expensive detector as a proxy for the socioeconomic status of an individual. So overall, it is likely that the wealth of an individual is, along with the enthusiasm, a very significant deciding factor for buying the relatively expensive metal detector; and that is the assumption we do for the *detector_expensive_dummy* variable. That is, higher the socioeconomic status, higher the probability that one owns an 'expensive' metal detector.

2. **H3**: The submission rate of finds is positively related to the density of archaeological localities in an individual's residing area.

   The idea behind this hypothesis is that if one lives in a region with a higher density of archaeological sites, represented by a proxy variable *localities_rate*, one has a higher probability of finding archaeological finds and thus a higher probability of submitting the finds (*'submission rate'*). This hypothesis might be connected to the supposed issue of a gradual destroying of the archaeological sites, hence losing valuable information about the past. In other words, if people living in the area with a higher probability of archaeological finds do not submit those finds more often, it might suggest that those are potentially building a collection of artifacts or coins found, or that it is not more likely to submit a find when having a higher probability of finding it. Nevertheless, we do not take into account the amount of travelling associated with metal detecting which might be a potentially important factor influencing the testing of this hypothesis.

5.2 Main Models

5.2.1 LPM

The Linear Probability Model (LPM) is a model estimated by Ordinary Least Squares (OLS), assuming a linear relationship between the independent variables and the response probability. The model can be represented by the following formula:

$$P(Y = 1|X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

The LPM has some limitations, such as fitted probabilities that can be less than zero or greater than one, and the partial effect of an explanatory variable being constant. These limitations can be overcome by using the following two binary response models (Wooldridge (2012)).

5.2.2 Probit

The Probit model assumes that the probability of the binary outcome follows a standard normal cumulative distribution function. The coefficients in this model are estimated using Maximum Likelihood Estimation (MLE), which maximizes the likelihood of the given outcomes. The model can be written as follows:

$$P(Y = 1|X) = \Phi(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)$$

Where: $\Phi(\cdot)$ is the standard normal cumulative distribution function expressed as an integral: $\Phi(z) = \int_{-\infty}^{z} \phi(v)dv$, where $\phi(z)$ is the standard normal density $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$.

The standard normal cumulative distribution function ensures that the fitted probabilities are strictly between zero and one for all values of the parameters and explanatory variables (Wooldridge (2012)).

5.2.3 Logit

Unlike the Probit model, the Logit model assumes that the response probability follows a logistic cumulative distribution function. The parameters are estimated using MLE, similar to the Probit model. The Logit model can be represented as:

$$P(Y = 1|X) = \Lambda(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)$$

$\Lambda(\cdot)$ is the cumulative distribution function for a standard logistic random variable: $\Lambda(z) = \frac{\exp(z)}{1+\exp(z)}$, which is between zero and one for all $z \in \mathbb{R}$ (Wooldridge (2012)).

In the above equations: $P(Y = 1|X)$ is the probability of the dependent variable Y taking the value 1 given the values of the explanatory variables X (response probability). $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are the coefficients (parameters) to be estimated. X is the set of explanatory variables $x_1, x_2, \ldots, x_k$.

Additionally, a notion of the general approach to obtaining an estimator using MLE can be found for example in Wooldridge (2012), page 630.

The assessment and interpretation of the Probit and Logit models is conducted with the help of measures and statistics described in the Assessment and Interpretation section.

5.3 Regression Assumptions

Correct Functional Form

Our models, particularly the OLS and Weighted Least Squares (WLS), might suffer from potential functional form misspecification. We tried various transformations, including logarithmic, inverse, Yeo-Johnson (a special case of Box-Cox), and Anscombe transformations. However, the logarithmic transformation appeared to be the most suitable. In the case of an incorrect functional form, both coefficients and standard errors might become unreliable. We take this into account when choosing the most suitable model. Moreover, the analysis relies on the properties of a large sample, thanks to the sizes of the datasets used.

Homoskedasticity

As evident from some of the scatter plots, there is likely to be heteroskedasticity. The Breusch-Pagan test also indicates the presence of heteroskedasticity in all the models with non-binary dependent variables. To mitigate the potential issues caused by this assumption violation, we first perform a logarithmic transformation of some variables to treat the heteroskedasticity and normality of the errors, at least to some extent. Additionally, we conduct OLS with heteroskedasticity robust standard errors and WLS. However, even these methods might not handle the possible heteroskedasticity correctly due to the high number of zeros in the sample.

Normality of Errors

The scatter plots indicate that the errors are non-normally distributed. However, since we have relatively large samples, the Central Limit Theorem (CLT) should apply, and the possible drawbacks caused by the non-normality of errors are assumed to be not significantly impactful.

Non-Multicollinearity

Another issue in estimating the models might be multicollinearity. To check for potential multicollinearity in our models, we use and assess the Variance Inflation Factor (VIF). In case it is higher than five for any of the variables, we adjust either the dataset or models so that the potential for multicollinearity is no longer present.

Exogeneity

The models might further suffer from endogeneity. We attempted to account for possible endogeneity by including a variety of variables. Since we do not have suitable Instrumental Variables, we use proxies for potentially unobservable variables contained in the error term. If not treated, these variables might make it hard to assess the relationship of one of the independent variables with the dependent variable.

5.4 Models Selection

Based on the assessment of the variety of estimated models, we decided to choose a narrower set of models that are likely to be consistent and overall fit the data well. First, considering the potential incorrect specification of the models, we conclude that the estimates of the OLS models might not be unbiased, and the associated standard errors (and thus p-values) might not be precise. Despite these concerns, the OLS-associated models performed quite well, with the p-values of the variables being very close to those of the LPM, Logit, and Probit models. Nevertheless, since we considered the OLS-associated models as part of the robustness and sensitivity control, although focusing mainly on the LPM, Logit, and Probit models, we include them in the discussion and description in the following sections.

From the OLS models, we trust the most the model with heteroskedasticity robust standard errors, as its adjusted R-squared is decent (about 0.1), the model is statistically significant at almost any significance level, and the p-values of the key independent variables indicate significance identically to the LPM, Logit, and Probit models. However, we are cautious about the WLS model, as its R-squared is often suspiciously too high, even close to one. The associated p-values of variables often vary significantly from other models, particularly for the key variables *real_net_monetary_index* and *localities_rate*.

Additionally, the condition number of the WLS models is often large, indicating high sensitivity to small changes in the data.

Hence, we decided to choose the LPM, Logit, and Probit models, as they seem to be more consistent in estimates compared to each other and might mitigate, though not fully, some of the issues of the OLS models, such as potential incomplete reporting of the submission rate and/or potentially incorrect functional form. The LPM, Logit, and Probit models had the same results for all datasets with respect to rejecting or not rejecting the hypotheses about the main independent variables, which are *real_net_monetary_index*, *localities_rate*, and *detector_expensive_dummy*.

Although the LPM is a simple model with the potential for predicting probabilities outside the [0,1] range, we use it along with the Logit and Probit models for comparison purposes. Next, we provide measures and statistics used for the evaluation of the main estimated models.

5.5 Assessment and Interpretation

5.5.1 Goodness of Fit

McFadden's pseudo R^2

McFadden's pseudo R^2 is defined as $1 - \frac{L_{ur}}{L_o}$, where L_{ur} is the log-likelihood function for the estimated model, and L_o is the log-likelihood function in the model with an intercept only (Wooldridge (2012)). An example of the derivation of a log-likelihood function can be also seen for example in Wooldridge (2012), page 630.

Likelihood Ratio Statistic (LR)

The Likelihood Ratio (LR) Statistic is defined as twice the difference in the log-likelihoods:

$$LR = 2(L_{ur} - L_r)$$

where L_{ur} is the value of the log-likelihood for the unrestricted model, and L_r is the log-likelihood value for the restricted model (Wooldridge (2012)).

The **Log-Likelihood Ratio (LLR) p-value** in the summaries of the Logit and Probit models is the p-value associated with the LR Statistic of those models.

Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC)

The Receiver Operating Characteristic (ROC) curve is a plot of the True Positive Rate (TPR) - sensitivity of the model, and the False Positive Rate (FPR) - specificity of the model. The TPR represents the proportion of true positive predictions out of all actual positive cases, while the FPR is the proportion of false positive predictions out of all actual negative cases. Every point on the ROC curve represents a different probability threshold for the classification of the observations. The classification thresholds range from low (low sensitivity, high specificity) to high (high sensitivity, low specificity) (Fawcett (2006)). For example, using the default threshold of 0.5, all values of the estimated probability of the observations that are higher or equal to 0.5 fall within the positive category, and all values below 0.5 fall into the negative category.

The Area Under the Curve (Area Under the Curve (AUC)) is the value of the area under the ROC curve. It ranges from 0 to 1, where an AUC of 1 represents a perfect model, 0.5 indicates random guessing, and an AUC below 0.5 indicates that the model performs worse than random guessing (Fawcett (2006)). Values higher than 0.5 and closer to 1 are desirable.

In our case, we work with the 0.5 probability threshold, firstly because the respective AUC has relatively high values in all our models (>0.80), and secondly, due to the number of models estimated. Additionally, we tried different thresholds, and generally only the thresholds from 0.3 to 0.5 seem to perform similarly to the 0.5 threshold probability. The respective ROC curves and AUCs of the models are available in the Appendix.

Confusion Matrix

Another measure connected to the ROC is the Confusion Matrix. It assesses the model's predictive performance by showing the distribution of the predicted and actual values. It divides the predictions into four categories: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) (Fawcett (2006)). In the case of a threshold probability for a confusion matrix, we also use the 0.5 probability, similar to the ROC/AUC. The confusion matrices of the respective models can be viewed in the Appendix as well.

Percentage Correctly Predicted

The Percentage Correctly Predicted is a goodness-of-fit measure that can be computed based on the Confusion Matrix. It is defined as $\frac{TP+TN}{FP+FN}$, where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives. By computing the Percentage Correctly Predicted, we can assess how well the model's latent variable $\tilde{y}_i$ predicts y_i across all observations (Wooldridge (2012)). The four possible outcomes are also listed in the above Confusion Matrix description.

5.5.2 Interpretation

Partial Effect at the Average (PEA)

The coefficients of the estimated logit and probit models provide the signs of the partial effects of each explanatory variable on the response probability and the statistical significance of each explanatory variable (Wooldridge (2012)). However, to compute the magnitudes of the partial effects, since they are not constant (unlike in the LPM), one commonly used method is to replace each explanatory variable with its sample average and use it as an adjustment factor:

$$g(\hat{\beta}_0 + \hat{\beta}_1\overline{x}_1 + \hat{\beta}_2\overline{x}_2 + \ldots + \hat{\beta}_k\overline{x}_k)$$

Where $g(\cdot)$ is the standard normal density in the case of probit, and $g(z) = \frac{\exp(z)}{(1+\exp(z))^2}$ in the logit case. By multiplying an estimated coefficient of the explanatory variable by the above adjustment factor, we obtain the partial effect of that explanatory variable for the average person in the sample, i.e., the Partial Effect at the Average (PEA) (Wooldridge (2012)). $\hat{\beta}$ is an estimated coefficient of the given explanatory variable.

Average Partial Effect (APE)

The Average Partial Effect (APE) is a measure of the marginal effect of the specific variable, similar to the PEA. However, the APE uses a different scale factor to multiply the estimated coefficient of the explanatory variable of interest. The scale factor for APE is given by:

$$\frac{1}{n}\sum_{i=1}^{n} g(\hat{\beta}_0 + x_i\hat{\beta})$$

where $g(\hat{\beta}_0 + x_i\hat{\beta}) = \phi(\hat{\beta}_0 + x_i\hat{\beta})$ in the case of Probit and $g(\hat{\beta}_0 + x_i\hat{\beta}) =$

$\frac{\exp(\hat{\beta}_0+x_i\hat{\beta})}{[1+\exp(\hat{\beta}_0+x_i\hat{\beta})]^2}$ in the case of Logit (Wooldridge (2012)).

In the formulas, $\hat{\beta} = \hat{\beta}_i$, $i = 1, 2, \ldots, n$.

5.6 Dataset 1: Artifacts, Full

5.6.1 Descriptive Statistics

Next, we provide some key characteristics of the first non-reduced dataset, consisting of all 7,622 observations. After analyzing the correlation matrix (Correlation Matrix 1 in Appendix) of the variables, we chose a set of 10 variables that had either the strongest correlation with the artifacts submission rate or were subject to the main hypotheses. The Table 5.1 provides an overview of the variables kept, including some non-numerical variables such as the *municipal_office* variable.

Variable	**Count**	**Unique**	**Mode**	**Frequency**
profile	7622	7622	Detek	1
link	7622	2	0.0	7422
experience	7622	1097	0.0	3064
contributions	7622	113	0.0	6190
comments	7622	737	0.0	847
artifacts	7622	294	1.0	1534
coins	7622	180	0.0	2289
residence_additional_info	7622	2	0.0	7538
municipality†	4174	893	Praha	363
municipal_office††	4160	354	Praha	408
real_net_monetary_index	7622	355	1.0	3462
artifs_rate ‡	7622	302	0.0	7056
coins_rate ‡	7622	90	0.0	7477
rate_artifs_dummy	7622	2	0.0	7056
rate_coins_dummy	7622	2	0.0	7477
detector_expensive_dummy	7622	2	0.0	7499
localities_rate	7622	278	0.021478	3462

Table 5.1: The main variables with the count of observations, the number of unique values, the mode and the frequency of the mode - Dataset 1

First, we can see that this dataset includes 893 municipalities out of the total of 6,254 municipalities existing in the Czech Republic (marked by †, Table 5.1). If it were possible, using the municipalities might increase the variability of our data compared to using municipal offices, which are present in 354 cases in our dataset (††, Table 5.1). Nevertheless, the dataset as it is includes the data for 354 municipal offices out of the total number of 393 municipal offices in the Czech Republic. Hence, our dataset includes observations of the *real_net_monetary_index* representing a considerable 9/10 of the whole Czech

Republic. Furthermore, we can see (marked by ‡, Table 5.1) that 7,056 and 7,477 out of 7,622 profiles did not submit any artifact or coin, respectively, meaning there is a large number of 0 observations of either artifacts or coins submission rate variable. In other words, 566 individuals submitted at least one artifact to the archeological authority, whereas only 145 individuals declared the submission of at least one coin. Hence, the large number of 0 observations might have an excessive influence on the results.

Nevertheless, we can see that the largest number of observations of *municipal_office* comes from Prague (Praha), specifically 408 out of the 7,622 observations. Using all the 7,622 observations with filled values enabled us to utilize the variety of observations as much as possible. Moreover, the distribution of municipal office observations in our dataset plausibly represents the distribution of population in the Czech Republic.

Next, we have the summary statistics of the dataset shown in Table 5.2. From this overview, we can spot the skewness to the right (mean > median) of all the variables, except *real_net_monetary_index*, *coins_rate*, and *localities_rate*. This observation reflects the already mentioned high frequency of low values, which is possibly addressed in later sections by further reducing the dataset. Moreover, the relatively high difference between the maximum value and the 75th percentile raises the second main concern, which is possible outliers. To address this and other issues related to the models, we use mainly scatter plots, histograms, and descriptive statistics in the following sections that describe the key variables separately.

Experience Variable

First, we examine the properties of the first independent variable, the *experience* score. This variable represents the volume of individual activity in identifying the objects uploaded on the website. The initial histogram of the observations of this variable (Figure 5.1 - left) excludes zero observations and displays the counts of values greater than zero only. The number of zero observations can be seen on the same plot displayed at the top right. This adjustment was made for the plots to be scaled in a way that they could be interpreted. Without this amendment, the plots would be adjusted for the overall extensive number of zeros in the dataset, and the counts of other observations might not be visible at all. The same reasoning applies to several other histograms of other variables.

In the left histogram of Figure 5.1, we can observe that, apart from the

Variable	Mean	Std	Min	25%	50%	75%	Max
link	0.03	0.16	0.00	0.00	0.00	0.00	1.00
experience	829.90	8892.82	0.00	0.00	10.00	78.00	498425.00
contributions	2.66	22.37	0.00	0.00	0.00	0.00	1489.00
comments	144.77	798.35	0.00	2.00	9.00	38.00	24463.00
artifacts	21.05	73.18	0.00	1.00	4.00	14.00	3332.00
coins	9.85	27.88	0.00	0.00	2.00	7.00	713.00
residence_additional_info	0.01	0.10	0.00	0.00	0.00	0.00	1.00
real_net_monetary_index	1.00	0.05	0.77	0.97	1.00	1.00	1.22
artifs_rate	0.01	0.08	0.00	0.00	0.00	0.00	1.00
coins_rate	0.00	0.04	0.00	0.00	0.00	0.00	1.00
rate_artifs_dummy	0.07	0.26	0.00	0.00	0.00	0.00	1.00
rate_coins_dummy	0.02	0.14	0.00	0.00	0.00	0.00	1.00
detector_expensive_dummy	0.02	0.13	0.00	0.00	0.00	0.00	1.00
localities_rate	0.02	0.01	0.00	0.02	0.02	0.03	0.14

Table 5.2: Summary Statistics of the Dataset 1

3064 zero values that the *experience* variable contains, there are approximately 4,500 to 4,600 observations having values from 1 to around 25,000. Then, there appear to be several observations attaining the experience score greater than 25,000 but lower than 50,000. The rest of the observations (*experience* $>$ 50,000) seem to be sparsely represented in the dataset.

Due to the described properties, we performed a logarithmic transformation of the *experience* variable using the transformation $log(x+1)$. This transformation is frequently used to make the original variable distribution closer to a normal distribution. Moreover, it is particularly appropriate when the variable x takes only positive values including zero, since when $x \sim 0$, $log(x+1) \sim 0$ (Wooldridge (2012)).

From the histogram of the log-transformed values (Figure 5.1 - right), we can observe that the data has a shape much closer to the desired bell-shaped normal distribution than the original one. However, we need to stress that we excluded the original 3,064 observations of the variable *experience* being 0; including them would slightly alter the appearance of the histogram. Nevertheless, the plots aim to highlight the general properties of the individual variables for conciseness.

Since the values on the x-axis of the histogram of log-transformed data are logarithms of the values on the x-axis of the original histogram plot, we can infer that the most frequent observations of the original experience score are approximately 13-20 points. However, when including the zero observations, there might be two peaks in the histogram, one close to zero and the second one approximately at the already mentioned values of 13-20 points. The log-transformed data are still skewed to the right; however, the improvement in the data distribution is apparent. Overall, as noticeable particularly on the histogram of the original data, the spread of the values on the x-axis might raise concerns about potential outliers. Most of the values are located in the range of 0 to 50,000 experience score, whereas the rest of the observations attain values from 50,000 up to 500,000, with their respective bars not being visible on the plot at all. Hence, we further investigate the data.

Figure 5.1: Experience: Histogram of Original Observations and the Log-transformed Data (both excluding 0 observations)

From the scatter plot of the *experience* observations and the dependent variable *artifs_rate* (Figure 5.2), we suspect the point close to the experience number of 500,000 to be an outlier. Additionally, the group of points with the *artifs_rate* of 1 might be a concern since there is a relatively big 'jump' from the last point attaining the value of about 0.8. Outliers were further examined using the interquartile range as well as the percentile cut-off. However, due to the sample properties, those methods yielded an unnecessarily large number of outliers, even when using, for example, a large cut-off like the 99.7th percentile, which contradicts our approach of using as many observations as possible for each of the samples. Nevertheless, thanks to the logarithmic transformation, as seen in Figure 5.1, there might be no need to delete many observations.

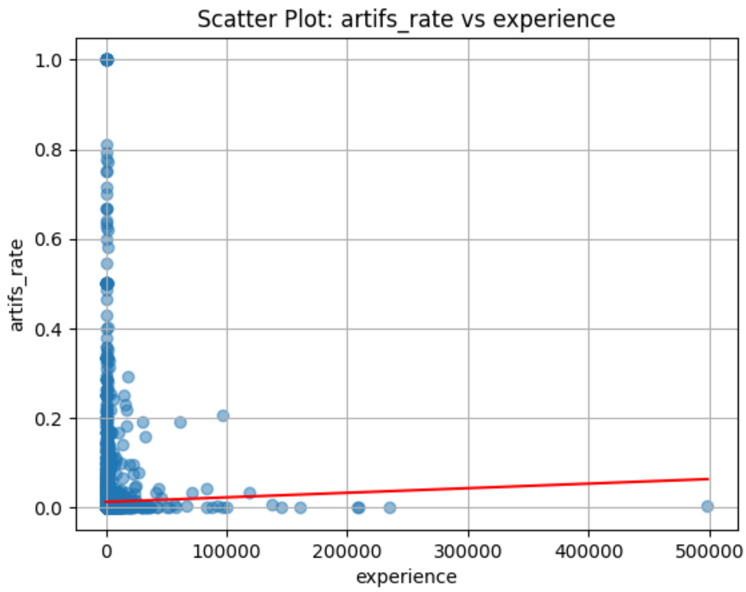

Figure 5.2: Scatter Plot of Experience

Contributions Variable

The *contributions* variable has similar properties as the *experience* variable (see Figure 5.3). The seeming improvement in the distribution compared to the *experience* variable is mainly due to the much larger portion of the data attaining the value of 0; in particular, 6,190 people did not contribute to the website by writing an article or club post. This observation might stem from the fact that the active creation of content (article/post) requires more effort than just identifying objects and subsequently gaining the experience score. Hence, in fact, the log-transformed data distribution of the *contributions* variable is less close to the normal distribution than the previous *experience* variable. Nevertheless, it is still much closer to the normal distribution than the original non-transformed data.

The scatter plot of *contributions* and *artifs_rate* (Figure 5.4) also yields similar results as the scatter plot of the previous variable. There is again one observation 'far away' from others, attaining the value of about 1,500 contributions. This time it is clearly an observation of a different profile than the previous one since the *artifs_rate* of this observation is higher. Additionally, the values attaining the *artifs_rate* of 1.0 seem a bit unusual due to the gap between them and the rest of the observations.

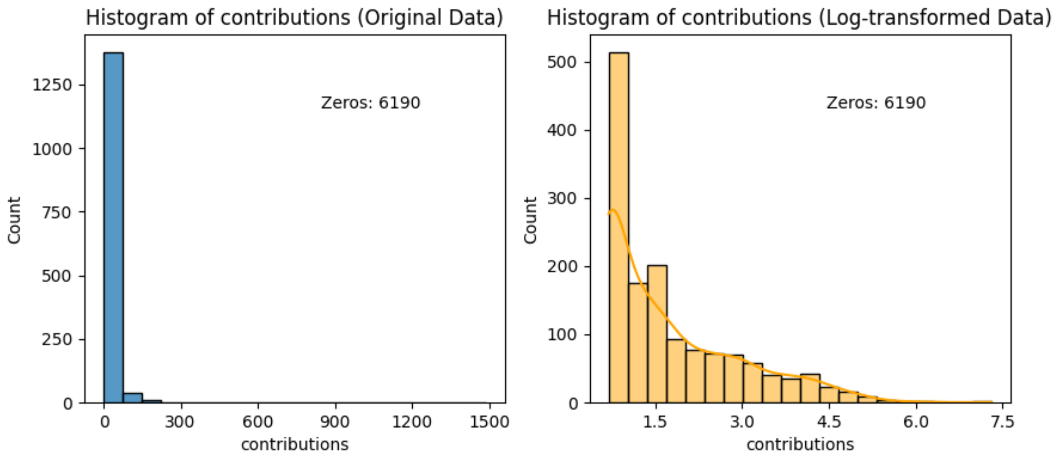

Figure 5.3: Contributions: Histogram of Original Observations and the Log-transformed Data (both excluding 0 observations)

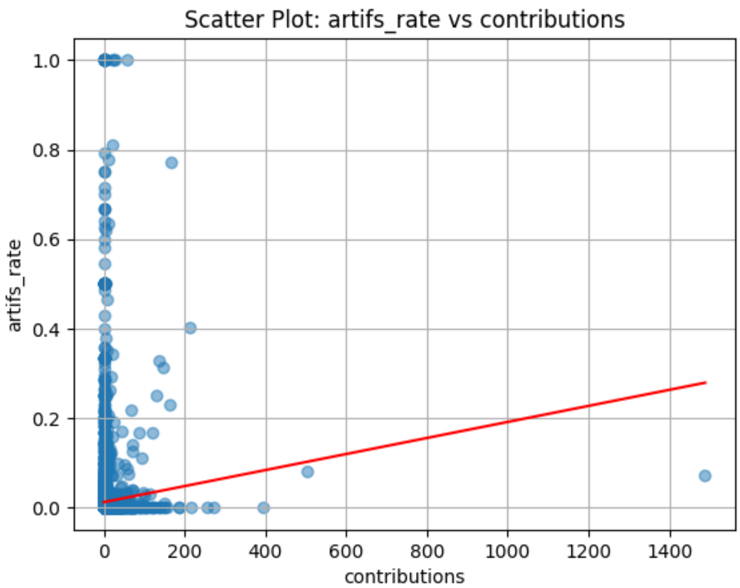

Figure 5.4: Scatter Plot of Contributions

Comments Variable

In the case of *comments* variable, it might be worth noting that compared to the previous two variables, only 847 people out of the 7,622 did not write any comments. Therefore it is more probable that one writes a comment than helping another individual with find identification. Otherwise, the comments variable has a similar distribution as the previous two variables (Figure 5.5). We might notice that basically all original variables stemming directly from the website have a distribution resembling the Poisson distribution with the mean parameter (λ) attaining a value that is very close to zero. Although this violates some of the assumptions of the Poisson distribution generating process, such as the randomness of the incoming members and their levels of activity - one can tell others to sign up to the website and might be more likely to collaborate with others as well; another violation might be that it is more likely that new categories of people appear with the longer period of the website functioning. Despite this, we might consider the observations of our original sample being created via a Poisson process, i.e., in a certain time gap of the website functioning, there appeared a number of people with certain levels of activity that are Poisson distributed.

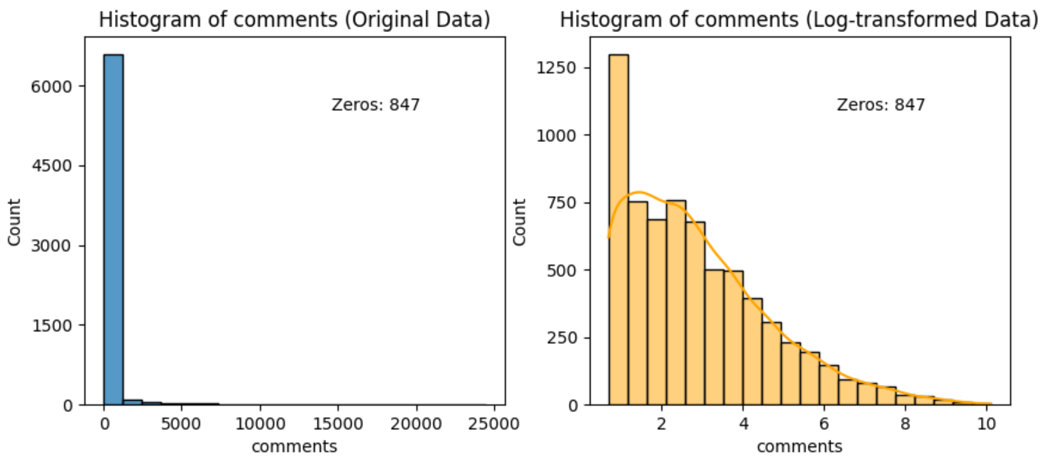

Figure 5.5: Comments: Histogram of Original Observations and the Log-transformed Data (both excluding 0 observations)

Looking at the scatter plot (Figure 5.6), there seem to be three observations of the number of comments that vary from the rest, all of them attaining values greater than 15,000. The one additional observation that is close to the number of 25,000 comments seems particularly extreme.

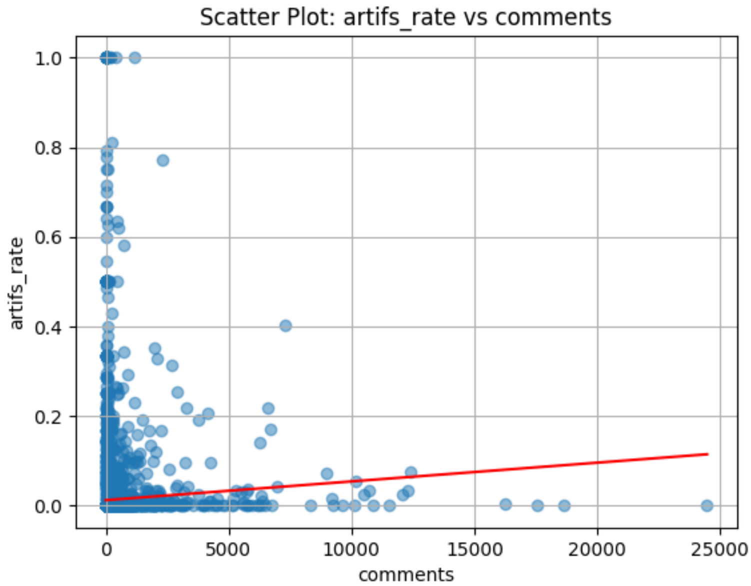

Figure 5.6: Scatter Plot of Comments

Artifacts Variable

Again, the *artifacts* variable properties, distribution, and extreme values seem to be similar to the already described variables. The distribution resembles a Poisson distribution, and the logarithmic transformation appears to have made the data more normal. There also seems to be one extreme value on the x-axis, along with a cluster of values at 1.0 on the y-axis. The respective plots can be seen in Figure 5.7 and 5.8.

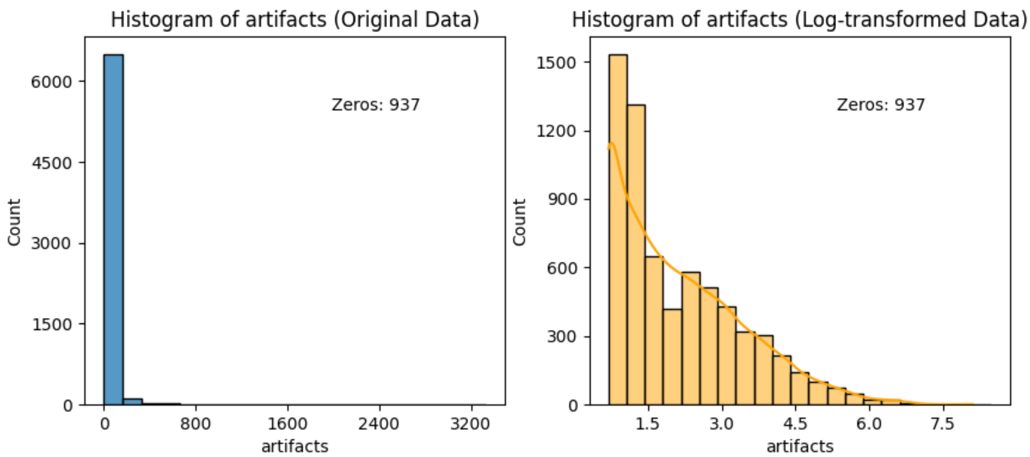

Figure 5.7: Artifacts: Histogram of Original Observations and the Log-transformed Data (both excluding 0 observations)

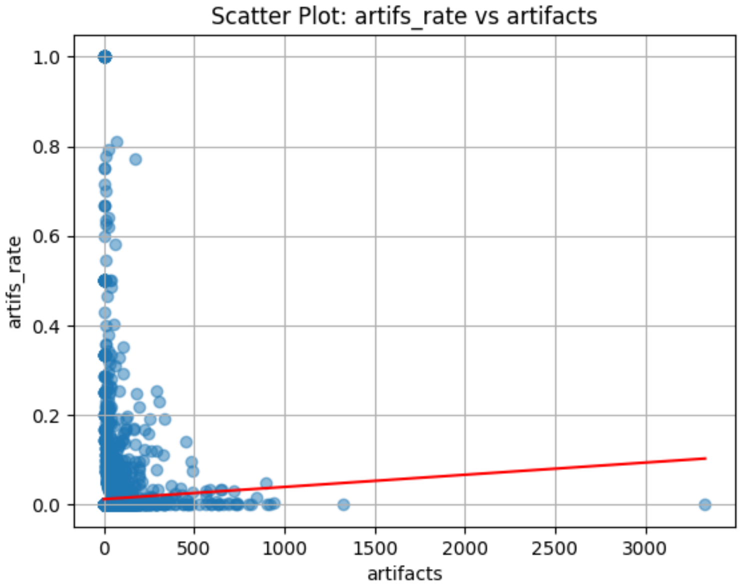

Figure 5.8: Scatter Plot of Artifacts

Real Net Monetary Index Variable

The histogram of the *real_net_monetary_index* in our sample of 7,622 observations appears relatively close to a normal distribution (Figure 5.9). However, there are clearly some flaws present. For example, the data with a value close to 1.0, as well as the data from about 1.022 to 1.111, seem to be underrepresented in the dataset, causing the distribution to be asymmetrical with multiple peaks. The peak at around the value of 1.111 to 1.133 might be present mainly due to the 408 observations representing the real net monetary index of Prague.

In the scatter plot of the *index* vs *artifs_rate* (Figure 5.10), five main groups of observations are apparent. First, the vertical line of observations obtaining a value of 1.0 on the x-axis. This is simply caused by the artificial filling of the data with an average *real_net_monetary_index* of the Czech Republic, which has a value of 1.0. This line represents the variety of the submission rate for those individuals who did not declare their residence on the website. Despite minor differences (such as the values about 0.4 or 0.8 for the submission rate), it seems to represent the density of other observations quite well. Hence, we may assume that the lack of residential information appears randomly and is not specific to any value or group of values of the submission rate. Moreover, the minor differences might be explained by the relatively large number of filled-in values, namely 3,462.

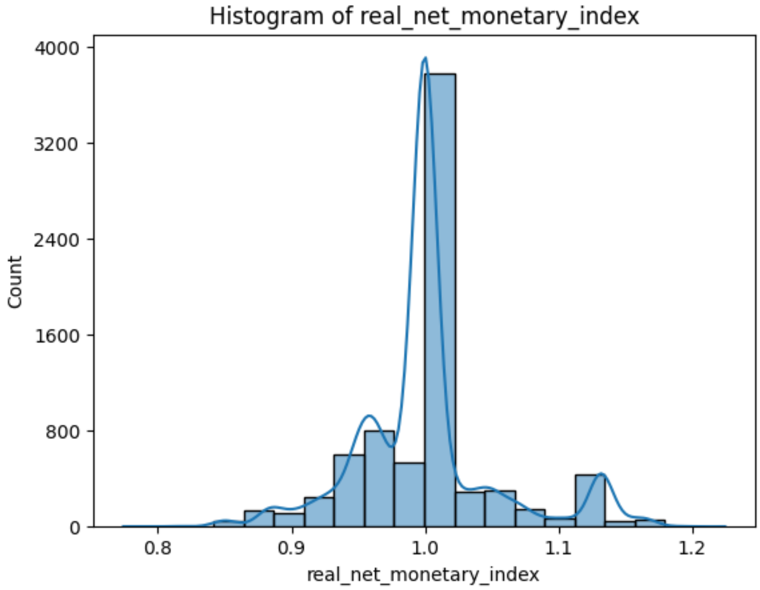

Figure 5.9: Histogram of the Real Net Monetary Index Observations

Second, the silhouette of the vertical line at about the point 1.14 on the x-axis likely represents the variety of observations mainly for Prague. Interestingly, the values of people from the area with this 1.14 observation of the *real_net_monetary_index* do not exceed 0.5.

Third, there appears a fictional horizontal line of points at the level of 0 on the y-axis (*artifs_rate*). This is simply caused by a large number of people not declaring submission of any artifact.

Lastly, there seem to be two other horizontal lines of points. One at the level of the submission rate of 0.5 and the second one at the level of 1.0. The presence of those lines at the levels of 0.5 and 1 might be caused by the fact that there are lots of observations of low numbers of artifacts uploaded (particularly the observations of 1 and 2). At the same time, it is more likely that those observations would attain values of the submission rate of 0.5 (1 artifact submitted out of 2) and 1 (one artifact submitted out of 1, or two submitted out of 2). Moreover, it may be possible that the preferences of individuals are artificially constrained by the definition of the submission rate variable (censoring). In other words, the preferences of individuals for submitting or not submitting finds, in reality, might be more diverse than the variable *artifs_rate* allows. Therefore, the observation values of 1 might not only represent the real preferences for submitting all the finds but also other unobserved preferences, such as additional activity to persuade others to submit their finds. Those individuals might then obtain a value higher than 1.0 if possible.

From all the above and also from the fact that we observed in all of the previous variables, the relatively big gap between the second highest value of the submission rate (at about 0.81) and the highest value (1.0), we consider deleting the observations attaining the value of the submission rate of artifacts of 1.0.

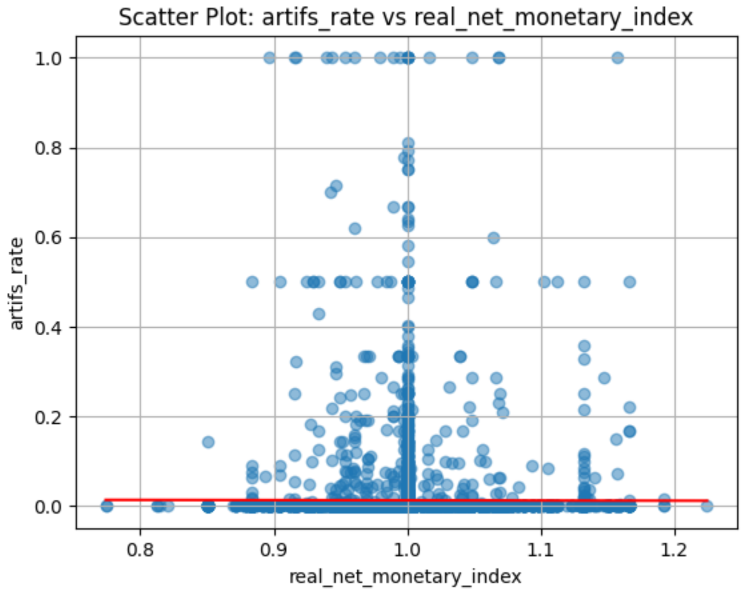

Figure 5.10: Scatter Plot of the Real Net Monetary Index

Localities Rate Variable

As can be seen in Figures 5.11 and 5.12, the variable *localities_rate* follows a similar pattern to the previous variable. The distribution of this variable has a similar shape as well, but since it is not defined the same way, it is shifted more towards zero. Also, there appears to be a greater number of 0 observations, thus the distribution looks less close to the normal distribution than the *real_net_monetary_index* variable distribution. Additionally, there appear to be some potentially extreme values at the point of 0.14 localities rate. Finally, there is no difference with other variables regarding the values of the submission rate of 1.0, which are considered for deletion in the later stage of the analysis.

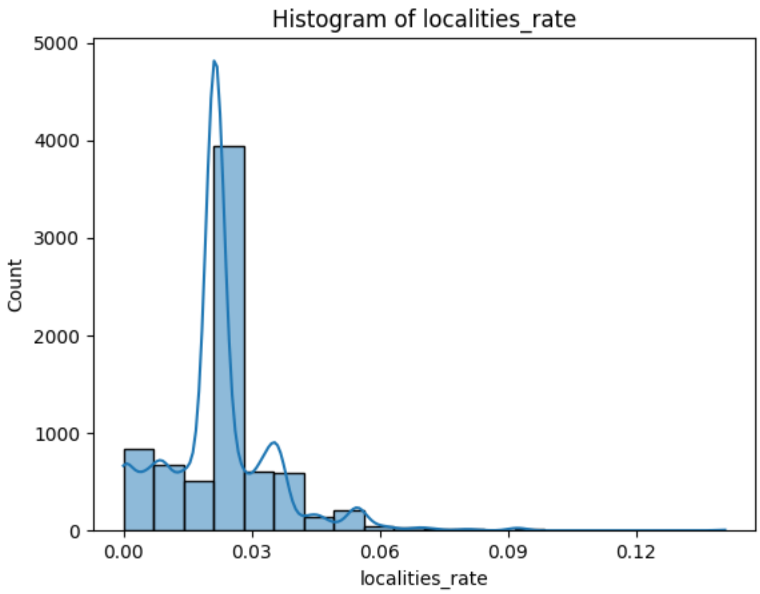

Figure 5.11: Histogram of the Localities Rate Observations

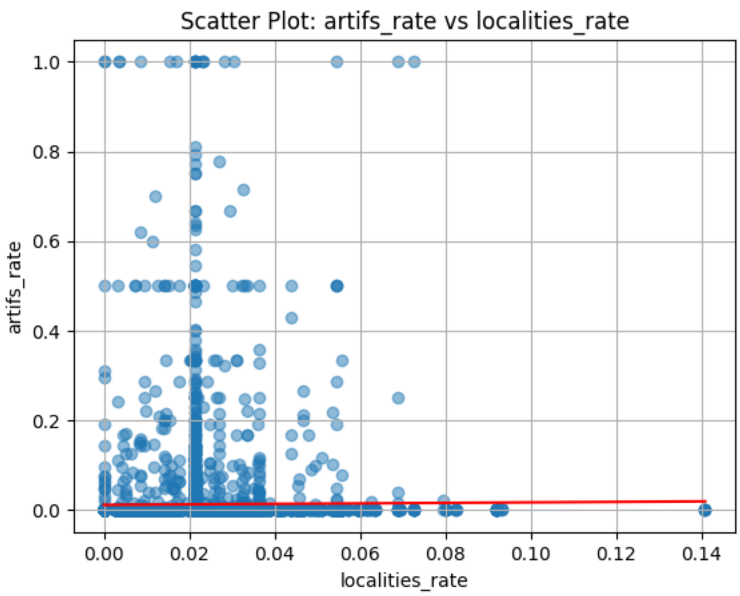

Figure 5.12: Scatter Plot of the Localities Rate

Link Variable (Dummy)

From Figure 5.1, we can see that there are overall 200 observations out of 7,622 (2.62%) that attain the value of 1 for the *link* dummy variable. That means that two hundred people uploaded the link to their other profiles/websites on other platforms, that is, made an effort to present themselves and disclose more personal information.

Residence Additional Info Variable (Dummy)

Overall, 84 people (1.1%) uploaded additional information, that is, age, telephone number, or additional residence (*residence_additional_info*), as can be seen in Figure 5.1 (since it is a dummy variable we simply subtract the count of the mode (0), which is 7,538, from the overall number of observations, which is 7,622).

Detector Expensive Dummy Variable

The count of owners of an 'expensive' metal detector (*detector_expensive_dummy*) in our sample of 7,622 observations is 123 (1.61%) (Figure 5.1). All those people own a metal detector that costs more than 30,000 CZK.

Coins Submission Rate Variable

The Figures 5.13 and 5.14 show the distribution, log-transformed distribution and scatter plot of the *coins_rate* variable, respectively. From the scatter plot we can see that there is likely greater variety of combinations of the *coins_rate* and *artifs_rate* variables compared to scatter plots of other variables, since the spread of the values is relatively big. The histogram of this variable is almost identical with the histogram of the following *artifs_rate* variable. Therefore, since the *artifs_rate* is our key dependent variable, its description is left to the following section.

The Dependent Variable - Submission Rate of Artifacts

The distribution of the dependent variable - the *artifs_rate* seems, similarly to other original variables, to be likely approximately Poisson distributed with the mean parameter (λ) having a value close to zero (however, in fact, the Poisson distribution assumes a count/integer variable; our submission rate is not).

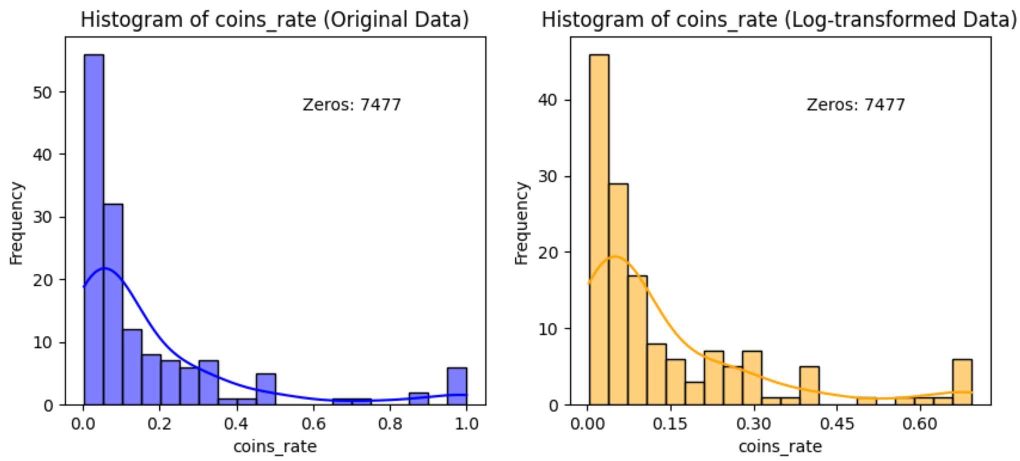

Figure 5.13: Histogram of The Coins Submission Rate Observations and Log-transformed Data

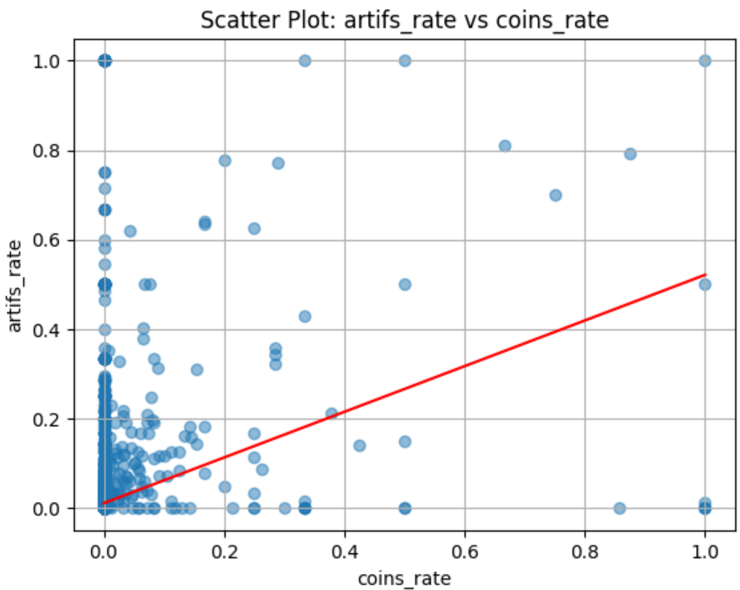

Figure 5.14: Scatter Plot of the Artifacts Rate vs Coins Rate Variables

As already discussed in the part describing the properties of the variable *real_net_monetary_index*, we can see the presence of a greater number of observations at the positions of 0.5 and 1.0 submission rates (x-axis, Figure 5.15). This presence might be explained by the more frequently occurring low counts of uploaded artifacts and the subsequent possible higher probability of the submission rate values of 0.5 and 1 occurring in the sample. Furthermore, although it does not help much with skewness this time, we create a log-transformation of the dependent variable. As can be seen in Figure 5.15, this might help decrease the impact of possible heteroskedasticity on our models by lowering the spread of the values.

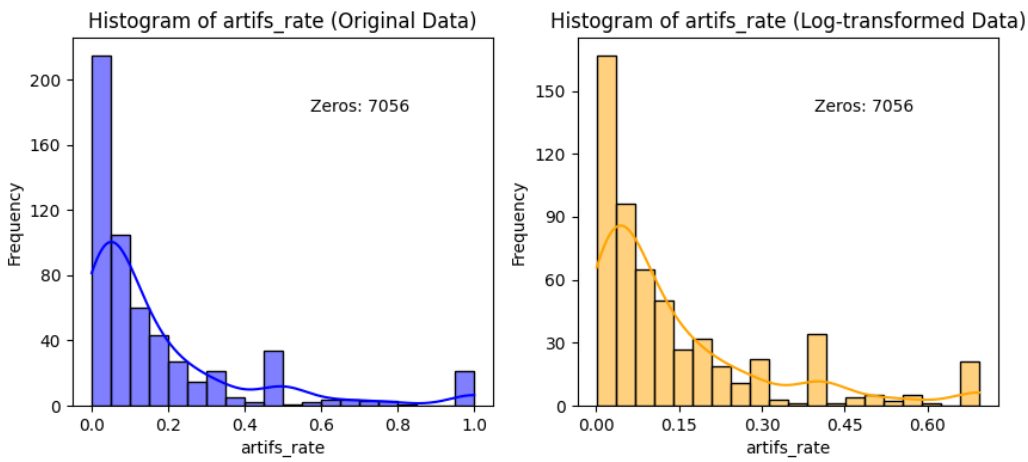

Figure 5.15: Histogram of The Artifacts Submission Rate Observations and Log-transformed Data

The Dependent Variable - Submission Rate of Artifacts (Dummy)

The other key dependent variable, used for Logit, Probit and LPM is the *rate_artifs_dummy* variable, taking the value of 1 whenever the artifacts submission rate is higher than zero, and taking the value of 0 otherwise. In this sample, there are 566 observations out of the 7,622 (7.43%) of this variable that take the value of 1 (see, for example, Figure 5.1).

5.6.2 Models

For this sample, we estimate the following models: OLS (Model 1), OLS with log-transformed variables (Model 2), OLS with heteroskedasticity robust standard errors (Model 2_robust), WLS (Model 2_WLS), LPM (Model 2_LPM), Logit (Model 2_Logit), and Probit (Model 2_Probit).

Additionally, we decided to delete individual outlying observations that were apparent from the scatter plots, and after examining their distance from the 2nd largest observations. We deleted the following observations:

- *experience* = 498,425 (approximately 260,000 higher than the 2nd largest observation)
- *contributions* = 1,489 (approximately 980 higher than the 2nd largest observation)
- *comments* = 24,463 (approximately 5,860 higher than the 2nd largest observation)
- *artifacts* = 3,332 (approximately 2,000 higher than the 2nd largest observation)

We observed that proportionally, those observations do not fit the rest of the data significantly. Additionally, due to the high number of zeros in the sample, it is not possible to delete outliers based on the Interquartile Range (IQR) or percentile cut-off. Those methods were examined and found unsuitable as they would cut off large amounts of values, even for high cut-off values like the 99.7 percentile.

Furthermore, we decided to delete all the following observations (except for the LPM, Logit, and Probit models):

- *artifs_rate* = 1

This means that whenever the observation attains a value of 1 for the artifacts submission rate, it is deleted. We made this choice based on the scatter plots (hinted in the description of several individual variables above) as well as further exploration of the dataset. We realized that people with a value of 1 are, in almost all cases, individuals with a low number of artifacts uploaded. Hence, for these people, it might be more likely that the submission rate taken from the website represents the actual real submission rate less accurately than for people with more artifacts uploaded. Here we further assume that it is more likely for people to stop uploading to the website than to stop metal detecting after finding, for example, one artifact. To sum up, these values might be disturbing, as apparent not only from the scatter and histogram plots. The final number of observations for the non-binary dependent variable models is 7,598.

For the binary dependent variable models, we do not delete the observations of *artifs_rate* = 1 since the definition of the dependent variable should account for all the observations that have at least one artifact submitted, regardless of the total number of artifacts uploaded. From this, we can see the possible advantage of the binary dependent variable models, which measure the willingness to submit an artifact overall, rather than the willingness to submit an artifact out of the total number of artifacts. We lose here the dimension of the proportion, which, in fact, might be beneficial, as it gets rid of the potential incomplete reporting of the true submission rate. The final number of observations for the binary dependent variable models is 7,619.

Furthermore, we decided to use a logarithmic transformation of the variables described above in all the models (except Model 1), as models containing these transformed variables perform better compared to the models with original variables. Hence, we built the following models:

OLS, WLS Models

1. Model 1 (OLS):

   Dependent Variable: *artifs_rate*
   Independent Variables: *experience*, *contributions*, *comments*, *artifacts*, *real_net_monetary_index*, *log_coins_rate*, *localities_rate*, *link*, *residence_additional_info*, *detector_expensive_dummy*

   Model Equation:

$$\begin{aligned} artifs\_rate = \beta_0 &+ \beta_1 experience + \beta_2 contributions \\ &+ \beta_3 comments + \beta_4 artifacts + \beta_5 real\_net\_monetary\_index \\ &+ \beta_6 log\_coins\_rate + \beta_7 localities\_rate \\ &+ \beta_8 link + \beta_9 residence\_additional\_info \\ &+ \beta_{10} detector\_expensive\_dummy \end{aligned}$$

2. Model 2 (OLS - Log Transformed):

   Dependent Variable: *log_artifs_rate*
   Independent Variables: *log_experience*, *log_contributions*, *log_comments*, *log_artifacts*, *real_net_monetary_index*, *log_coins_rate*, *localities_rate*, *link*, *residence_additional_info*, *detector_expensive_dummy*

Model Equation:

$$\begin{aligned}
log_artifs_rate = \beta_0 &+ \beta_1 log_experience + \beta_2 log_contributions \\
&+ \beta_3 log_comments + \beta_4 log_artifacts \\
&+ \beta_5 real_net_monetary_index + \beta_6 log_coins_rate \\
&+ \beta_7 localities_rate + \beta_8 link \\
&+ \beta_9 residence_additional_info \\
&+ \beta_{10} detector_expensive_dummy
\end{aligned}$$

3. Model 2_Robust (Robust OLS - Log Transformed):

   Model Equation:

$$\begin{aligned}
log_artifs_rate = \beta_0 &+ \beta_1 log_experience + \beta_2 log_contributions \\
&+ \beta_3 log_comments + \beta_4 log_artifacts \\
&+ \beta_5 real_net_monetary_index + \beta_6 log_coins_rate \\
&+ \beta_7 localities_rate + \beta_8 link \\
&+ \beta_9 residence_additional_info \\
&+ \beta_{10} detector_expensive_dummy
\end{aligned}$$

4. Model 2_WLS (Weighted Least Squares - Log Transformed):

   Model Equation:

$$\begin{aligned}
log_artifs_rate = \beta_0 &+ \beta_1 log_experience + \beta_2 log_contributions \\
&+ \beta_3 log_comments + \beta_4 log_artifacts \\
&+ \beta_5 real_net_monetary_index + \beta_6 log_coins_rate \\
&+ \beta_7 localities_rate + \beta_8 link \\
&+ \beta_9 residence_additional_info \\
&+ \beta_{10} detector_expensive_dummy
\end{aligned}$$

LPM, Logit, Probit Models

Dependent Variable: *rate_artifs_dummy*
Independent Variables: *log_experience*, *log_contributions*, *log_comments*, *log_artifacts*, *real_net_monetary_index*, *log_coins_rate*, *localities_rate*, *link*, *residence_additional_info*, *detector_expensive_dummy*

1. Model 2_LPM (Linear Probability Model - Dummy Variable):

$$\begin{aligned} rate\_artifs\_dummy = {} & \beta_0 + \beta_1 log\_experience + \beta_2 log\_contributions \\ & + \beta_3 log\_comments + \beta_4 log\_artifacts \\ & + \beta_5 real\_net\_monetary\_index \\ & + \beta_6 log\_coins\_rate \\ & + \beta_7 localities\_rate + \beta_8 link \\ & + \beta_9 residence\_additional\_info \\ & + \beta_{10} detector\_expensive\_dummy \end{aligned}$$

2. Model 2_Logit (Logit Model - Dummy Variable):

$$\begin{aligned} P(rate\_artifs\_dummy = 1|x) = {} & \Lambda(\beta_0 + \beta_1 log\_experience \\ & + \beta_2 log\_contributions \\ & + \beta_3 log\_comments + \beta_4 log\_artifacts \\ & + \beta_5 real\_net\_monetary\_index \\ & + \beta_6 log\_coins\_rate \\ & + \beta_7 localities\_rate + \beta_8 link \\ & + \beta_9 residence\_additional\_info \\ & + \beta_{10} detector\_expensive\_dummy) \end{aligned}$$

3. Model 2_Probit (Probit Model - Dummy Variable):

$$
\begin{aligned}
P(rate\_artifs\_dummy = 1|x) = \Phi(&\beta_0 + \beta_1 log\_experience \\
&+ \beta_2 log\_contributions \\
&+ \beta_3 log\_comments + \beta_4 log\_artifacts \\
&+ \beta_5 real\_net\_monetary\_index \\
&+ \beta_6 log\_coins\_rate \\
&+ \beta_7 localities\_rate + \beta_8 link \\
&+ \beta_9 residence\_additional\_info \\
&+ \beta_{10} detector\_expensive\_dummy)
\end{aligned}
$$

As can be seen, all the models contain the same variables, differing in individual cases by either having log-transformed variables (Model 2, Model 2_Robust, Model 2_WLS), or by having a binary dependent variable (Model 2_LPM, Model 2_Logit, Model 2_Probit).

Finally, the current section was made more general to account for the rest of the samples as much as possible; therefore, for the above initial sample (7,622 observations), we provided a more detailed description. Concerning other datasets, we provide a rather concise overview, often referring to the figures in the Appendix.

5.7 Dataset 2: Coins, Full

5.7.1 Descriptive Statistics

This dataset, consisting of 5,323 observations, was created by deleting the observations that did not include any submitted coins from the original dataset of 7,728 observations. In other words, this dataset consists of observations of profiles that have at least one coin uploaded on the website. Hence, this dataset is used for the analysis of coins and coins submission rate only.

From Table 5.3, we can see that compared to the previous dataset used for the analysis of artifacts submission rate, the number of distinct municipal offices (*municipal_office*) in the dataset decreased slightly - from 354 to 329. However, more concerning might be the decrease in the observations that were successfully matched with the respective municipal office. This number decreased from 4,160 (out of 7,622) for the artifacts-specific dataset to 2,765 (out of 5,323) for this dataset, respectively. This means that almost half of the dataset, specifically 2,558 observations, were filled with the values of an average of the Czech Republic. Therefore, although there is not a big decrease in the variety of observations, there is a decrease in the actual number of observations that include those varying observations and a simultaneous increase in the frequency of one specific observation, which is the average of the Czech Republic.

Despite this, the correlation matrix of the independent variables and the dependent variable (*coins_rate*) follows a very similar pattern to the correlation matrix specific to the submission rate of artifacts from the previous section. However, there is one notable difference, which is the sign and magnitude of correlation with the *real_net_monetary_index* variable. The *coins_rate* now has a small positive correlation of 0.01 with the real net monetary index, in comparison to the negligible -0.0029 correlation of *artifs_rate* with the real net monetary index in the previous dataset. For further details, see Correlation Matrices 1 and 2 in the Appendix.

Furthermore, as can be seen in Table 5.4, the variables in this dataset bear very similar characteristics as in the first dataset. Due to this, we do not further discuss additional characteristics of the individual variables. The respective plots and further data properties are also very similar to the first dataset, specific to artifacts. The respective figures can be seen in the Appendix.

Variable	**Count**	**Unique**	**Mode**	**Frequency**
profile	5322	5322	Detek	1
link	5322	2	0.0	5164
experience	5322	1067	0.0	1658
contributions	5322	107	0.0	4088
comments	5322	719	0.0	343
artifacts	5322	293	0.0	936
coins	5322	179	1.0	1356
residence_additional_info	5322	2	0.0	5253
municipality	2778	696	Praha	243
municipal_office	2765	328	Praha	269
real_net_monetary_index	5322	329	1.0	2557
submitted_number_artifs	4378	44	0.0	3864
number_artifs	4378	291	1.0	457
artifs_rate	5322	299	0.0	4808
submitted_number_coins	5322	14	0.0	5177
number_coins	5322	178	1.0	1362
coins_rate	5322	90	0.0	5177
rate_artifs_dummy	5322	2	0.0	4808
rate_coins_dummy	5322	2	0.0	5177
detector_expensive_dummy	5322	2	0.0	5224
localities_rate	5322	258	0.021478	2557

Table 5.3: The main variables with the count of observations, the number of unique values, the mode, and the frequency of the mode - Dataset 2

Variable	Mean	Std	Min	25%	50%	75%	Max
link	0.03	0.17	0.00	0.00	0.00	0.00	1.00
experience	1160.45	10609.45	0.00	0.00	21.00	157.00	498425.00
contributions	3.42	26.13	0.00	0.00	0.00	0.00	1489.00
comments	199.56	946.87	0.00	4.00	16.00	65.00	24463.00
artifacts	28.54	86.26	0.00	1.00	7.00	23.00	3332.00
coins	14.10	32.46	1.00	1.00	4.00	12.00	713.00
residence_additional_info	0.01	0.11	0.00	0.00	0.00	0.00	1.00
real_net_monetary_index	1.00	0.05	0.77	0.98	1.00	1.00	1.22
submitted_number_artifs	0.71	4.37	0.00	0.00	0.00	0.00	136.00
number_artifs	34.69	94.03	1.00	4.00	10.00	29.00	3338.00
artifs_rate	0.01	0.07	0.00	0.00	0.00	0.00	1.00
submitted_number_coins	0.08	0.97	0.00	0.00	0.00	0.00	54.00
number_coins	14.09	32.45	1.00	1.00	4.00	12.00	713.00
coins_rate	0.00	0.05	0.00	0.00	0.00	0.00	1.00
rate_artifs_dummy	0.10	0.30	0.00	0.00	0.00	0.00	1.00
rate_coins_dummy	0.03	0.16	0.00	0.00	0.00	0.00	1.00
detector_expensive_dummy	0.02	0.13	0.00	0.00	0.00	0.00	1.00
localities_rate	0.02	0.01	0.00	0.02	0.02	0.02	0.14

Table 5.4: The Main Summary Statistics of the Dataset 2

5.7.2 Models

For this sample of observations specific to coins, we conduct an analysis using the same models as in the first sample. The only difference is that instead of the *artifs_rate*, *log_artifs_rate*, and *rate_artifs_dummy* dependent variables, we use the *coins_rate*, *log_coins_rate*, and *rate_coins_dummy*, respectively. Similarly, in the set of independent variables, we change the *artifacts* variable to the *coins* variable, the *log_artifacts* to the *log_coins* variable, and finally, we exchange the independent variable *coins_rate* for *artifs_rate*. We delete outliers using the exact same approach as in the first sample, leaving us with 5,313 and 5,319 observations for the non-binary dependent variable and binary dependent variable models, respectively. The models estimated for this dataset are as follows:

OLS, WLS Models (Non-Binary Dependent Variable Models)

1. Model 1 (OLS):

   Dependent Variable: *coins_rate*

   Independent Variables: *experience*, *contributions*, *comments*, *coins*, *real_net_monetary_index*, *log_coins_rate*, *localities_rate*, *link*, *residence_additional_info*, *detector_expensive_dummy*

Model Equation:

$$\begin{aligned} coins\_rate = \beta_0 &+ \beta_1 experience \\ &+ \beta_2 contributions + \beta_3 comments \\ &+ \beta_4 coins + \beta_5 real\_net\_monetary\_index \\ &+ \beta_6 log\_coins\_rate + \beta_7 localities\_rate \\ &+ \beta_8 link + \beta_9 residence\_additional\_info \\ &+ \beta_{10} detector\_expensive\_dummy \end{aligned}$$

2. Model 2 (OLS - Log Transformed):

   Dependent Variable: *log_coins_rate*
   Independent Variables: *log_experience*, *log_contributions*, *log_comments*, *log_coins*, *real_net_monetary_index*, *log_artifs_rate*, *localities_rate*, *link*, *residence_additional_info*, *detector_expensive_dummy*

   Model Equation:

$$\begin{aligned} log\_coins\_rate = \beta_0 &+ \beta_1 log\_experience \\ &+ \beta_2 log\_contributions + \beta_3 log\_comments \\ &+ \beta_4 log\_coins + \beta_5 real\_net\_monetary\_index \\ &+ \beta_6 log\_artifs\_rate + \beta_7 localities\_rate \\ &+ \beta_8 link + \beta_9 residence\_additional\_info \\ &+ \beta_{10} detector\_expensive\_dummy \end{aligned}$$

3. Model 2_Robust (Robust OLS - Log Transformed):

Model Equation:

$$\begin{aligned}
log_coins_rate = \beta_0 + \beta_1 log_experience \\
+ \beta_2 log_contributions + \beta_3 log_comments \\
+ \beta_4 log_coins + \beta_5 real_net_monetary_index \\
+ \beta_6 log_artifs_rate + \beta_7 localities_rate \\
+ \beta_8 link + \beta_9 residence_additional_info \\
+ \beta_{10} detector_expensive_dummy
\end{aligned}$$

4. Model 2_WLS (Weighted Least Squares - Log Transformed):

   Model Equation:

$$\begin{aligned}
log_coins_rate = \beta_0 + \beta_1 log_experience \\
+ \beta_2 log_contributions + \beta_3 log_comments \\
+ \beta_4 log_coins + \beta_5 real_net_monetary_index \\
+ \beta_6 log_artifs_rate + \beta_7 localities_rate \\
+ \beta_8 link + \beta_9 residence_additional_info \\
+ \beta_{10} detector_expensive_dummy
\end{aligned}$$

LPM, Logit, Probit Models (Binary Dependent Variable Models)

Dependent Variable: *rate_coins_dummy*
Independent Variables: *log_experience*, *log_contributions*, *log_comments*, *log_coins*, *real_net_monetary_index*, *log_artifs_rate*, *localities_rate*, *link*, *residence_additional_info*, *detector_expensive_dummy*

1. Model 2_LPM (Linear Probability Model - Dummy Variable):

Model Equation:

$$\begin{aligned} rate\_coins\_dummy = \beta_0 &+ \beta_1 log\_experience \\ &+ \beta_2 log\_contributions + \beta_3 log\_comments \\ &+ \beta_4 log\_coins + \beta_5 real\_net\_monetary\_index \\ &+ \beta_6 log\_artifs\_rate + \beta_7 localities\_rate \\ &+ \beta_8 link + \beta_9 residence\_additional\_info \\ &+ \beta_{10} detector\_expensive\_dummy \end{aligned}$$

2. Model 2_Logit (Logit Model - Dummy Variable):

   Model Equation:

$$\begin{aligned} P(rate\_coins\_dummy = 1|x) = \Lambda(\beta_0 &+ \beta_1 log\_experience \\ &+ \beta_2 log\_contributions \\ &+ \beta_3 log\_comments + \beta_4 log\_coins \\ &+ \beta_5 real\_net\_monetary\_index \\ &+ \beta_6 log\_artifs\_rate + \beta_7 localities\_rate \\ &+ \beta_8 link + \beta_9 residence\_additional\_info \\ &+ \beta_{10} detector\_expensive\_dummy) \end{aligned}$$

3. Model 2_Probit (Probit Model - Dummy Variable):

   Model Equation:

$$\begin{aligned} P(rate\_coins\_dummy = 1|x) = \Phi(\beta_0 &+ \beta_1 log\_experience \\ &+ \beta_2 log\_contributions \\ &+ \beta_3 log\_comments + \beta_4 log\_coins \\ &+ \beta_5 real\_net\_monetary\_index \\ &+ \beta_6 log\_artifs\_rate + \beta_7 localities\_rate \\ &+ \beta_8 link + \beta_9 residence\_additional\_info \\ &+ \beta_{10} detector\_expensive\_dummy) \end{aligned}$$

5.8 Dataset 3: Artifacts, Reduced

5.8.1 Descriptive Statistics

This dataset consists of 4,160 individual observations which include uploaded artifacts or coins and at the same time, the respective municipal office was successfully matched. Hence, this data sample was created by deleting the observations that did not include a recognizable municipality and at the same time included at least one uploaded artifact or coin. By defining the condition for either artifact or coin, we seek to keep the number of observations as high as possible, and at the same time, ensure that all the people included in the sample were active on the website. The same holds for the initial sample for artifacts (7,622 observations) that included the same condition, except for the successfully matched municipal office.

From the full initial dataset of 7,728 observations, we are down by 3,568 (46.17%) to the current number of 4,160 observations. As can be seen in Table 5.5 below, we have the same variability of municipal offices as in the first sample, namely 354. Important in the following figure is the dummy dependent variable; *rate_artifs_dummy* takes the value of 1 in 256 cases. This means that our sample includes 256 observations that have a submission rate of artifacts higher than zero.

Similarly, as for the initial two samples, we delete the top outlying observation for each of the *experience*, *contributions*, *comments*, and *artifacts* variables since they vary significantly from the rest of the sample. Although this could be sensed from Table 5.6 below as well (max. values are very high with respect to the third quartile), we delete them after individually examining the distance between the respective observation and the rest of the sample. This can be seen in the code created for the analysis (link in the Appendix). Moreover, since the histograms and scatter plots follow a similar pattern as for the initial two samples (see Appendix for further detail), we delete the observations of the submission rate of artifacts, *artifs_rate* variable attaining the value of one. Furthermore, we can see in Table 5.6 that the summary statistics have not changed significantly compared to the previous two datasets. Cutting the outliers leaves us with 4,142 observations for the models with the *artifs_rate* dependent variable and 4,157 observations for the models with the *rate_artifs_dummy* dependent variable.

Variable	Count	Unique	Mode	Frequency
profile	4160	4160	Detek	1
link	4160	2	0.0	4042
experience	4160	601	0.0	1840
contributions	4160	73	0.0	3519
comments	4160	383	0.0	546
artifacts	4160	181	1.0	949
coins	4160	123	0.0	1388
residence_additional_info	4160	2	0.0	4125
municipality	4160	888	Praha	363
municipal_office	4160	354	Praha	408
real_net_monetary_index	4160	354	1.131666	408
artifs_rate	4160	144	0.0	3904
coins_rate	4160	39	0.0	4102
rate_artifs_dummy	4160	2	0.0	3904
rate_coins_dummy	4160	2	0.0	4102
detector_expensive_dummy	4160	2	0.0	4088
localities_rate	4160	277	0.0	452

Table 5.5: The main variables with the count of observations, the number of unique values, the mode, and the frequency of the mode - Dataset 3

Variable	Mean	Std	Min	25%	50%	75%	Max
link	0.03	0.17	0.00	0.00	0.00	0.00	1.00
experience	467.05	5413.11	0.00	0.00	4.00	50.00	207830.00
contributions	1.95	25.56	0.00	0.00	0.00	0.00	1489.00
comments	85.69	594.30	0.00	2.00	6.00	25.00	24463.00
artifacts	15.70	68.53	0.00	1.00	3.00	11.00	3332.00
coins	7.18	21.03	0.00	0.00	1.00	6.00	472.00
residence_additional_info	0.01	0.09	0.00	0.00	0.00	0.00	1.00
real_net_monetary_index	1.00	0.07	0.77	0.95	0.98	1.05	1.22
artifs_rate	0.01	0.08	0.00	0.00	0.00	0.00	1.00
coins_rate	0.00	0.04	0.00	0.00	0.00	0.00	1.00
rate_artifs_dummy	0.06	0.24	0.00	0.00	0.00	0.00	1.00
rate_coins_dummy	0.01	0.12	0.00	0.00	0.00	0.00	1.00
detector_expensive_dummy	0.02	0.13	0.00	0.00	0.00	0.00	1.00
localities_rate	0.02	0.02	0.00	0.01	0.02	0.04	0.14

Table 5.6: Summary Statistics of the Dataset 3

5.8.2 Models

For the closeness of both samples in terms of summary statistics and also for comparison purposes, we estimate the exact same models as for Dataset 1. Similarly to the description of the Dataset 1 in the respective section, also the models were created to fit different samples as much as possible, serving for better comparison of the results.

5.9 Dataset 4: Coins, Reduced

5.9.1 Descriptive Statistics

The fourth dataset consists of 2,774 observations. It was created by deleting the observations which did not include matched municipalities and at the same time did not have at least one coin uploaded; deleting those out of the original dataset of 7,728 observations. Therefore, this dataset consists of observations representing the people who uploaded at least one coin to the website and also provided their municipality in a recognizable way.

Variable	**Count**	**Unique**	**Mode**	**Frequency**
profile	2774	2774	Detek	1
link	2774	2	0.0	2680
experience	2774	580	0.0	970
contributions	2774	67	0.0	2237
comments	2774	368	0.0	218
artifacts	2774	181	0.0	505
coins	2774	122	1.0	778
residence_additional_info	2774	2	0.0	2747
municipality	2774	692	Praha	244
municipal_office	2774	328	Praha	270
real_net_monetary_index	2774	328	1.131666	270
artifs_rate	2774	143	0.0	2551
coins_rate	2774	39	0.0	2716
rate_artifs_dummy	2774	2	0.0	2551
rate_coins_dummy	2774	2	0.0	2716
detector_expensive_dummy	2774	2	0.0	2718
localities_rate	2774	257	0.0	293

Table 5.7: The main variables with the count of observations, the number of unique values, the mode, and the frequency of the mode - Dataset 4

The data follow similar patterns to the datasets described so far. Therefore, we conducted the deletion of outliers in the same way as for all the previous datasets. We deleted the top one observation for the variables *experience*, *contributions*, and *comments* (see 'max.' column, Table 5.8). Additionally, we deleted the observations with a *coins_rate* value of one for the non-binary dependent variable models. This process left us with datasets consisting of 2,767 and 2,771 observations for the non-binary dependent variable and binary dependent variable models, respectively.

However, there is one significant change in the analysis of this dataset. Due to the low number of observations (27 out of 2,774, 0.90%) taking the value of one for the *residence_additional_info* dummy variable (see Table 5.7), we decided not to include this variable in our models for this section. Moreover, logit and probit models were unable to estimate the equation (including the *residence_additional_info* variable) as was done for all the previous datasets, further confirming the unsuitability of this variable for the models.

Variable	**Mean**	**Std**	**Min**	**25%**	**50%**	**75%**	**Max**
link	0.03	0.18	0.00	0.00	0.00	0.00	1.00
experience	687.26	6616.97	0.00	0.00	13.00	98.00	207830.00
contributions	2.57	30.98	0.00	0.00	0.00	0.00	1489.00
comments	122.07	723.19	0.00	3.00	12.00	42.00	24463.00
artifacts	21.86	83.07	0.00	1.00	5.00	18.00	3332.00
coins	10.77	24.99	1.00	1.00	3.00	10.00	472.00
residence_additional_info	0.01	0.10	0.00	0.00	0.00	0.00	1.00
real_net_monetary_index	1.00	0.07	0.77	0.95	0.98	1.05	1.22
artifs_rate	0.01	0.07	0.00	0.00	0.00	0.00	1.00
coins_rate	0.00	0.05	0.00	0.00	0.00	0.00	1.00
rate_artifs_dummy	0.08	0.27	0.00	0.00	0.00	0.00	1.00
rate_coins_dummy	0.02	0.14	0.00	0.00	0.00	0.00	1.00
detector_expensive_dummy	0.02	0.14	0.00	0.00	0.00	0.00	1.00
localities_rate	0.02	0.02	0.00	0.01	0.02	0.04	0.14

Table 5.8: Summary Statistics of the Dataset 4

5.9.2 Models

Following the deletion of the independent variable *residence_additional_info*, we estimate the following models with nine independent variables for this dataset:

OLS, WLS Models (Non-Binary Dependent Variable Models)

1. Model 1 (OLS):

   Dependent Variable: *coins_rate*
   Independent Variables: *experience*, *contributions*, *comments*, *coins*, *real_net_monetary_index*, *log_artifs_rate*, *localities_rate*, *link*, *detector_expensive_dummy*

   Model Equation:

$$
\begin{aligned}
coins\_rate = \beta_0 &+ \beta_1 experience \\
&+ \beta_2 contributions + \beta_3 comments \\
&+ \beta_4 coins + \beta_5 real\_net\_monetary\_index \\
&+ \beta_6 log\_artifs\_rate + \beta_7 localities\_rate \\
&+ \beta_8 link + \beta_9 detector\_expensive\_dummy
\end{aligned}
$$

2. Model 2 (OLS - Log Transformed):

   Dependent Variable: *log_coins_rate*
   Independent Variables: *log_experience*, *log_contributions*, *log_comments*, *log_coins*, *real_net_monetary_index*, *log_artifs_rate*, *localities_rate*, *link*, *detector_expensive_dummy*

   Model Equation:

$$
\begin{aligned}
log\_coins\_rate = \beta_0 &+ \beta_1 log\_experience \\
&+ \beta_2 log\_contributions + \beta_3 log\_comments \\
&+ \beta_4 log\_coins + \beta_5 real\_net\_monetary\_index \\
&+ \beta_6 log\_artifs\_rate + \beta_7 localities\_rate \\
&+ \beta_8 link + \beta_9 detector\_expensive\_dummy
\end{aligned}
$$

3. Model 2_Robust (Robust OLS - Log Transformed):

Model Equation:

$$\begin{aligned} log\_coins\_rate = \beta_0 + &\beta_1 log\_experience \\ &+ \beta_2 log\_contributions + \beta_3 log\_comments \\ &+ \beta_4 log\_coins + \beta_5 real\_net\_monetary\_index \\ &+ \beta_6 log\_artifs\_rate + \beta_7 localities\_rate \\ &+ \beta_8 link + \beta_9 detector\_expensive\_dummy \end{aligned}$$

4. Model 2_WLS (Weighted Least Squares - Log Transformed):

   Model Equation:

$$\begin{aligned} log\_coins\_rate = \beta_0 + &\beta_1 log\_experience \\ &+ \beta_2 log\_contributions + \beta_3 log\_comments \\ &+ \beta_4 log\_coins + \beta_5 real\_net\_monetary\_index \\ &+ \beta_6 log\_artifs\_rate + \beta_7 localities\_rate \\ &+ \beta_8 link + \beta_9 detector\_expensive\_dummy \end{aligned}$$

LPM, Logit, Probit Models (Binary Dependent Variable Models)

Dependent Variable: *rate_coins_dummy*
Independent Variables: *log_experience*, *log_contributions*, *log_comments*, *log_coins*, *real_net_monetary_index*, *log_artifs_rate*, *localities_rate*, *link*, *detector_expensive_dummy*

1. Model 2_LPM (Linear Probability Model - Dummy Variable):

   Model Equation:

$$\begin{aligned} rate\_coins\_dummy = \beta_0 + &\beta_1 log\_experience \\ &+ \beta_2 log\_contributions + \beta_3 log\_comments \\ &+ \beta_4 log\_coins + \beta_5 real\_net\_monetary\_index \\ &+ \beta_6 log\_artifs\_rate + \beta_7 localities\_rate \\ &+ \beta_8 link + \beta_9 detector\_expensive\_dummy \end{aligned}$$

2. Model 2_Logit (Logit Model - Dummy Variable):

Model Equation:

$$\begin{aligned}P(rate\_coins\_dummy = 1|x) = \Lambda(&\beta_0 + \beta_1 log\_experience \\ &+ \beta_2 log\_contributions \\ &+ \beta_3 log\_comments + \beta_4 log\_coins \\ &+ \beta_5 real\_net\_monetary\_index \\ &+ \beta_6 log\_artifs\_rate + \beta_7 localities\_rate \\ &+ \beta_8 link \\ &+ \beta_9 detector\_expensive\_dummy)\end{aligned}$$

3. Model 2_Probit (Probit Model - Dummy Variable):

Model Equation:

$$\begin{aligned}P(rate\_coins\_dummy = 1|x) = \Phi(&\beta_0 + \beta_1 log\_experience \\ &+ \beta_2 log\_contributions \\ &+ \beta_3 log\_comments + \beta_4 log\_coins \\ &+ \beta_5 real\_net\_monetary\_index \\ &+ \beta_6 log\_artifs\_rate + \beta_7 localities\_rate \\ &+ \beta_8 link \\ &+ \beta_9 detector\_expensive\_dummy)\end{aligned}$$

5.10 Dataset 5: Ancient Artifacts, Full

5.10.1 Descriptive Statistics

Using this last dataset, we attempt to verify our estimates obtained by the analysis especially of Dataset 1 and Dataset 3. Particularly, we are concerned if the majority of the artifacts uploaded on the website are actually 'valuable' enough to be even considered for submission to the archaeological authority. Therefore, we created this last dataset that contains the uploaded artifacts originating in certain historical periods, so that they are likely considered valuable. Specifically, the periods that the artifacts in this dataset come from are the Bronze Age, Iron Age, Roman period, Migration of peoples, Avar-Slavic period 5th to 9th century, Early Middle Ages, and The Middle Ages. Therefore, out of all the artifacts uploaded on the website (203,846), this dataset contains 12,958 (15.73%) of them. It is important to note that this dataset is different from all the datasets used so far. Up to now, we used the datasets where individual observations were the profiles on the website. This time, the individual observations represent the individual artifacts uploaded. From Table 5.9, we can see the properties of the individual variables which we describe separately later on. Worth noting might be the variability of the *municipal_office* variable. This time, the dataset includes 252 different municipal offices and their respective characteristics.

From the summary statistics of the dataset (Table 5.10), we can see similar properties (large skew, potential outliers) of variables that we know from the other datasets. Those variables were matched to the respective individual artifact finds based on the profile that uploaded them. Those variables are namely: *link*, *experience*, *contributions*, *comments*, *artifacts*, *residence_additional_info*, *real_net_monetary_index*, *artifs_rate*, *detector_expensive_dummy*, and *localities_rate*. The properties of those variables are likely to be enhanced by the fact that we filled in the missing values of the *real_net_monetary_index*, *average_age*, and *localities_rate* variables (8,261 observations) with an average of the Czech Republic. When comparing an initial correlation matrix without the filled values and the correlation matrix of the dataset with filled average values, the respective correlations did not significantly change. In Table 5.10, we can further see new variables specific to this dataset that are briefly described in the following part.

Variable	Count	Unique	Mode	Frequency
artif_name	12958	12958	*	1
likes	12958	100	1.0	1573
viewed	12958	2615	673.0	23
comments_under	12958	84	0.0	2464
profile	12958	2426	Kvasak	322
uploaded	12958	14	2020	1483
detector_used	9187	1925	XP Deus	245
submitted_to	2415	94	**	167
period	12958	7	7.0	7114
link	12958	2	0.0	11488
experience	12958	928	0.0	1091
contributions	12958	98	0.0	5594
comments	12958	651	24463.0	322
artifacts	12958	288	3332.0	322
residence_additional_info	12958	2	0.0	12641
municipality	4712	386	Praha	459
municipal_office	4697	252	Praha	511
real_net_monetary_index	12958	253	1.0	8261
artifs_rate	12958	300	0.0	7076
average_age	12958	253	42.473492	8261
detector_expensive_dummy	12958	2	0.0	12598
localities_rate	12958	207	0.021478	8261
uploaded_year	12958	14	11.0	1483
submitted_to_dummy	12958	2	0.0	10543

Table 5.9: The main variables with the count of observations, the number of unique values, the mode, and the frequency of the mode - Dataset 5;
* Podkova lesní; ** Muzeum Komenského v Přerově

Variable	Mean	Std	Min	25%	50%	75%	Max
likes	7.1	10.3	0.0	2.0	4.0	8.0	229.0
viewed	1091.7	732.9	72.0	631.0	908.0	1336.8	12438.0
comments_under	5.2	8.0	0.0	1.0	3.0	6.0	209.0
uploaded	2017.6	3.0	2010.0	2015.0	2018.0	2020.0	2023.0
period	4.8	2.6	1.0	2.0	7.0	7.0	7.0
link	0.1	0.3	0.0	0.0	0.0	0.0	1.0
experience	8451.1	26413.4	0.0	91.0	957.5	5610.0	498425.0
contributions	30.3	93.2	0.0	0.0	1.0	17.0	1489.0
comments	1915.3	4254.4	0.0	76.0	347.0	1877.0	24463.0
artifacts	253.9	528.5	1.0	35.0	105.0	261.8	3332.0
residence_additional_info	0.0	0.2	0.0	0.0	0.0	0.0	1.0
real_net_monetary_index	1.0	0.0	0.8	1.0	1.0	1.0	1.2
artifs_rate	0.1	0.1	0.0	0.0	0.0	0.1	1.0
average_age	42.4	0.7	37.4	42.5	42.5	42.5	45.4
detector_expensive_dummy	0.0	0.2	0.0	0.0	0.0	0.0	1.0
localities_rate	0.0	0.0	0.0	0.0	0.0	0.0	0.1
uploaded_year	8.6	3.0	1.0	6.0	9.0	11.0	14.0
submitted_to_dummy	0.2	0.4	0.0	0.0	0.0	0.0	1.0

Table 5.10: Summary Statistics of the Dataset 5

1. *likes*

   This variable contains the number of likes that an individual artifact obtained from other users. From Table 5.9, we can see that the most frequent number of likes is 1 (1,573 observations). Overall, the number of likes ranges from 0 to 229 with a median of 4 likes. The histograms of this variable highlight the overall large number of low values (see Appendix).

2. *viewed*

   The *viewed* variable represents the number of times the specific artifact was viewed. The mean number of views is almost 1,100 per artifact. Interestingly, as might be inferred from looking at the histogram of this variable and its logarithmic transformation in the Appendix, though discrete, this variable resembles a log-normal distribution; since after the logarithmic transformation, the histogram looks like a normal distribution.

3. *comments_under*

   This variable accounts for the number of comments that were written under the specific individual artifact post. The median number is 3 comments, and there are overall 2,464 artifacts with no comments. The logarithmic transformation of this variable seems to have helped with the skew of its distribution.

4. *period*

   The *period* categorical variable accounts for the historical period to which a certain artifact belongs. The respective historical periods are represented by the following values:

   Bronze Age = 1
   Iron Age = 2
   Roman period = 3
   Migration of peoples = 4
   Avar-Slavic period 5th to 9th century = 5
   Early Middle Ages = 6
   The Middle Ages = 7

   The histogram in the Appendix shows that the highest number of observations comes from the Middle Ages (7,114), followed by the Bronze Age period. We do not transform this variable.

5. *uploaded_year*

   Another categorical variable *uploaded_year* represents the year in which the artifact was uploaded. The artifacts were uploaded during the period of 14 years, and this variable is therefore defined as follows:

   2010 = 1
   2011 = 2
   2012 = 3
   2013 = 4
   2014 = 5
   2015 = 6
   2016 = 7
   2017 = 8
   2018 = 9
   2019 = 10
   2020 = 11
   2021 = 12
   2022 = 13
   2023 = 14

The histogram of this variable, as can be seen in the Appendix, suggests the volume of uploading the artifacts from the respective periods over time. The trend of uploading those artifacts seems to be decreasing since the year 2020.

6. *submitted_to_dummy* - The Dependent Variable

   The dependent variable used for all the models specific to this dataset is the binary dummy variable called *submitted_to_dummy*. Since we have observations about whether each of the artifacts was submitted to the archeological authority or not, we define this variable as having value 1 if the artifact was submitted, 0 otherwise:

$$submitted\_to\_dummy = \begin{cases} 1 & \text{artifact submitted} \\ 0 & \text{artifact not submitted} \end{cases}$$

5.10.2 Models

Since our only dependent variable is binary, we estimate only three of the models, LPM, Logit, and Probit. After assessing the impact of potentially influential variables, we decided to delete them in the same way as for all the previous datasets. The main reason was that, when estimating the LPM model with those outliers, we spotted potential multicollinearity (VIF $>$ 5 for two variables). After deleting the potential over-influential observations, the multicollinearity was no longer present. Moreover, the fit of the models improved. We also compared the models without and with the logarithmic transformation of some variables. The conclusion is that the models with logarithmic transformation perform better (higher R^2, AUC) than those without. Hence, the estimated models are the following:

LPM, Logit, Probit Models (with Logarithmic Transformation)

Dependent Variable: *submitted_to_dummy*
Independent Variables: *period*, *uploaded_year*, *log_likes*, *log_viewed*, *log_comments_under*, *link*, *log_experience*, *log_contributions*, *log_comments*, *log_artifacts*, *residence_additional_info*, *real_net_monetary_index*, *log_artifs_rate*, *average_age*, *detector_expensive_dummy*, *localities_rate*

1. Model2_LPM (Linear Probability Model - Dummy Variable):

Model Equation:

$$\begin{aligned} submitted\_to\_dummy = \beta_0 &+ \beta_1 period \\ &+ \beta_2 uploaded\_year + \beta_3 log\_likes \\ &+ \beta_4 log\_viewed + \beta_5 log\_comments\_under \\ &+ \beta_6 link + \beta_7 log\_experience \\ &+ \beta_8 log\_contributions + \beta_9 log\_comments \\ &+ \beta_{10} log\_artifacts \\ &+ \beta_{11} residence\_additional\_info \\ &+ \beta_{12} real\_net\_monetary\_index \\ &+ \beta_{13} log\_artifs\_rate \\ &+ \beta_{14} average\_age \\ &+ \beta_{15} detector\_expensive\_dummy \\ &+ \beta_{16} localities\_rate \end{aligned}$$

2. Model2_LOGIT (Logit Model - Dummy Variable):

   Model Equation:

$$\begin{aligned} P(submitted\_to\_dummy = 1|x) = \Lambda(\beta_0 &+ \beta_1 period \\ &+ \beta_2 uploaded\_year + \beta_3 log\_likes \\ &+ \beta_4 log\_viewed \\ &+ \beta_5 log\_comments\_under \\ &+ \beta_6 link + \beta_7 log\_experience \\ &+ \beta_8 log\_contributions \\ &+ \beta_9 log\_comments \\ &+ \beta_{10} log\_artifacts \\ &+ \beta_{11} residence\_additional\_info \\ &+ \beta_{12} real\_net\_monetary\_index \\ &+ \beta_{13} log\_artifs\_rate \\ &+ \beta_{14} average\_age \\ &+ \beta_{15} detector\_expensive\_dummy \\ &+ \beta_{16} localities\_rate) \end{aligned}$$

3. Model2_PROBIT (Probit Model - Dummy Variable):

   Model Equation:

$$
\begin{aligned}
P(submitted\_to\_dummy = 1|x) = \Phi(&\beta_0 + \beta_1 period \\
&+ \beta_2 uploaded\_year + \beta_3 log\_likes \\
&+ \beta_4 log\_viewed \\
&+ \beta_5 log\_comments\_under \\
&+ \beta_6 link + \beta_7 log\_experience \\
&+ \beta_8 log\_contributions \\
&+ \beta_9 log\_comments \\
&+ \beta_{10} log\_artifacts \\
&+ \beta_{11} residence\_additional\_info \\
&+ \beta_{12} real\_net\_monetary\_index \\
&+ \beta_{13} log\_artifs\_rate \\
&+ \beta_{14} average\_age \\
&+ \beta_{15} detector\_expensive\_dummy \\
&+ \beta_{16} localities\_rate)
\end{aligned}
$$

Chapter 6

Results

This chapter presents the results of the main hypotheses tested using the described models and respective measures. The results are presented for each of the datasets separately, comparing them stepwise to the results obtained from the other datasets. We start with Dataset 1 and follow up to the final Dataset 5, using it mainly as a robustness check to evaluate the performance and consistency of the models used. Additionally, a sensitivity analysis is provided by comparing the model estimates of LPM, Logit, and Probit for each of the datasets.

6.1 Dataset 1: Artifacts, Full

For this part, we use Dataset 1 to estimate the probability of submitting an artifact (*rate_artifs_dummy* dependent variable) using logit and probit models. We also use the LPM model with heteroskedasticity-robust standard errors. The LPM models in all other sections are also heteroskedasticity-robust. The results are given in Table 6.1.

All three models are statistically significant, with their respective p-values less than 0.01. The pseudo-R2 in the LPM is a standard R-squared as reported for OLS. Specifically, in our case, it attains a value of 0.183, indicating that about 18% of the dependent variable variation is explained by the independent variables included in the model. The pseudo-R2 for Logit and Probit models are McFadden's pseudo-R2s, based on the log-likelihoods described in one of the earlier sections. Specifically, those attain the values of 0.2646 and 0.2656 for Logit and Probit, respectively. Overall, the Probit model seems to perform the best out of the three models, as its R2, Log-likelihood, Percentage correctly predicted, and AUC are slightly higher than those of the other two models. On the other hand, though having only slightly lower statistics of goodness of fit, the LPM seems to perform the worst out of the three models. The respective AUC, Percentage correctly predicted, Log-likelihood, and Pseudo R2 of the models can be seen in Table 6.1; the ROC and Confusion Matrix of the Logit and Probit models can be seen in the Appendix.

The estimates from all three models are consistent compared to each other, with slight differences between the LPM and the Logit and Probit models. For example, the *log_contributions* variable is statistically significant at the 5% significance level in the LPM model, but not in Logit and Probit. On the other hand, the *residence_additional_info* variable is significant at the 5% significance level in Logit and Probit, but not in the LPM. The same holds for an intercept (constant) term. So, overall, both Logit and Probit are consistent with each other, while LPM indicates the same direction of all the variables' coefficients but differs in the significance of two of the variables. Nevertheless, those variables are not very important for our analysis.

Moreover, the *log_experience* variable is statistically significant at the 5% significance level with positive coefficients in all models. This possibly means that if one is more active and successful in helping others with identifying their finds, one is more likely to submit their own artifact. This is an interesting observation since this variable possibly indicates a certain degree of pro-sociality. The *log_comments* variable exhibits very similar properties. However, the *log_contributions* variable, which we assumed might be most correlated with pro-sociality, is not significant at the 5% significance level in the Logit and Probit models. Furthermore, inferring from the properties of the *log_artifacts* variable estimates, it is statistically significant with a positive coefficient. This potentially means that if one has more artifacts, one is more likely to submit at least one of them. Following on to the *log_coins_rate* variable, its estimates indicate that this variable is significantly positively associated with the submission of an artifact. This possibly means that if one is more likely to submit a coin, then one is more likely to submit an artifact as well. This might not be surprising, as coins might be considered a subset of artifacts. Next, providing a *link* to the website does not significantly influence the submission of an artifact. On the other hand, if one provided additional information such as one's age, it proved statistically significant at the 5% significance level with a positive estimate, hence when one is more likely to disclose personal information, they can be more likely to submit an artifact.

Next, as mentioned before, the magnitudes of the coefficients across the models are not directly comparable (Wooldridge (2012)). Therefore, we would compute the scale (adjustment) factors respective to the PEA and APE. Nevertheless, in our case, it might not be necessary since we are primarily interested in the *real_net_monetary_index*, *detector_expensive_dummy*, and *localities_rate* variables, and those have the following properties.

Although having a negative coefficient in all three models, our key independent variable, the *real_net_monetary_index*, is not statistically significant in either of the models. Hence, we reject the **H1** that the higher the *real_net_monetary_index*, the higher the probability of submitting an artifact at the 5% significance level. In other words, we do not have enough evidence to reject the null hypothesis that the *real_net_monetary_index* variable has no effect (its coefficient is equal to 0) on submitting an artifact at the 5% significance level. On the other hand, the next proxy for the socio-economic status, the *detector_expensive_dummy*, is statistically significant, however, only at the 10% significance level in both the Logit and Probit models. It has a positive coefficient, compared to the previous variable, potentially indicating that if one owns an 'expensive' metal detector, one is more likely to submit an artifact. This holds at the 10% significance level, however, meaning that we reject the **H2** stating that owning a more expensive metal detector means a higher probability of submitting an artifact at the 5% significance level.

Similarly, we take a look at the *localities_rate* variable, as this variable is subject to our third hypothesis, **H3**. It is significant at the 10% significance level not only in the Logit and Probit models but also in the LPM. It also has a positive coefficient in all three cases, potentially indicating that the higher the density of archaeological localities in the area one lives in, the higher the probability of submitting a find. Nevertheless, this holds for the 10% significance level. Hence, also in this case, we reject the hypothesis **H3** that the higher the *localities_rate* in a given area one lives in, the higher the probability that this individual submits an artifact, at the 5% significance level.

The above results might raise a question if some of the key independent variables are jointly statistically significant. Therefore, we conduct a Wald test which is used for testing the joint significance of a specific set of variables in the Tobit, Logit, and Probit models (Wooldridge (2012)).

First, the Wald test of joint significance of all three main independent variables raises a p-value of approximately 0.058 and 0.085 for Logit and Probit, respectively, not rejecting the null hypothesis that the *real_net_monetary_index*, *detector_expensive_dummy*, and *localities_rate* variables are not statistically significant at the 5% significance level. Second, based on the Wald test p-values (0.161 and 0.187 for Logit and Probit, respectively) of the joint significance of the *localities_rate* and *real_net_monetary_index*, we do not have enough evidence to state that those two variables are jointly significant. And finally, testing the joint significance of the *detector_expensive_dummy* and *localities_rate* variables, we get the p-values of the Wald test for Logit of 0.0243 and 0.0367 for Probit. Therefore, with 95% confidence, we reject the null hypothesis that those two variables are not jointly statistically significant. Therefore, the *detector_expensive_dummy* and *localities_rate* variables are jointly statistically significant, potentially meaning that if one owns an 'expensive' metal detector and at the same time lives in an area with a higher density of archaeological localities, that individual has a higher probability of submitting an artifact; since the estimated coefficients of both variables are also positive in all of the models. One might argue that this could be potentially explained by a higher probability of a valuable find thanks to the potential higher density of those valuable finds in the area; and, at the same time, there might be also a higher probability of a valuable find thanks to the better technical equipment of the detector (i.e. a more expensive metal detector likely has better sensitivity, for example). Moreover, we know that there is a positive association between the number of artifacts and the submission of the artifact, which might, together with the higher probability of finding an artifact thanks to either the better detector or density of the sites, form the effect. Nevertheless, in our model, it is controlled for the number of artifacts, hence it is likely not the higher chance of finding a valuable artifact (either thanks to locality or detector) that primarily drives the effect. Therefore, together with the assumption that an individual owning an 'expensive' metal detector has a higher socioeconomic status, we might conclude that if one is richer and lives in an area with a higher density of archaeological localities, that individual is more likely to submit an artifact. This observation might indirectly support the hypothesis about the more prevailing hobbyist motivation for metal detecting amongst the richer.

Dependent Variable: rate_artifs_dummy			
Independent Variables	LPM (OLS)	Logit (MLE)	Probit (MLE)
log_experience	0.0109***	0.1260***	0.0686***
	(0.002)	(0.026)	(0.013)
log_contributions	0.0260***	0.0670	0.0514*
	(0.006)	(0.052)	(0.029)
log_comments	0.0055**	0.1602***	0.0689***
	(0.003)	(0.044)	(0.023)
log_artifacts	0.0416***	0.5192***	0.2762***
	(0.003)	(0.052)	(0.027)
real_net_monetary_index	-0.0080	-0.2855	-0.1665
	(0.048)	(1.019)	(0.510)
log_coins_rate	1.4722***	10.6671***	5.2977***
	(0.233)	(1.313)	(0.554)
localities_rate	0.3305*	7.7846*	3.7713*
	(0.198)	(4.093)	(2.063)
link	0.0195	0.0248	0.0225
	(0.022)	(0.245)	(0.132)
residence_additional_info	0.0728*	0.8341**	0.4576**
	(0.040)	(0.327)	(0.182)
detector_expensive_dummy	0.0346	0.6023*	0.2999*
	(0.027)	(0.310)	(0.167)
constant	-0.0570	-4.9257***	-2.6005***
	(0.048)	(1.008)	(0.503)
Percentage correctly predicted	92.84%	93.25%	93.29%
Log-likelihood value	-	-1478.8	-1476.8
Pseudo R-squared	0.183	0.2646	0.2656
AUC	-	0.8578	0.8579

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors are shown in parentheses.

Table 6.1: LPM, Logit, and Probit Estimates - Dataset 1

6.2 Dataset 2: Coins, Full

In this section, we estimate the models related to the submission of coins using Dataset 2. The dependent variable is *rate_coins_dummy*, which takes the value 1 if one submitted at least one coin, 0 otherwise. As in Dataset 1, we use the LPM with (White's) heteroskedasticity-robust standard errors, as well as Logit and Probit models. The actual specific estimated models can be seen in the section describing the properties of Dataset 2. The results of the estimated models are given in Table 6.2.

All three models are statistically significant, with p-values much lower than 1%. When compared to the models in Dataset 1 all three models perform better. For example, the Percentage Correctly Predicted is about 97.5% for the models in Dataset 2, whereas about 93% in Dataset 1. The same holds for the AUCs. In the previous dataset, those attained a value of about 0.86, this time it is about 0.90 which is closer to one, hence better. Overall, it seems that the models in this coin-specific dataset fit the data better than in Dataset 1, which can be observed on higher Log-likelihood values as well as higher Pseudo R-squareds. In the case of the Logit and Probit, about 34.6% of the variation is explained.

Comparing the models specific to Dataset 2, they seem to perform quite well and are comparable in terms of Percentage Correctly Predicted. The Probit and Logit models are very similar in terms of log-likelihoods and R-squareds, with Logit leading slightly. Nevertheless, the AUC is higher for the Probit model. The estimates and associated p-values of the variables are consistent within the three models, with only LPM differing in the *log_contributions*, *log_comments*, and a constant term. Consistent with the previous dataset's estimate for the number of artifacts, the *log_coins* variable measuring the number of coins uploaded is significantly positively related to submitting at least one coin. Similarly, the rate of artifact submission is significant and positively related to the submission of coins.

Now, we focus on our key independent variables, the *real_net_monetary_index*, *localities_rate*, and *detector_expensive_dummy*. First, we reject the **H1** that the *real_net_monetary_index* is positively associated with submitting the coin at the 5% significance level. In fact, we can reject this hypothesis even at the 80% level of significance for all three models. We continue with the *localities_rate* variable, rejecting the **H3** that a higher density of localities is positively associated with submitting at least one coin. Nevertheless, the *detector_expensive_dummy* has a positive and statistically significant coefficient with a p-value less than 5% for the LPM, and p-values being even less than 1% for the Logit and Probit models. Hence, we do not reject **H2**, that if one owns an 'expensive' metal detector, that individual is more likely to submit an artifact, at the 5% level of significance. Interestingly, the *detector_expensive_dummy* is statistically significant (except the WLS model which is deemed not reliable as explained earlier) at the 5% significance level also in all the other models we estimated (OLS without logarithmic transformation, OLS with log transformation, heteroskedasticity robust OLS). Therefore, additionally, for this variable, we compute the PEA and APE for the Logit and Probit models, which can be seen in Table 6.3.

Dependent Variable: rate_coins_dummy			
Independent Variables	LPM (OLS)	Logit (MLE)	Probit (MLE)
log_experience	0.0005	0.0426	0.0199
	(0.001)	(0.051)	(0.023)
log_contributions	0.0170***	0.1143	0.0601
	(0.004)	(0.085)	(0.042)
log_comments	0.0009	0.2522***	0.1056***
	(0.002)	(0.084)	(0.040)
log_coins	0.0174***	0.5661***	0.2437***
	(0.003)	(0.112)	(0.052)
real_net_monetary_index	0.0060	-0.1142	0.1528
	(0.033)	(2.133)	(0.974)
log_artifs_rate	1.0789***	11.4004***	5.6698***
	(0.116)	(0.844)	(0.407)
localities_rate	0.0949	9.3570	2.3490
	(0.148)	(8.544)	(3.937)
link	0.0230	-0.0816	0.0412
	(0.018)	(0.403)	(0.202)
residence_additional_info	0.0108	0.1827	0.0066
	(0.032)	(0.612)	(0.290)
detector_expensive_dummy	0.0603**	1.4639***	0.6621***
	(0.027)	(0.411)	(0.210)
constant	-0.0390	-7.0666***	-3.5581***
	(0.033)	(2.112)	(0.966)
Percentage correctly predicted	97.44%	97.57%	97.55%
Log-likelihood value	-	-432.66	-432.73
Pseudo R-squared	0.194	0.3462	0.3461
AUC	-	0.9007	0.9029

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors are shown in parentheses.

Table 6.2: LPM, Logit, and Probit Estimates - Dataset 2

From the results of the PEA and APE, we can see that this adjusted coefficients of the *detector_expensive_dummy* in the Logit and Probit models are quite close to each other, especially the PEA attaining the values of 0.0287 and 0.0278 for Logit and Probit, respectively. On the other hand, the LPM coefficient is more than two times higher. We might interpret the coefficient of the *detector_expensive_dummy* variable for the LPM as if one owns an 'expensive' metal detector, the probability of submission of the coin is about 0.0603 higher than without owning an 'expensive' metal detector. However, there are certain drawbacks of the LPM estimates which we covered earlier. On the other hand, the Logit and Probit estimates are deemed more reliable. So, we might interpret for example the Probit PEA of the *detector_expensive_dummy* variable such that the ownership of an expensive metal detector is estimated to increase the probability of submitting the coin by 0.0278. Overall, we might infer that, based on this dataset and the estimated models, owning an 'expensive' metal detector possibly increases the probability of submitting the coin at the 5% significance level. Similarly, as for Dataset 1, if we assume that ownership of an expensive metal detector is positively associated with the wealth of an individual, it might mean that if one has a higher socioeconomic status, one is more likely to submit a coin. This time, however, in contrast to Dataset 1, it is based not only on the joint but also individual significance of the *detector_expensive_dummy* variable.

Dependent Variable: rate_coins_dummy		
Independent Variable:	detector_expensive_dummy	p-value
LPM	0.0603	0.026
	(0.027)	
Logit (PEA)	0.0287	0.000
	(0.008)	
Probit (PEA)	0.0278	0.002
	(0.009)	
Logit (APE)	0.0118	0.001
	(0.004)	
Probit (APE)	0.0146	0.003
	(0.005)	

Standard errors are shown in parentheses.

Table 6.3: PEA/APE for the detector_expensive_dummy Variable - Dataset 2

6.3 Dataset 3: Artifacts, Reduced

The analysis of Dataset 3 is conducted similarly to the previous two datasets. Particularly, this dataset focuses on the estimation of submitting or not submitting the artifacts, similar to Dataset 1. Hence, the dependent variable is *rate_artifs_dummy*, taking the value 1 if an individual submitted at least one artifact, 0 otherwise. In this section, we analyze the data that do not include filled-in data of the average of the Czech Republic for the socioeconomic and demographic characteristics, hence examining the potential impact of filling those values in the first sample. The results of the models specific to this dataset can be seen in Table 6.4.

All three models are statistically significant at almost any significance level, based on the F-statistic p-value in the case of LPM and LLR p-value in the case of Logit and Probit. If we compare the performance of the models to the models in the first dataset, all three models perform slightly better in terms of Percentage correctly predicted, attaining a value of about 94% correctly predicted cases, which is 1% higher than before. However, we need to be careful with the interpretation of this measure, since, due to the reduction of the data, we now have fewer observations of the dependent variable taking the value of 1; thus we might have a higher Percentage correctly predicted in this sample when none of the ones were correctly predicted than in the same scenario in Dataset 1. The AUC, on the other hand, is about 0.84 for the Logit and Probit models, which is almost a 2% decrease compared to Dataset 1. When looking at the R-squareds, the models in the first dataset also seem to fit the data better. Overall, the performance of the models respective to this dataset might be slightly worse than in Dataset 1. Nevertheless, all the measures are still very close to the measures of the models obtained in the first dataset.

Comparing the three models, LPM, Logit, and Probit, the last mentioned seem to perform the best in all terms, though there are no major differences compared to the other two models. Similarly for the parameter estimates. The only differences are the significance of the *log_contributions* variable at the 10% significance level compared to non-significance of the variable in the Logit and Probit models at this significance level. On the other hand, the *log_comments* variable in the LPM is not significant at the 5% level unlike in Logit and Probit. Except from the constant term, there is one more variable whose significance differs in LPM; the *localities_rate* variable is significant at the 10% level of significance in the Logit and Probit only.

When we compare the estimates of the models to the estimates from Dataset 1, the results do not significantly differ. Although, in the LPM, the significance of four variables decreased slightly, in the Logit and Probit, all variables, except *residence_additional_info*, are either significant or not significant at the 5% significance level the same way as in the model results in Dataset 1. Moreover, the estimated coefficients of the significant variables are very close to each other in both datasets.

Now, we focus on the key independent variables related to our hypotheses. First, although having a negative coefficient in all three models and in both datasets, the *real_net_monetary_index* variable is not significant. Therefore, we reject the **H1**. This is the same as in the first dataset. Second, the *localities_rate* variable is significant in Logit and Probit only at the 10% significance level. Thus, we also reject the **H3** at the 5% level of significance. And finally, the *detector_expensive_dummy* became insignificant even at the 10% significance level, compared to the first dataset. Therefore, we reject also **H2** at the 5% significance level. Those results are in line with Dataset 1. However, testing the joint significance of the three variables using the Wald test, no combination of those three key independent variables is statistically significant at the 5% significance level. In the first dataset, we concluded with the joint significance of the *localities_rate* and *detector_expensive_dummy.* Indeed, even here, the combination of those two proved the most significant; however not enough to reject the null hypothesis of not significance (although very close) even at the 10% level of significance. Since the *localities_rate* variable is significant in the same way as in the first dataset, it is likely that this joint insignificance (at 5% significance level, see Dataset 1), of the *detector_expensive_dummy* and *localities_rate* variables, is due to the loss of observations of the *detector_expensive_dummy.* In fact, in Dataset 1 there are 123 observations of the *detector_expensive_dummy* attaining the value of 1, whereas in this Dataset 3, there are only 72 observations of the *detector_expensive_dummy* variable having a value of 1. This means a loss of 51 observations of this variable having the value of one, which is about a 41.5% decrease in the valuable information that is associated with this dummy variable compared to the first dataset. Moreover, since this dummy variable is not subject to filling-in the average values of the Czech Republic, and this dataset was primarily supposed to verify the estimates of the variables that were subject to this filling-in of the variables, we conclude that the estimates of this dataset with respect to the *detector_expensive_dummy* variable are not as trustworthy as in the first dataset. Nevertheless, we take them into account.

Dependent Variable: rate_artifs_dummy			
Independent Variables	LPM (OLS)	Logit (MLE)	Probit (MLE)
log_experience	0.0115***	0.1572***	0.0825***
	(0.002)	(0.040)	(0.019)
log_contributions	0.0159*	-0.0066	0.0171
	(0.009)	(0.086)	(0.047)
log_comments	0.0070*	0.1618**	0.0692**
	(0.004)	(0.067)	(0.033)
log_artifacts	0.0324***	0.4623***	0.2427***
	(0.004)	(0.078)	(0.039)
real_net_monetary_index	-0.0047	-0.2618	-0.1503
	(0.048)	(1.004)	(0.503)
log_coins_rate	1.4684***	10.2656***	4.9539***
	(0.336)	(1.874)	(0.758)
localities_rate	0.3139	7.3308*	3.4668*
	(0.197)	(4.061)	(2.048)
link	0.0273	0.2322	0.1150
	(0.027)	(0.334)	(0.179)
residence_additional_info	0.0113	0.5290	0.3310
	(0.047)	(0.660)	(0.325)
detector_expensive_dummy	0.0288	0.4769	0.2556
	(0.035)	(0.425)	(0.223)
constant	-0.0454	-4.8348***	-2.5384***
	(0.048)	(1.000)	(0.499)
Percentage correctly predicted	94.08%	94.37%	94.51%
Log-likelihood value	-	-735.99	-735.15
Pseudo R-squared	0.151	0.2302	0.2311
AUC	-	0.8406	0.8411

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors are shown in parentheses.

Table 6.4: LPM, Logit, and Probit Estimates - Dataset 3

6.4 Dataset 4: Coins, Reduced

This dataset specific to coins consists of the observations of individuals that uploaded at least one coin and at the same time their respective municipal office was matched. Hence, the dependent variable, in this case, is *rate_coins_dummy*, and we compare the model results of this dataset with the results of Dataset 2. The results of the estimated LPM, Logit, and Probit can be found in Table 6.5.

First, unlike in the previous section comparing the two datasets (Dataset 1 and 3) subject to artifacts, this time the estimated models of the datasets with respect to coins (Dataset 2 and 4) seem to be more consistent between each other. The respective AUCs of the models of slightly above 0.9 are almost the same as well as the Percentage correctly predicted. Overall, it seems that the models respective to this Dataset 4 fit the data even slightly better than the original Dataset 2. We have now, despite dropping the *residence_additional_info* variable off the models, even greater pseudo R2 for the Logit and Probit models of 0.3742 and 0.3718, respectively.

The individual estimates in the LPM, Logit, and Probit models are all in line with the previous dataset subject to coins, having similar estimated coefficients as well as the respective variables' significance at the 5% significance level. Moving on to the estimates of our key independent variables, the *real_net_monetary_index* variable is again not significant; thus we reject the **H1**. Similarly, we reject the **H3** due to the non-significance of the *localities_rate* variable. Finally, the *detector_expensive_dummy* variable is significant. Thus, we reject the null hypothesis that its coefficient is zero at the 5% level of significance, meaning that we do not reject our **H2**. All those results are completely in line with the second dataset. Interestingly, the drop in the number of ones in the *detector_expensive_dummy* variable did not have an influence on the significance of the results; what is more, the estimated coefficients have an even greater magnitude this time. Therefore, we again compute the PEA/APE for better comparison of the estimated parameters' magnitude. The results can be seen in Table 6.6.

Interestingly, the estimated marginal effects are remarkably close to the marginal effects in Table 6.3. This holds especially for both, Logit and Probit models APE, which for Dataset 2 are 0.0118 and 0.0146 for Logit and Probit, respectively, and 0.0114 and 0.0140 for Logit and Probit, respectively, for this Dataset 4. Hence, using, for example, the APE of the *detector_expensive_dum-*

Dependent Variable: rate_coins_dummy			
Independent Variables	LPM (OLS)	Logit (MLE)	Probit (MLE)
log_experience	0.0008	0.0621	0.0271
	(0.001)	(0.086)	(0.038)
log_contributions	0.0137**	-0.0253	-0.0017
	(0.006)	(0.151)	(0.073)
log_comments	0.0036	0.4411***	0.1938***
	(0.002)	(0.140)	(0.064)
log_coins	0.0115**	0.4013**	0.1584**
	(0.005)	(0.176)	(0.078)
real_net_monetary_index	0.0094	-0.0216	0.1808
	(0.033)	(2.190)	(1.001)
log_artifs_rate	0.9125***	11.1976***	5.4585***
	(0.153)	(1.275)	(0.592)
localities_rate	0.1048	10.4593	2.4676
	(0.147)	(8.912)	(4.085)
link	0.0284	-0.0537	0.0193
	(0.020)	(0.598)	(0.304)
detector_expensive_dummy	0.1053**	2.1507***	0.9749***
	(0.041)	(0.519)	(0.266)
constant	-0.0399	-7.6755***	-3.7709***
	(0.033)	(2.202)	(1.009)
Percentage correctly predicted	98.09%	98.20%	98.16%
Log-likelihood value	-	-173.85	-174.52
Pseudo R-squared	0.190	0.3742	0.3718
AUC	-	0.9056	0.9092

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors are shown in parentheses.

Table 6.5: LPM, Logit, and Probit Estimates - Dataset 4

my in the Probit model, we might state that if one owns an ‘expensive’ metal detector, one is 0.014 more likely to submit a coin. This is a consistent observation with the original dataset, supporting the hypothesis that, assuming the more expensive metal detector implies higher socioeconomic status, richer individuals have a higher probability of submitting coins when found.

Dependent Variable: rate_coins_dummy		
Independent Variable:	detector_expensive_dummy	p-value
LPM	0.1053	0.010
	(0.041)	
Logit (PEA)	0.0315	0.000
	(0.008)	
Probit (PEA)	0.0316	0.001
	(0.009)	
Logit (APE)	0.0114	0.002
	(0.004)	
Probit (APE)	0.0140	0.004
	(0.005)	

Standard errors are shown in parentheses.

Table 6.6: PEA/APE for the detector_expensive_dummy Variable - Dataset 4

6.5 Dataset 5: Ancient Artifacts, Full

Using a different dataset on the individual artifacts level allowed us to add an additional set of variables to the models. Therefore, here we do not present only the robustness check of the hypotheses testing in the previous datasets but also other potentially valuable insights into the trends connected to the submission of artifacts. In the LPM, Logit, and Probit models, we use a dependent variable called *submitted_to_dummy* which takes a value of 1 if a particular artifact was submitted, and 0 if not. The results of the estimated models are given in Table 6.7.

First, the LPM, Logit, and Probit seem to perform well and consistently with respect to each other. The Percentage correctly predicted might be in this case considered a more reliable measure of fit than in the previous datasets, since the proportion of ones in our dependent variable is higher. That is also reflected in the Confusion matrix (in the Appendix), where the models (Logit and Probit) estimate true positives (TP) in approximately 1,400 cases and True Negatives (TN) in approximately 9,800 cases. Hence the fraction of TP and TN is much higher than in the previous datasets. The respective Percentage correctly predicted is slightly above 89 percent for the Logit and Probit. Furthermore, we obtained a relatively high R2 compared to all other datasets, now with the variation of the independent variables explaining about 44.3% dependent variable variation in the case of LPM and 47.8% and 47.6% in the case of Logit and Probit, respectively. The AUC of the two latter models of almost 0.94 is also the highest among all models estimated so far.

Second, we compare the estimates of the common independent variables with other datasets; most often with Datasets 1 and 3. Although the *link* variable was insignificant in all other datasets, this time it proved significant for Logit and Probit. Moreover, its estimates have a negative sign, potentially meaning that if one provided a link to other social media, one might be less likely to submit an artifact from the particular group of ancient artifacts that this dataset consists of. Next, the *log_experience* variable is significant in all the datasets respective to artifacts, holding also for this dataset. Nevertheless, this variable is significant neither in Dataset 2 nor in Dataset 4, which are both subject to coins. Therefore, interestingly, if one is helping others on the website with identifying their finds, one might be more likely to submit an artifact but not a coin. Next, we have a *log_contributions* variable, which is also significant, unlike in the previous datasets. The positive estimated coefficients of this variable possibly mean that if one is more active in writing contributions on the website, one is more likely to submit an ‘ancient’ artifact. Then, the *log_comments* variable was significant in all previously estimated models, which is also the case here. However, this time it has a negative estimate, potentially meaning that, overall, if one is more active in commenting finds of others, one is more likely to submit a find; however, when it comes to ‘ancient’ artifacts, it might be vice versa, i.e. one could be less likely to submit an artifact in this category. Then, the significance of *log_artifacts* variable suggests that overall if one has more artifacts, one might be more likely to submit one of them. This is in line with other datasets. Next, we have a *log_artifs_rate* variable, which we added to the models. It proved significant, likely meaning that if one is more likely to submit an artifact overall (in the full group of artifacts, not specific to any period), one is more likely to submit an artifact in the ‘ancient’ artifacts category as well. Finally, we added one more variable, which is an *average_age* in the respective municipal office one resides in. This variable, however, did not prove significant.

Third, in Table 6.7 we can see that there are five other variables that did not appear in any of the previous datasets. These are *period*, *uploaded_year*, *log_likes*, *log_viewed*, and *log_comments_under*. Their description is provided in the respective chapter on descriptive statistics of Dataset 5. Those variables might provide valuable insights into the trends of submission of artifacts. First, we have the *period* variable. This variable is statistically significant (at the 5% significance level), and its interpretation might be that it is more likely for the artifacts to be submitted as they are older. That might mean, for example, that if one finds an artifact from the Bronze Age, one is more likely to submit this artifact than when one finds, for example, an artifact from the Middle Ages. The negative signs of this variable’s estimates in Table 6.7 are due to the definition of the variable attaining a higher number when coming from a more recent historical period. Next, we have the *uploaded_year* variable. This variable is also significant at the 5% level of significance and similarly to the previous variable, attains a higher number representing a more recent period. This time, however, the variable indicates a year in which a given artifact was uploaded to the website. The positive estimated coefficient may be interpreted such that since the year 2010 with each additional year, it is more probable that an artifact (from the ‘ancient’ group) will be submitted; all up to the year 2023. This might support the claims about the improving relationships between archaeologists and metal detectorists (supporting, for example, Komoróczy (2022)). Moving on, the estimated parameter of *log_likes* is also positive and statistically significant. This might mean that if an artifact is submitted, it is positively associated with an increased number of likes; thus, the submission of the artifact is likely regarded by the website community as an exemplary act (supporting, for example, Hajšman *et al.* (2019)). Then, the significance and positive estimate for the *log_viewed* variable might indicate that if an artifact is submitted, the number of views of that particular artifact is higher. This might perhaps be the case since that specific artifact appears to be more interesting. Hence possibly, if an artifact is more interesting, it might be more likely to be submitted. On the other hand, one can see right away on the website, even without viewing the particular artifact if it was submitted or not, so it might be difficult to distinguish if that particular artifact was viewed more due to the fact that it is interesting, or, due to the fact that it was submitted. And the final newly added variable is the *log_comments_under*. It is also significant at the 5% significance level; however, it has a negative estimate. This might be a surprising finding, potentially interpreted as if the specific artifact was submitted, there were fewer comments written under the post of this artifact. Why is it the case is up for debate. The only potential explanation might be that other unsubmitted artifacts obtained more comments that were asking why the given artifact is not submitted, hence being probably more controversial.

Finally, we move on to the independent variables of our primary interest. Those are *real_net_monetary_index*, *detector_expensive_dummy*, and *localities_rate*. Neither of those is statistically significant at the 5% significance level. Therefore, we do not have enough evidence not to reject **H1**, **H2**, as well as **H3**. Moreover, testing the joint significance of all possible combinations of those three variables using the Wald test, we do not have enough evidence to reject the null hypothesis of their joint insignificance. Those findings support the hypotheses testing results in the previous datasets, especially Datasets 1 and 3 that were focused on artifacts.

Dependent Variable: submitted_to_dummy			
Independent Variables	LPM (OLS)	Logit (MLE)	Probit (MLE)
period	-0.0174***	-0.1943***	-0.1035***
	(0.001)	(0.013)	(0.007)
uploaded_year	0.0153***	0.1589***	0.0913***
	(0.001)	(0.015)	(0.008)
log_likes	0.0292***	0.3314***	0.1759***
	(0.004)	(0.048)	(0.026)
log_viewed	0.1499***	1.7013***	0.9310***
	(0.007)	(0.079)	(0.042)
log_comments_under	-0.0239***	-0.3191***	-0.1707***
	(0.004)	(0.046)	(0.025)
link	-0.0096	-0.7630***	-0.2898***
	(0.010)	(0.110)	(0.057)
log_experience	0.0087***	0.1312***	0.0717***
	(0.001)	(0.019)	(0.010)
log_contributions	0.0074***	0.0949***	0.0480***
	(0.002)	(0.027)	(0.015)
log_comments	-0.0149***	-0.1909***	-0.0983***
	(0.002)	(0.034)	(0.018)
log_artifacts	0.0327***	0.4839***	0.2606***
	(0.003)	(0.048)	(0.026)
residence_additional_info	-0.0320	-0.3298	-0.1567
	(0.021)	(0.231)	(0.116)
real_net_monetary_index	0.0067	-0.4234	-0.0978
	(0.060)	(0.911)	(0.488)
log_artifs_rate	1.8318***	17.4656***	8.8668***
	(0.033)	(0.453)	(0.204)
average_age	-0.0060	0.0060	0.0068
	(0.005)	(0.042)	(0.023)
detector_expensive_dummy	-0.0134	-0.3184	-0.1000
	(0.018)	(0.212)	(0.111)
localities_rate	0.3055	-0.4650	0.8172
	(0.279)	(3.310)	(1.803)
const	-0.8780***	-17.2255***	-9.8186***
	(0.223)	(2.359)	(1.276)
Percentage correctly predicted	87.77%	89.38%	89.01%
Log-likelihood value	-	-3207.7	-3223.2
Pseudo R-squared	0.443	0.4781	0.4755
AUC	-	0.9382	0.9372

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors are shown in parentheses.

Table 6.7: LPM, Logit, and Probit Estimates - Dataset 5

6.6 Summary of the Results

In this section, we provide an overview of all datasets' results in the form of a Table 6.8. The table columns are represented by the five datasets' names. For example, 'D1A' means Dataset 1 Artifacts, and 'D2C' means Dataset 2 Coins. The rest of the column names are defined the same way. Then the rows represent all variables included in our models. When a variable was not included in the models respective to the specific dataset, the corresponding cell contains 'na'. If the variable was not significant in all three models (LPM, Logit, Probit) at the 5% level of significance, the respective cell contains the value of '0'. If the variable had a negative coefficient and was significant at the 5% level in all three models for the respective dataset, it attains a value of '$-$'. On the other hand, if the variable's estimate was positive and significant, it attains a value of '+'. And finally, if the results were not consistent within the three models in the corresponding dataset, the cell contains the signs representing the significance and direction of the estimates of each of the three models.

The dependent variables in the models included in Table 6.8 are the dummy variables representing the submission of artifacts or coins on the individual detectorists and individual artifacts level. The dependent variables are respectively *rate_artifs_dummy* (D1A, D3A), *rate_coins_dummy* (D2C, D4C), and *submitted_to_dummy* (D5A). The key independent variables are highlighted in Table 6.8 in italics.

We can see that the estimates of the variables, in general, are relatively consistent within the datasets. Most importantly, the results of our three key variables are all consistent. The *real_net_monetary_index* is not significant in any of the datasets. The same holds for the variable *localities_rate*. Similarly, the *detector_expensive_dummy* is also not significant in the artifacts-specific datasets. However, interestingly, it is significant and positively related to the submission of coins.

Therefore, under our assumptions, it is likely that the poorer metal detectorists create collections out of coins more than their richer peers. Due to the potentially more collectible nature of coins when compared to artifacts in general, it might mean that the poorer individuals see part of their finds as a potentially valuable investment collection.

Summary of the Results					
Variables	**D1A**	**D2C**	**D3A**	**D4C**	**D5A**
log_experience	+	0	+	0	+
log_contributions	+00	+00	0	+00	+
log_comments	+	0++	0++	0++	−
log_artifacts	+	na	+	na	+
log_coins	na	+	na	+	na
real_net_monetary_index	0	0	0	0	0
log_coins_rate	+	na	+	na	na
log_artifs_rate	na	+	na	+	+
localities_rate	0	0	0	0	0
link	0	0	0	0	0−−
residence_additional_info	0++	0	0	na	0
detector_expensive_dummy	0	+	0	+	0
average_age	na	na	na	na	0
period	na	na	na	na	−
uploaded_year	na	na	na	na	+
log_likes	na	na	na	na	+
log_viewed	na	na	na	na	+
log_comments_under	na	na	na	na	−

Table 6.8: Summary of The Models' Results

6.7 Potential Drawbacks

6.7.1 Incomplete Reporting

The data used for our analysis, especially the datasets one to four, are likely subject to incomplete reporting both from above and from below since we do not observe the true submission rates. In reality, detectorists might have more finds, as well as a different proportion of finds submitted, which could be higher or lower. Moreover, each individual has a different lower/upper bound for the reported values, so there is no threshold that applies to all individuals. Instead, individual observations are incompletely reported in their unique way. For example, some individuals may have 100 artifacts but only upload 10. They submitted 1 out of the 100 artifacts, but what is observable on the website is the 1/10 submission rate, not the actual 1/100 submission rate. The same may hold for incomplete reporting 'from above'. Another individual may have 10 artifacts in reality, uploads all of them, but only submitted one without declaring it on the website. As a result, the observed submission rate is 0, but in reality, it is 1/10. Whereas the incomplete reporting 'from below' is likely addressed by using binary classification models, such as Logit and Probit, the incomplete reporting 'from above' might still appear to be a concern since there might be people that in reality submitted one of their finds, but their observable submission dummy variable attains a value of zero. Nevertheless, we assume that the proportion of those individuals is low enough not to influence the results significantly.

6.7.2 Non-random Sample

The data used for the analysis are not a random sample since an individual can tell other individuals to sign up for the website. Not only because of that might some people be more likely to register on the website. Also, the knowledge and availability of technology might influence the probability of registering on the website. For example, in the case when an individual detectorist lives without a connection to the internet, that individual is possibly more likely not to register on the website compared to a metal detectorist who has an internet connection. However, we assume those situations rarely occur so that the sample is overall representative of the metal-detecting population in the Czech Republic.

6.7.3 Time Variant Characteristics

The observations of individuals and finds that we use for the analysis are accumulated over a certain period, from 2010 to 2023. This might raise concern since some people might be active sooner, some later. In our analysis, we do not account for the time dimension, though we downloaded from the website also data that allow for panel-data analysis; and panel-data methods might address some of the issues arising in the cross-sectional analysis. However, it was primarily due to the unavailability of the panel data on the independent variables, such as the *real_net_monetary_index* variable serving as a key proxy for the economic status of an individual. Therefore, we provide just a static analysis of the cross-sectional data, whose results might, however, be influenced by the varying preferences of individuals throughout time. For this reason, we also make an assumption for the current analysis that the preferences of individuals in the metal-detecting population are predominantly time-invariant. Also, additionally, the current cross-sectional analysis might be improved, instead of using binary classification models, by using models that can handle zero-inflated data and at the same time are applicable to non-discrete dependent variables.

Chapter 7

Conclusion

The contribution of this thesis lies not only in introducing the topic of the metal detecting hobby as part of the social sciences and providing initial insights but also in providing the data obtained. This data might have a high potential for further analyses, not only within the metal detecting hobby, or the field of numismatics, but could also be used, for example, to estimate the economic activity and trade in different regions of the Czech Republic in the past.

In our analysis, the key hypothesis is that individuals with a higher socioeconomic status submit finds to the museums more and therefore are less likely to create private collections out of their finds. The underlying rationale was that if one is richer, one has less need to collect the finds found while metal detecting, therefore having more preferences towards submitting the artifacts, showing the higher hobbyist motivation for metal detecting. On the other hand, if one has a lower socioeconomic status, one might think that collecting finds might help improve their socioeconomic status; either hoping to sell the collection in the future, or simply by 'consuming' the collection, which might be the case when one simply has the collection displayed on a shelf.

The main hypothesis was tested using two different independent variables at the same time. This divided the key hypothesis, under certain assumptions, into the following two. First (**H1**) was, that if one lives in an area with a higher real net monetary index, which serves as a proxy for an individual's socioeconomic status, that individual is more likely to submit a find to the archaeological authority. And second (**H2**), when one owns a more expensive metal detector, one is more likely to submit a find, either artifact or coin. Here the value of the metal detector serves as a proxy for the socioeconomic status of an individual. Hence, the above two hypotheses attempted to verify if there are varying preferences for finds submission within different socioeconomic groups. And finally, the third hypothesis attempted to test if, in the presence of a higher chance of a valuable find, there is a higher probability of submitting the find. This third hypothesis was thus formulated as if one lives in an area with a higher density of archaeological localities, one is more likely to submit a find (**H3**). This hypothesis might potentially support the supposed overall prevailing hobbyist motivation for metal detecting.

Next, we collected the data from the renowned Czech metal detecting website, using the technique of web scraping, cleaned the obtained data, and moved on to the analysis, testing primarily the above-mentioned hypotheses. The first hypothesis (**H1**) proved not significant at the 5% significance level in any of the fifteen main models used. Therefore, overall, we reject the **H1**, that an individual living in an area with a higher real net monetary index (higher purchasing power) is more likely to submit a find, either artifact or coin. Then, the second hypothesis (**H2**) was rejected in all the models focused on estimating the submission of artifacts. There was only one exception, which occurred in the first dataset, suggesting a potential joint significance of the density of localities in an area one lives in and a higher value of the metal detector one uses. This might mean that if those two, the higher density of localities and a higher price of the metal detector, occur together, there is a higher chance that this individual submits an artifact. This might potentially support our main hypothesis, under the declared assumptions, that an individual with a higher socioeconomic status is more likely to submit an artifact, being in an area with a higher density of archaeological localities, compared to the rest of the metal detectorists. Nevertheless, this hypothesis was not significant in either of the two remaining control datasets. On the other hand, the key **H2** proved significant at the 5% level of significance in all six models based on the datasets respective to the submission of coins. Not rejecting the **H2** in this case potentially means that if one owns an expensive metal detector, one is more likely to submit a coin. Under the assumption that ownership of the more expensive metal detector is associated with a higher socioeconomic status, it might mean that if one is richer, one is more likely to submit a coin. This might be an interesting conclusion when compared to the results of the models focusing on artifacts, which did not indicate the significance of this hypothesis. Therefore, we might conclude overall that if one owns an 'expensive' metal detector, one is more likely to submit a coin but not an artifact. This outcome might be the most trustworthy out of the all outcomes of this analysis, since, the detectors used by the metal detectorists in our datasets are not an approximation as in the case of localities rate and real net monetary index but the real metal detectors that the individuals declared to use.

It might be reasonable to mention that both proxy variables for the socioeconomic status of an individual were estimated alongside each other in all the models. Although both are proxies for the same variable, each of them absorbs different effects. Whereas the higher *real_net_monetary_index* represents a relatively more affluent environment, the *detector_expensive_dummy* represents rather an individual socioeconomic status. With the former indicating rather an environment an individual lives in, one with a higher real net monetary index might potentially have a higher awareness of the public cultural goods. On the other hand, an individual owning an 'expensive' metal detector might have a higher socioeconomic status on a personal level. This might mean, as opposed to the more wealthy environment, that the person might have a relatively lower awareness of public cultural goods. Interestingly, our estimates are positive and significant for a proxy of personal wealth, and not for the proxy of the socioeconomic environment. This observation might imply that the higher likelihood of finds submission is positively related to personal wealth rather than the awareness about cultural goods (socioeconomic environment). This supports our hypothesis further, suggesting that individual wealth plays a significant role in making the collections of finds.

Next, except for the joint significance together with the expensive detector dummy variable in the first dataset, which was not further supported by the evidence from the other two datasets, the **H3** was, similar to the **H1**, rejected in all main fifteen models used. Hence, we do not have enough evidence that an increased density of archaeological localities in an area one lives in is associated with an increased likelihood of submitting an artifact. Here, we cannot be sure that a localities rate is a good proxy variable for an actual density of potential archaeological finds. Therefore, based on our analysis, we can reject the **H3**, that overall, there is a prevailing hobbyist motivation among metal detector users, only assuming, that the localities rate is a good approximation of the density of archaeological finds in an area.

Moreover, since our hypotheses are unique and have likely never been tested before, we cannot directly address any literature, that is, put our results in line with other studies that we are aware of. Nevertheless, indirectly addressing the outcomes of the survey of metal detectorists in the Czech Republic conducted by Komoróczy (2022), we come to similar conclusions. First, the coins are less likely to be submitted to the archaeological institutions than artifacts, based on the proportions of respective submitted and unsubmitted finds (about 8.5% for artifacts and 2.7% for coins). Furthermore, our models show a significant differentiation of preferences among users for submitting or not submitting coins, which are a relatively uniform group of finds, as opposed to artifacts. Based on that, coins are potentially more suitable than artifacts for measuring preferences for ownership or non-ownership of finds. The varying preferences for finds submission are also observed by Komoróczy (2022), either in the case of not an insignificant number of metal detectorists declaring that they are motivated by the vision of enriching their collections, or by declaring that the financial reward for handed-over artifacts should be paid. Nevertheless, our analysis goes beyond this and attempts to match the socioeconomic characteristics of those individuals to their respective preferences. The results might serve as a hint for potential policies and actions taken with respect to the metal detecting hobby. One of the examples might be the identification of prospective metal detectorists for collaboration with archaeological institutions. One of the other examples might be the potential taxation policies addressing the metal detecting hobby.

To sum up, we found out in our analysis, under the mentioned assumptions, first, that the preferences for submission particularly of artifacts do not differ significantly among potentially different socioeconomic groups of metal detectorists, nor do they differ when living in an area with a possibly higher density of archaeological finds. That might mean that all the metal detectorists, regardless of the socioeconomic group they belong to, may have a rather uniform motivation for submitting or not submitting and the potential creation of artifacts collections either for investment or consumption purposes. On the other hand, in the case of coins, under the declared assumptions, the preferences for submitting and not submitting coins likely differ, provided that the detector an individual detectorist uses may be a more accurate measure of one's wealth than the real net monetary index in an area an individual resides in. In other words, assuming the value of the detector that one uses is the most accurate measure of one's wealth among our independent variables, we conclude that potentially richer individuals tend not to create collections of coins as much as relatively poorer individuals do. Hence, the motivation of richer individuals with respect to finding coins is likely more hobbyist than the motivation of poorer individuals, who appear to create collections of coins more, either for consumption or investment purposes. Overall, we did find evidence that the preferences of individuals with different prices of metal detectors for submission of finds do not differ in the case of artifacts but are significantly different in the case of coins. Since, unlike artifacts, coins are items that form a relatively homogenous group, the latter observation of lower preferences of richer individuals for the ownership of finds, in general, is considered to be the most reliable result of our analysis. Therefore, we indeed found out that relatively poorer agents participating in the contest for valuable finds are likely motivated by the potential value of the finds. Moreover, for the relatively poorer, the finds are potentially more likely to serve the purpose of creating collections.

Bibliography

ADDYMAN, P. V. (2009): "Before the Portable Antiquities Scheme." *Thomas, S., Stone, P. G. (Eds.), Metal detecting and archaeology* pp. 51–62.

AUSTIN, T. (2005): "Building bridges between metal detectorists and archaeologists." *Thomas, S., Stone, P. G. (Eds.), Metal detecting and archaeology* pp. 119–124.

VANDEN BROUCKE, S. & B. BAESENS (2018): *Practical Web Scraping for Data Science: Best Practices and Examples with Python.* Berkeley, CA, United States: Apress Berkeley.

CODIGNOLA, F. & P. MARIANI (2022): "Investigating preferences in art collecting: the case of the François Pinault Collection." *Italian Journal of Marketing* p. 107–133.

CORNELISON, J. E. & G. S. SMITH (2009): "Archaeology, Metal Detecting and the Development of Battlefield Archaeology in the United States." *Thomas, S., Stone, P. G. (Eds.), Metal detecting and archaeology* pp. 33–50.

FAWCETT, T. (2006): "An introduction to ROC analysis." *Pattern Recognition Letters* **27(8)**: pp. 861–874.

FU, Q. & Z. WU (2019): "Contests: Theory and Topics." *Oxford Research Encyclopedia of Economics and Finance* .

GLYN, D. (1967): *The origins and growth of archaeology.* Harmondsworth, Middlesex, England: Penguin Books Ltd.

HAJŠMAN, J., M. ŘEZÁČ, P. SOKOL, & R. TRNKA (2019): *Příručka amatérského archeologa, aneb, Do mrtvých se nekope (In English: Handbook of the Amateur Archaeologist, Alias, Do Not Dig Into the Dead).* Prague: Libri.

HARDY, S. A. (2017): "Quantitative analysis of open-source data on metal detecting for cultural property: Estimation of the scale and intensity of metal detecting and the quantity of metal-detected cultural goods." *Cogent Social Sciences* **3(1)**.

HAVLOVICOVÁ, A. (2020): "Evaluation of Contemporary art as an alternative investment." *Bachelor thesis*, Charles University, Faculty of Social Sciences, Institute of Economic Studies.

HODDER, I. (1984): "Archaeology in 1984." *Antiquity* **58**: pp. 25–32.

HUTH, C. (2013): "Vom rechten Umgang mit Sondengängern: Das 'Portable Antiquities Scheme' in England und Wales und seine Folgen." *Archäologische Informationen* **36**.

IACUS, S. M. (2015): "Book Review: JSS Journal of Statistical Software." *Journal of Statistical Software* **68(3)**.

KOBYLINSKY, Z. & P. SZPANOWSKI (2009): "Metal detector users in Poland: The Current State of Affairs." *Thomas, S., Stone, P. G. (Eds.), Metal detecting and archaeology* pp. 20–21.

KOCOUREK, A., J. ŠIMANOVÁ, & J. ŠMÍDA (2021): "Money Income and the Cost of Living of the Population: A Detailed View of the Czech Republic." *Technical report*, Technical University of Liberec.

KOMORÓCZY, B. (2022): "Archaeology, Metal Detecting, and Citizen Science in the Czech Republic." *Advances in archaeological practice: a journal of the Society of American archeaology* **10(3)**: pp. 322–335.

KOSKINEN-KOIVISTO, E. & S. THOMAS (2017): "Lapland's Dark Heritage: Responses to the Legacy of World War II." *Heritage in Action* pp. 121–133.

KRÁSNÝ, F. (2014): "Problematika detektorů kovů v archeologii (In English: 'The Problem of Metal Detecting in Archaeology'." *Diploma thesis*, Charles University, Faculty of Arts.

VAN DER LANDE, J. (2021): "The Antiquities Trade: A reflection on the past 25 years." *Cultural Property News* .

Maaranen, P. (2016): "Metal Detecting and Archaeology in Finland: An Overview of the Hobby and its Consequences." *New Sites, New Methods. Proceedings of the Finnish-Russian Archaeological Symposium, Helsinki, 19–21 November, 2014* **21**.

Maines, R. (2009): *Hedonizing Technologies: Paths to Pleasure in Hobbies and Leisure*. Baltimore: Johns Hopkins University Press.

Mason, R. O. (1986): "Four Ethical Issues of the Information Age." *MIS Quarterly* **10(1)**: pp. 5–12.

Mazilu (2022): "Web Scraping and Ethics in Automated Data Collection." *Education, Research and Business Technologies: Proceedings of 20th International Conference on Informatics in Economy* p. 292.

Moilanen, U. (2023): "The Role of Experiences in Valuing Metal-Detecting Finds among Finnish Hobbyists." *Public Archaeology* .

Moltaš, Z. (2007): *Detektory kovů prakticky aneb zapni a hledej (In English: Metal Detectors Practically alias Turn It On and Search)*. Prague: BEN - technická literatura.

Navrátil, A. (2015): "Česká archeologie a čtvrt století užívání detektorů kovů (In English: 'Czech Archeology and a Quarter Century of Metal Detector Use'. Research Review 56)." *Přehled výzkumů* **56(1)**: pp. 119–130.

Ngo, V. L. (2013): "The theory of contests: A unified model and review of the literature." *European Journal of Political Economy* **32**: pp. 161–181.

Ofiu, K. (2013): *Boost Your Earnings from Home*. Melbourne: John Wiley Sons Australia.

Pitblado, B. (2014): "An argument for ethical, proactive, archaeologist-artifact collector collaboration." *American Antiquity* **79(3)**: pp. 385–400.

Redesdale, R. (2008): "Foreword." *Thomas, S., Stone, P. G. (Eds.), Metal detecting and archaeology* pp. 1–3.

Reeves, M. (2015): "Sleeping with the 'Enemy': Metal Detecting Hobbyists and Archaeologists." *Advances in Archaeological Practice* **3(3)**: pp. 263–274.

Richards, J. D. & J. Naylor (2009): "The Real Value of Buried Treasure. VASLE: The Viking and Anglo-Saxon Landscape and Economy Project." *Thomas, S., Stone, P. G. (Eds.), Metal detecting and archaeology* pp. 167–179.

Roberts, R. T. (1999): "Review of The History of Metal Detectors." *Geotech* pp. 1–5.

Simmel, G. (2004): *The Philosophy of Money*. London: Routledge, 3rd edition.

Spencer, P. D. (2009): "The Construction of Histories: Numismatics and Metal Detecting." *Thomas, S., Stone, P. G. (Eds.), Metal detecting and archaeology* pp. 125–136.

Taylor, B. (1995): "Amateurs, Professionals and the Knowledge of Archaeology." *The British Journal of Sociology* **46(3)**: p. 499–508.

Thomas, S. (2009): "Metal detecting and archaeology." *Thomas, S., Stone, P. G. (Eds.), Metal detecting and archaeology* pp. 1–3.

Thomas, S. (2016): "The Future of Studying Hobbyist Metal Detecting in Europe: A Call for a Transnational Approach." *Open Archaeology* **2(1)**: pp. 140–149.

Thompson, E. L. (2016): *Possession: The curious history of private collectors from antiquity to the present*. Yale University Press.

Tite, M. S. (1972): "Methods of physical examination in archaeology." pp. 32–33.

Winkley, F. (2016): "Talking to Metal Detectorists in the Field: A Methodology for Analysing Motivations and Attitudes to Landscape." *Public archaeology* **15(4)**: pp. 186–213.

Wooldridge, J. M. (2012): *Introductory Econometrics: A Modern Approach, Fifth Edition*. South-Western 5191 Natorp Boulevard Mason, OH 45040 USA: Cengage Learning.

Appendix A

Additional Figures

A.1 Dataset 1

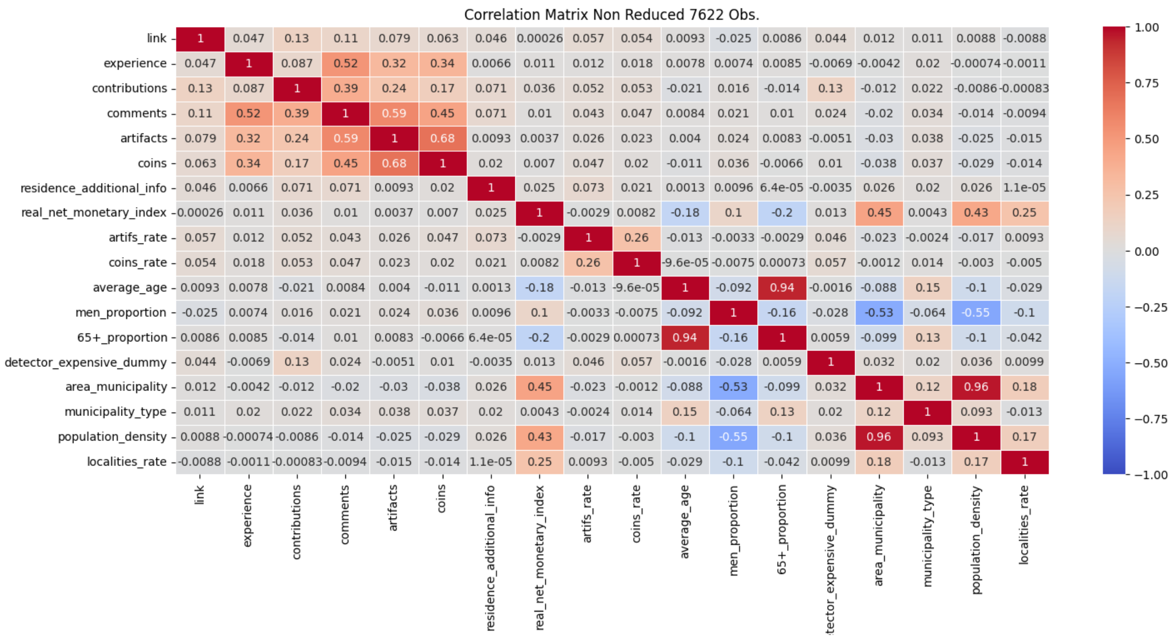

Figure A.1: Correlation Matrix - Dataset 1

Predicted	**0**	**1**
Actual 0	7046	9
Actual 1	536	28

Table A.1: Confusion Matrix - LPM - Dataset 1

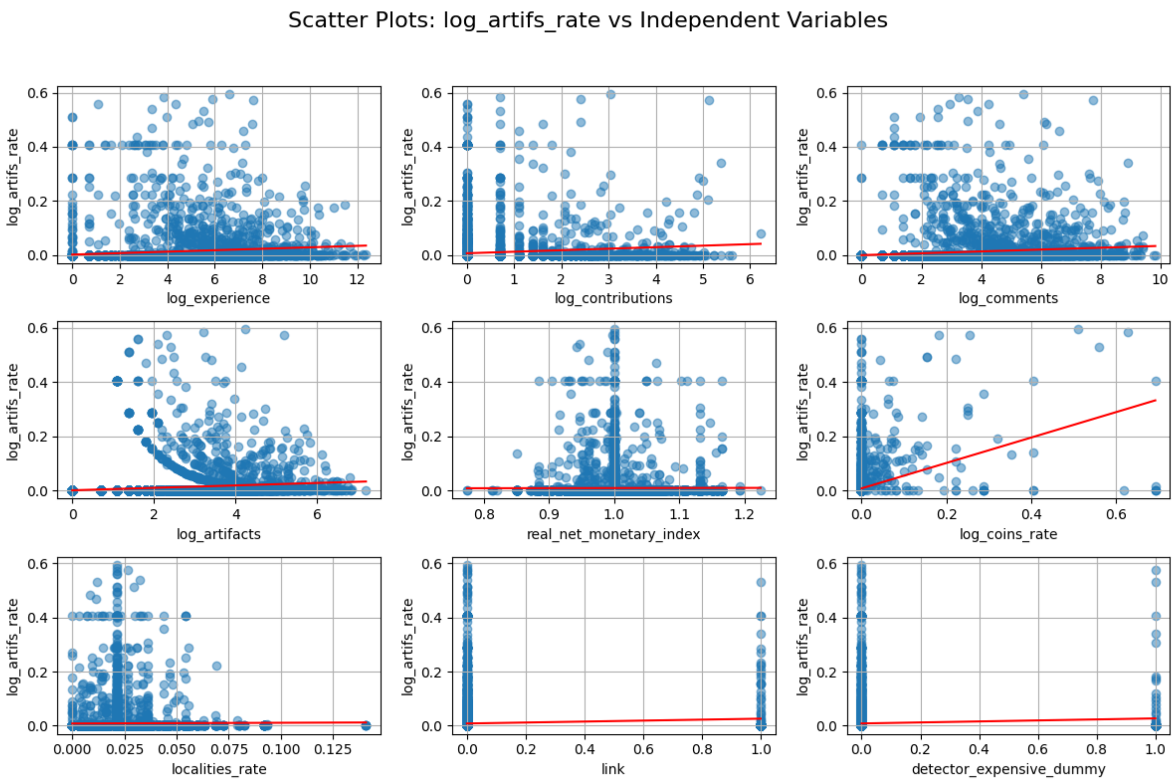

Figure A.2: Log-transformed Scatter Plots with deleted outliers - Dataset 1

Predicted	**0**	**1**
Actual 0	6997	58
Actual 1	456	108

Table A.2: Confusion Matrix - Logit - Dataset 1

Predicted	**0**	**1**
Actual 0	7008	47
Actual 1	464	100

Table A.3: Confusion Matrix - Probit - Dataset 1

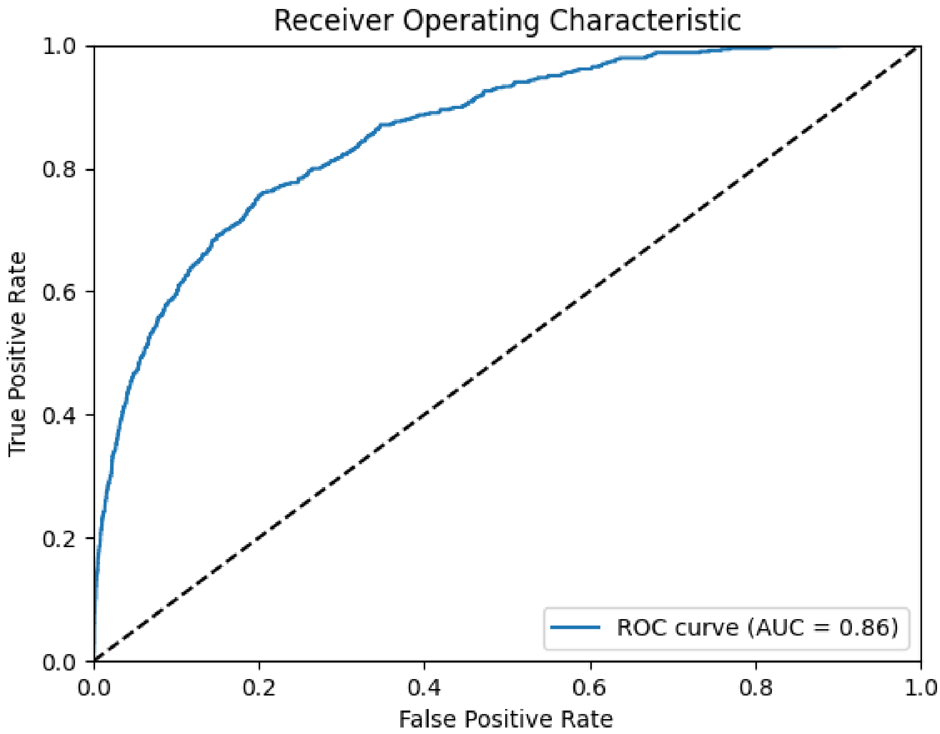

Figure A.3: Receiver Operating Characteristic - Logit - Dataset 1

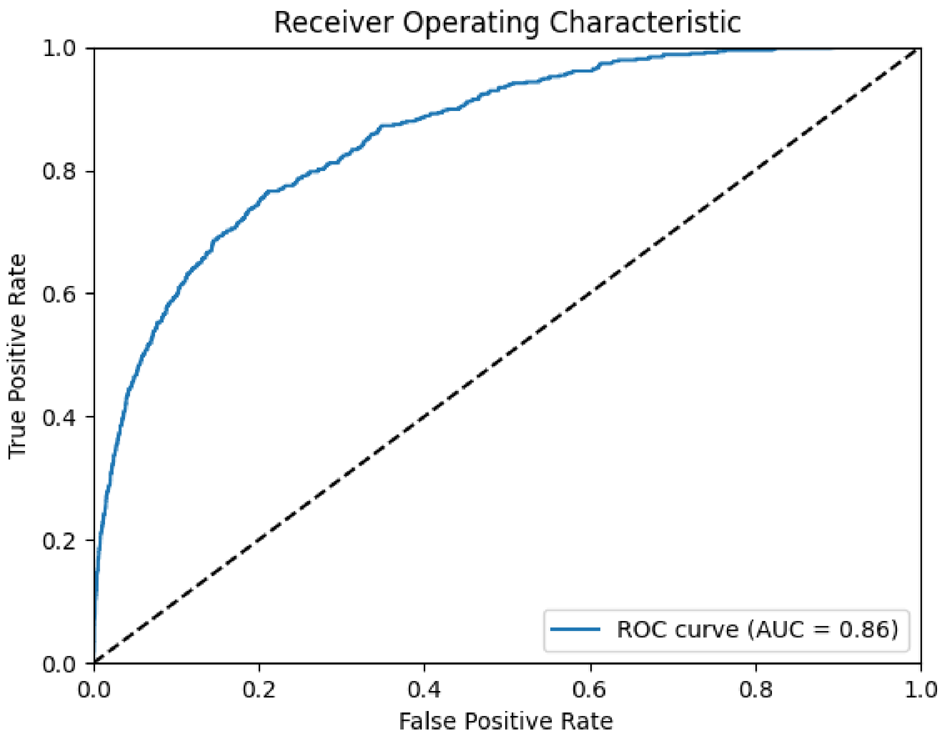

Figure A.4: Receiver Operating Characteristic - Probit - Dataset 1

A.2 Dataset 2

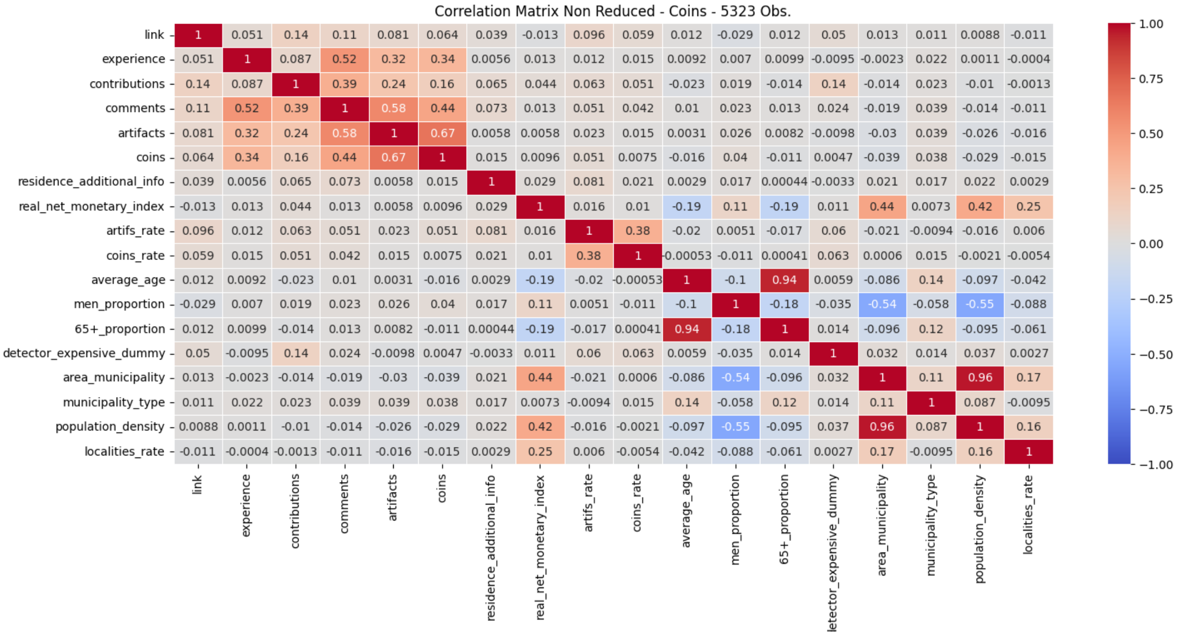

Figure A.5: Correlation Matrix - Dataset 2

Predicted	**0**	**1**
Actual 0	5169	6
Actual 1	130	14

Table A.4: Confusion Matrix - LPM - Dataset 2

Predicted	**0**	**1**
Actual 0	5163	12
Actual 1	117	27

Table A.5: Confusion Matrix - Logit - Dataset 2

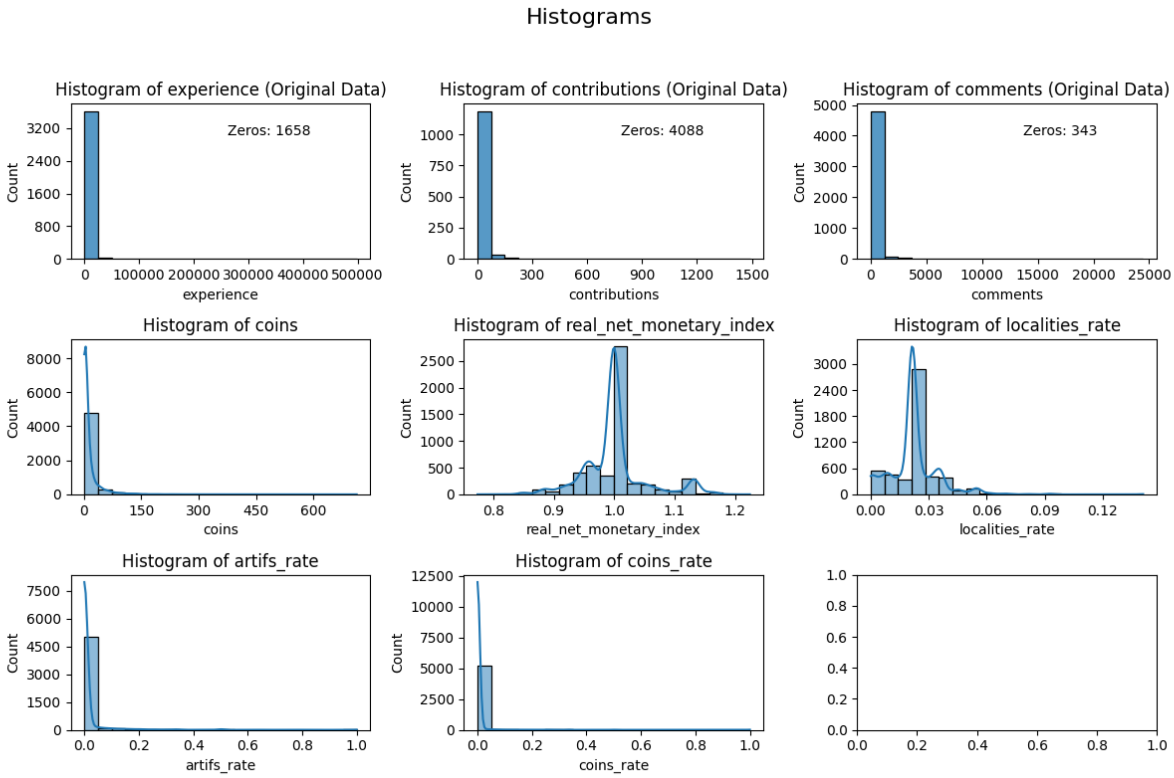

Figure A.6: Histograms - Dataset 2

Predicted	**0**	**1**
Actual 0	5166	9
Actual 1	121	23

Table A.6: Confusion Matrix - Probit - Dataset 2

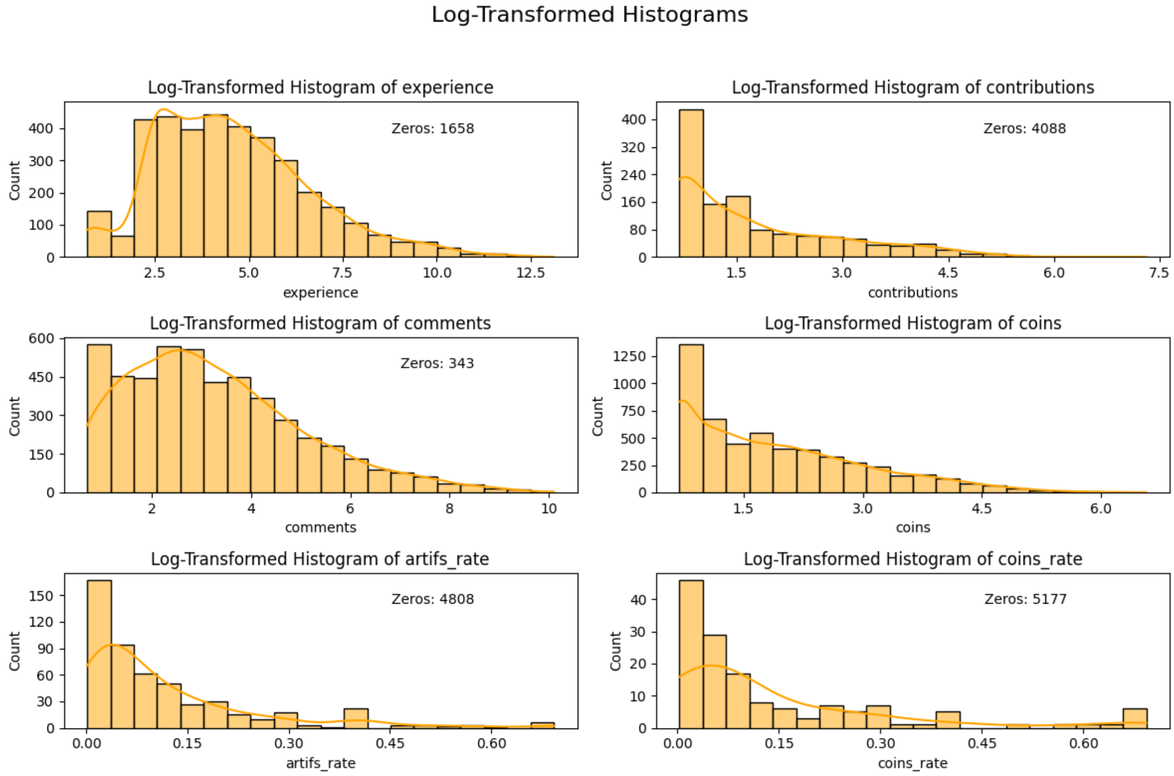

Figure A.7: Log-transformed Histograms - Dataset 2

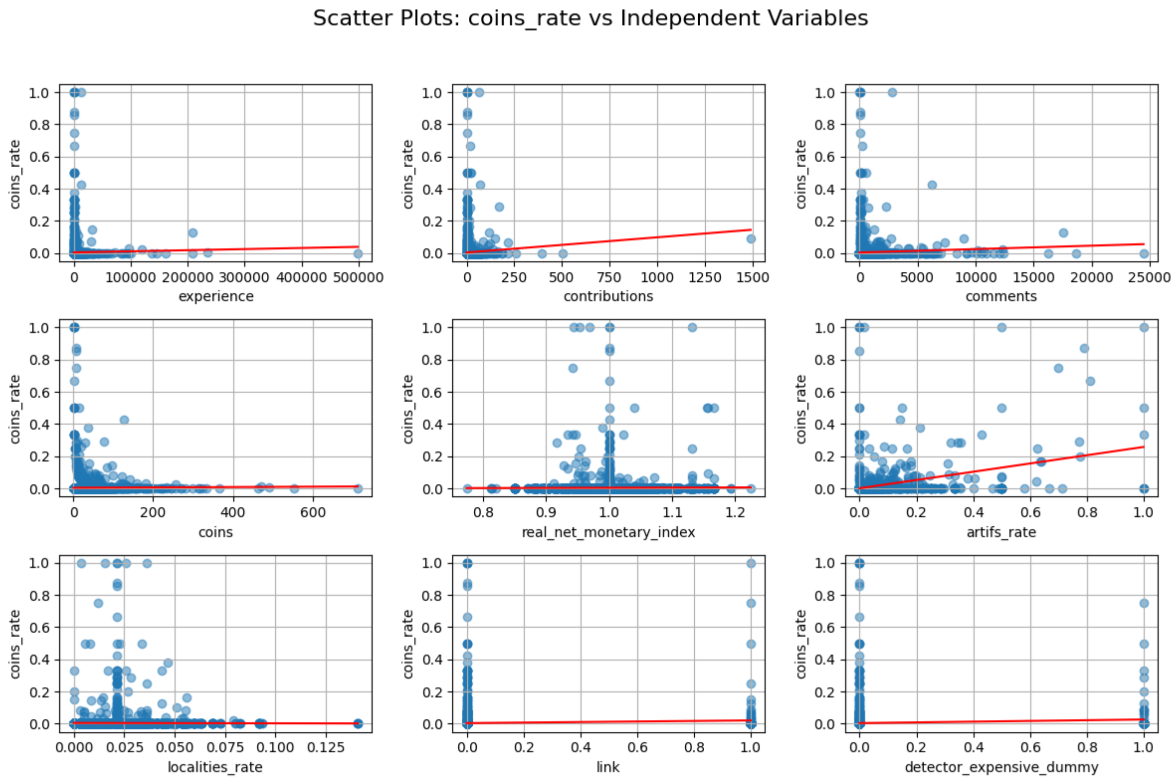

Figure A.8: Scatter Plots - Dataset 2

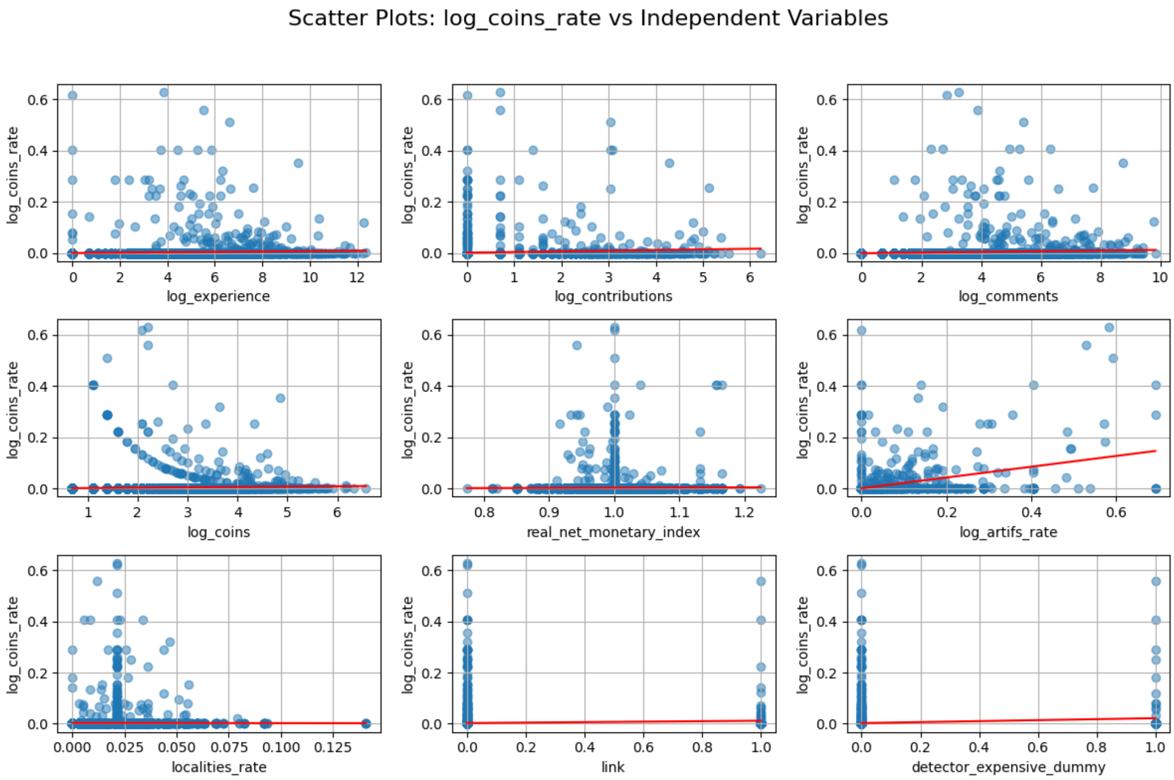

Figure A.9: Log-transformed Scatter Plots with deleted outliers - Dataset 2

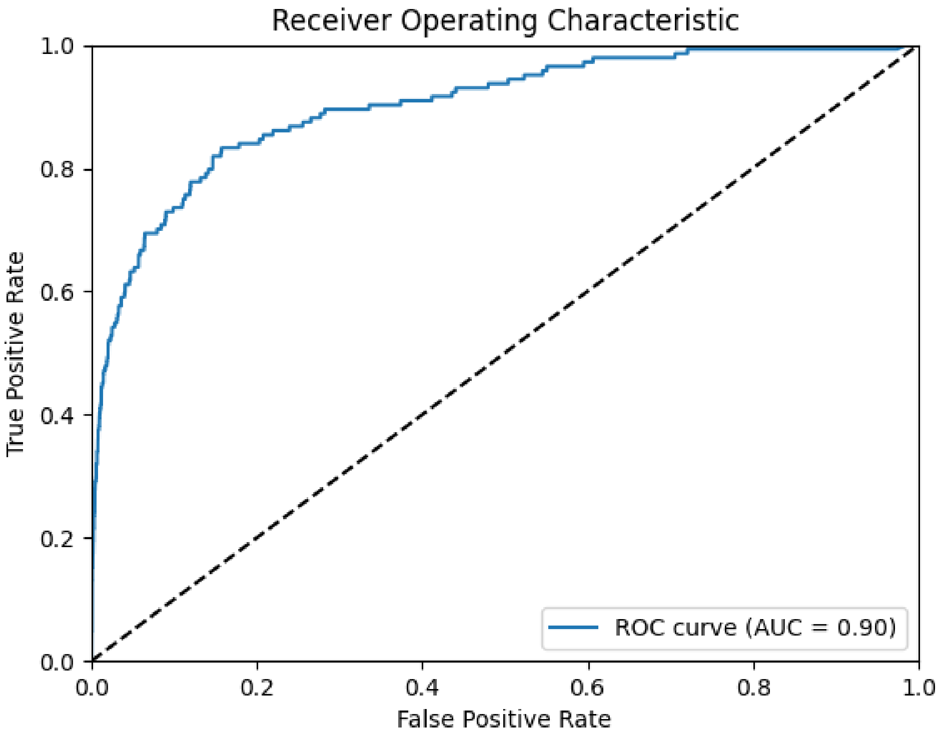

Figure A.10: Receiver Operating Characteristic - Logit - Dataset 2

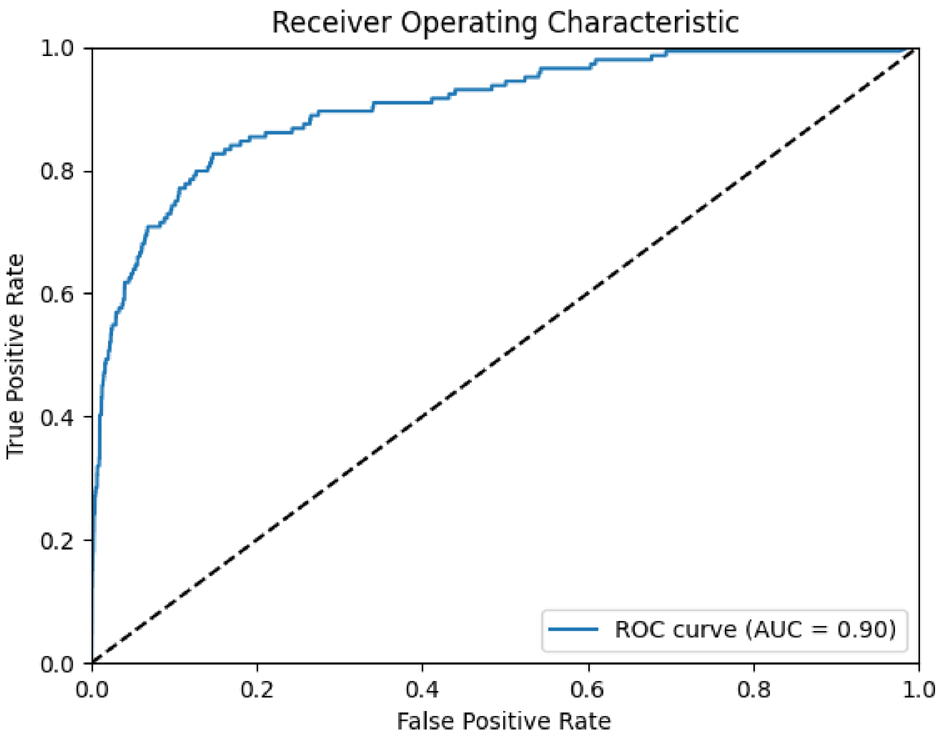

Figure A.11: Receiver Operating Characteristic - Probit - Dataset 2

A.3 Dataset 3

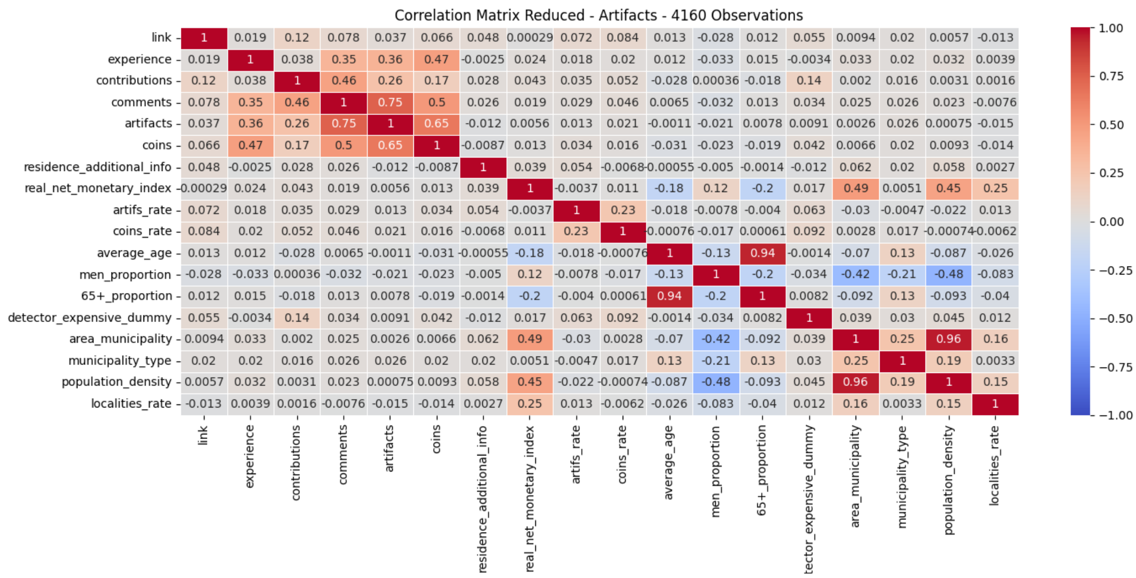

Figure A.12: Correlation Matrix - Dataset 3

Predicted	**0**	**1**
Actual 0	3900	3
Actual 1	243	11

Table A.7: Confusion Matrix - LPM - Dataset 3

Predicted	**0**	**1**
Actual 0	3886	17
Actual 1	217	37

Table A.8: Confusion Matrix - Logit - Dataset 3

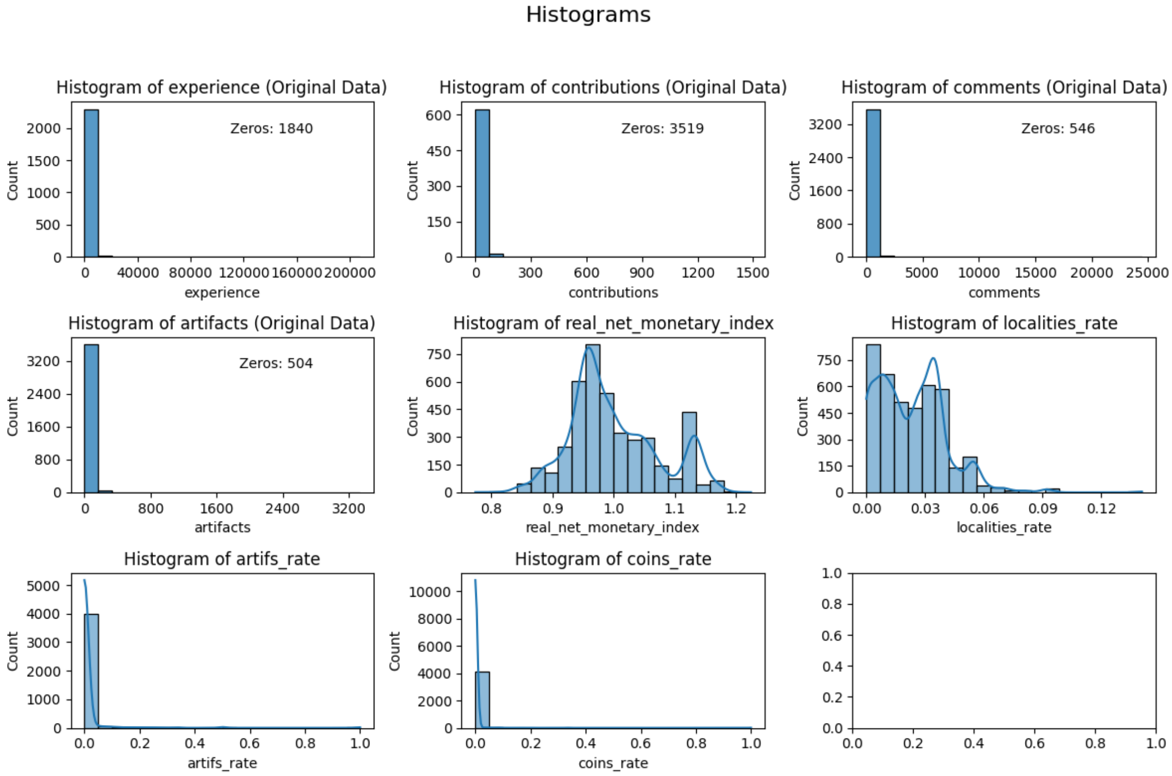

Figure A.13: Histograms - Dataset 3

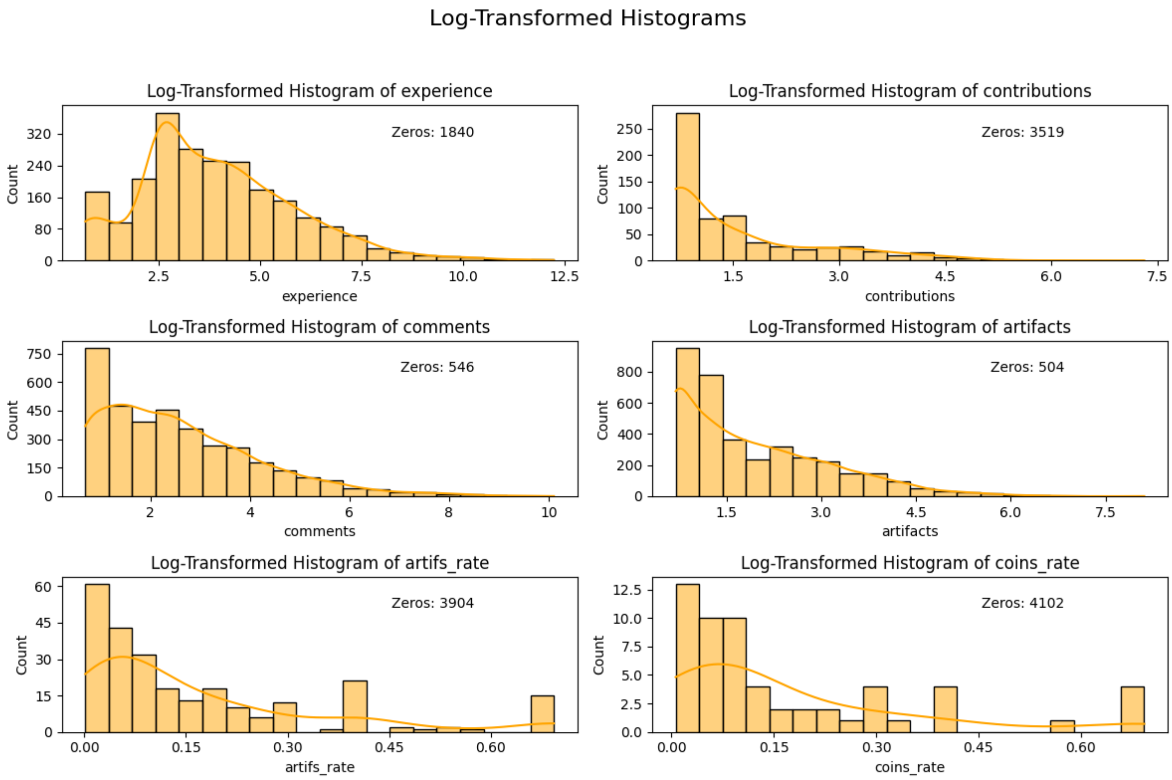

Figure A.14: Log-transformed Histograms - Dataset 3

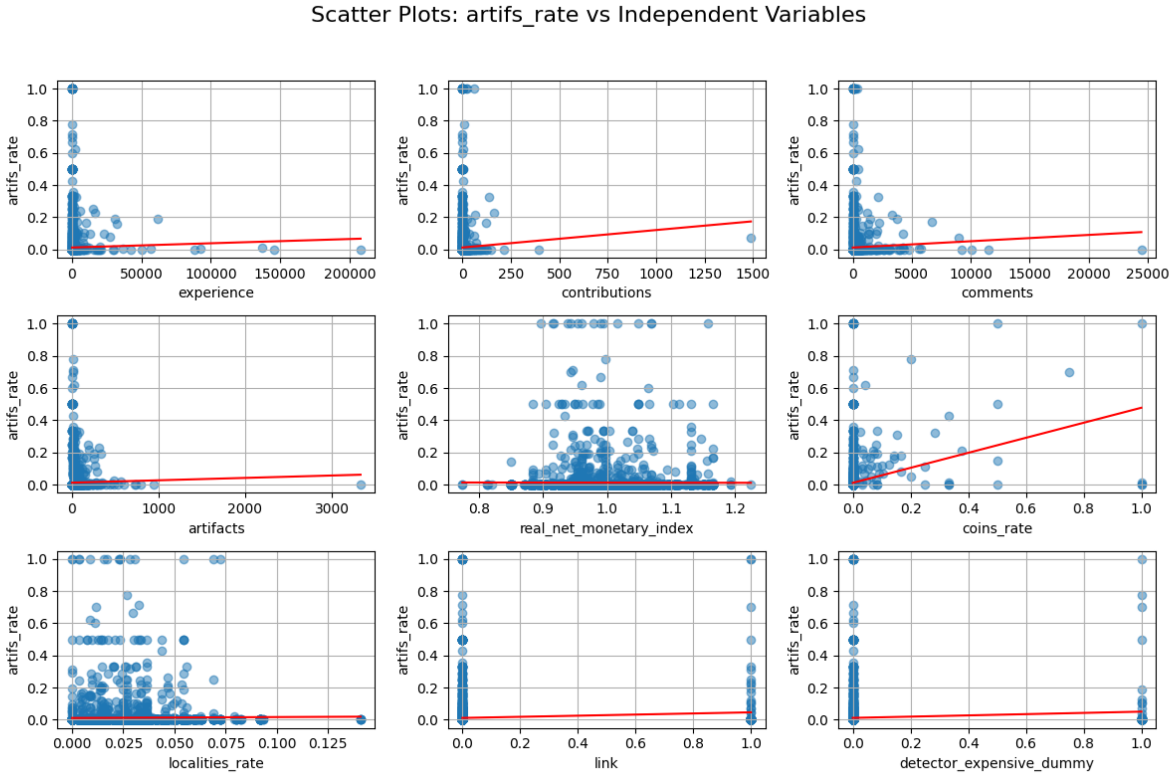

Figure A.15: Scatter Plots - Dataset 3

Predicted	0	1
Actual 0	3893	10
Actual 1	218	36

Table A.9: Confusion Matrix - Probit - Dataset 3

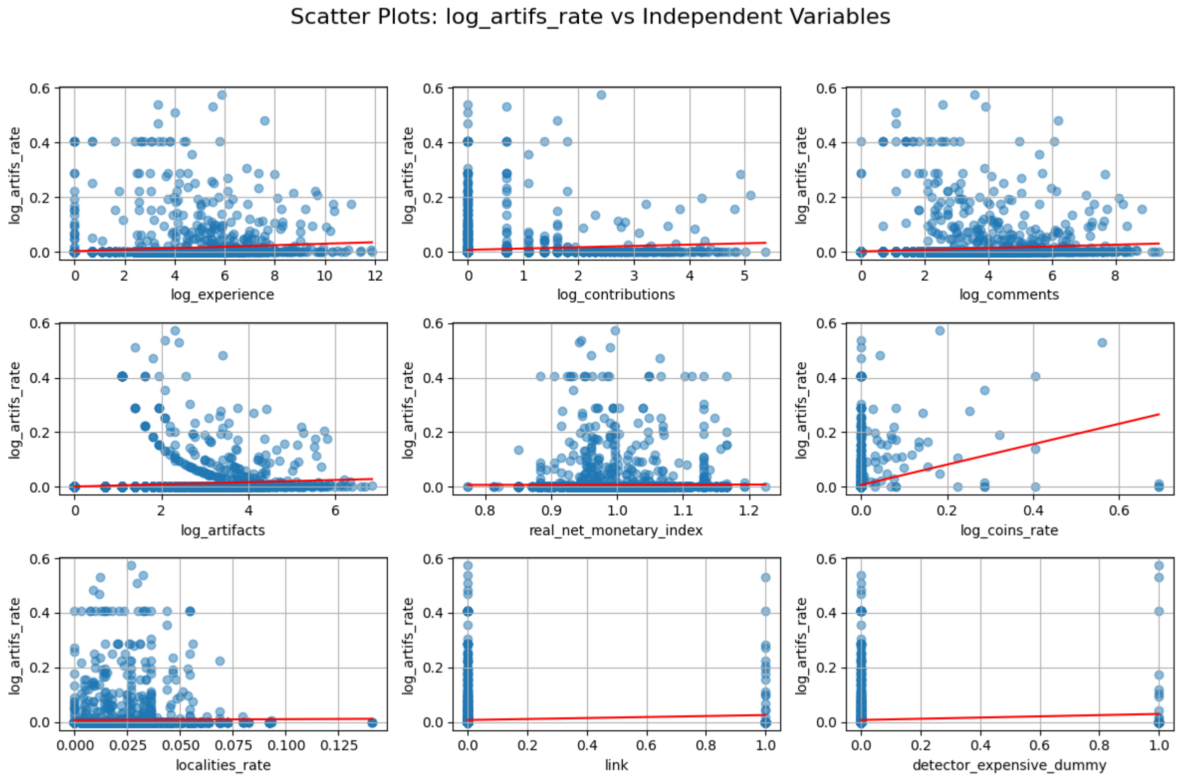

Figure A.16: Log-transformed Scatter Plots with deleted outliers - Dataset 3

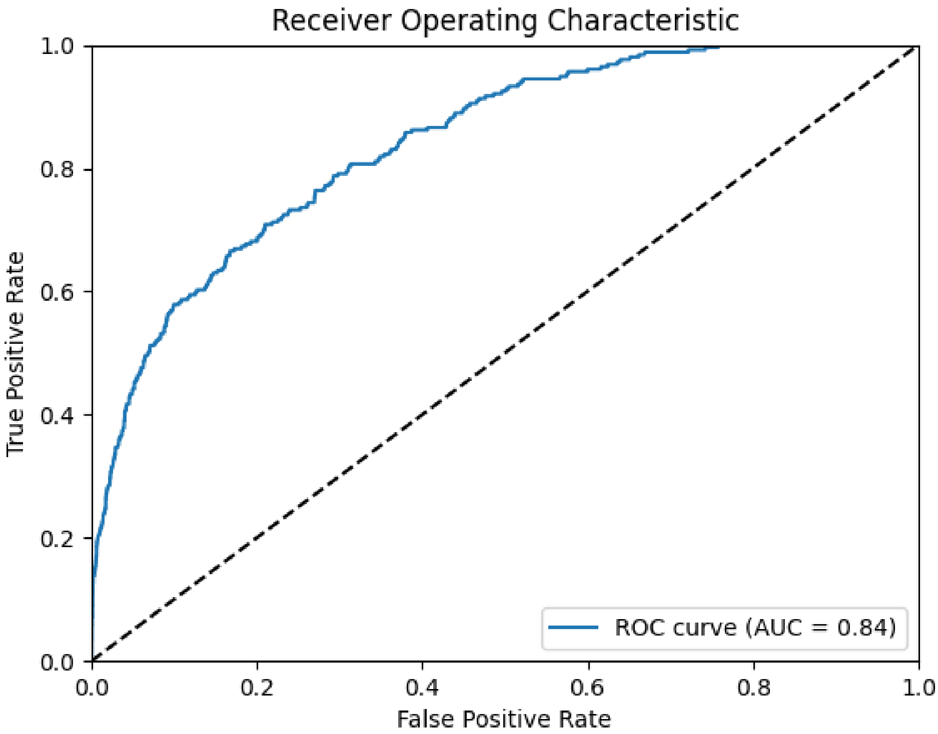

Figure A.17: Receiver Operating Characteristic - Logit - Dataset 3

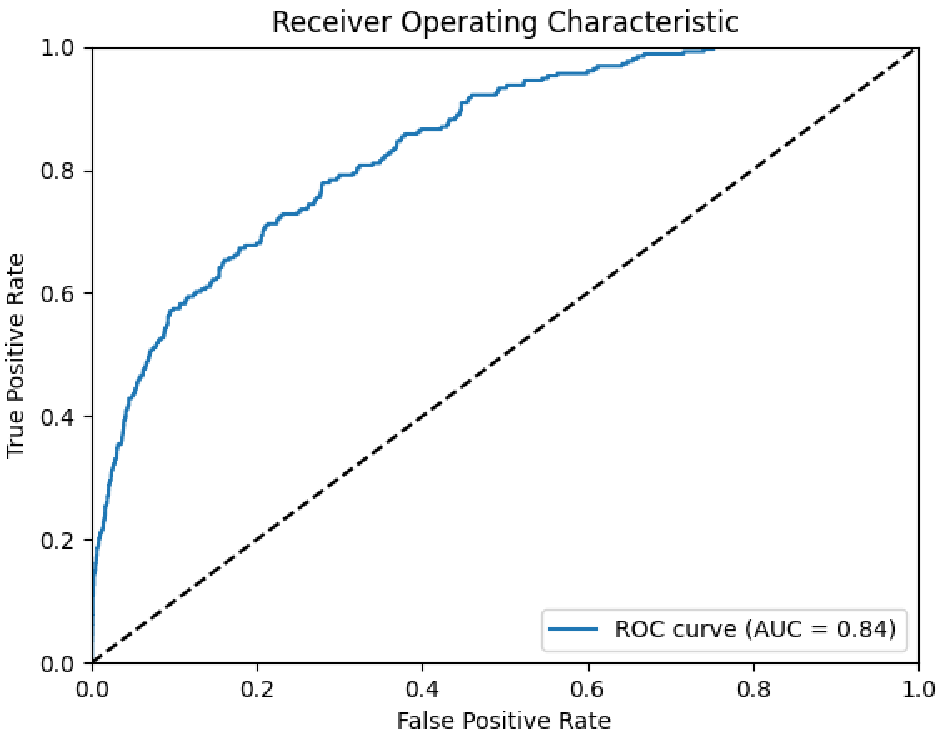

Figure A.18: Receiver Operating Characteristic - Probit - Dataset 3

A.4 Dataset 4

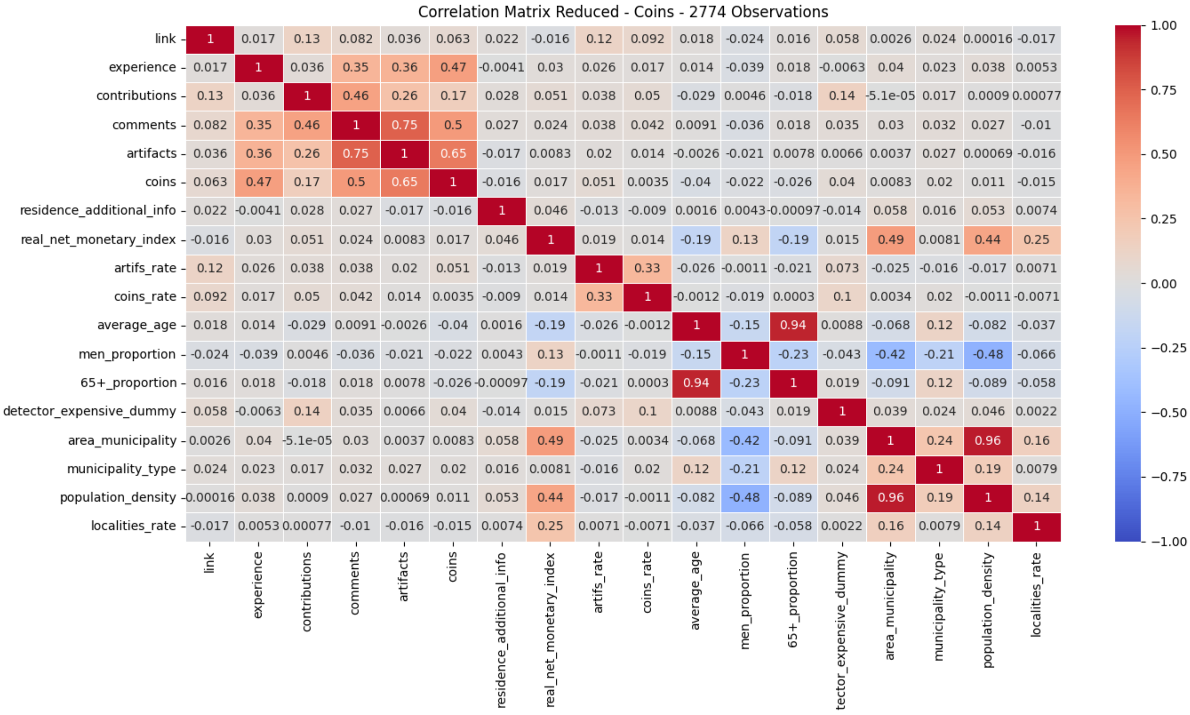

Figure A.19: Correlation Matrix - Dataset 4

Predicted	**0**	**1**
Actual 0	2712	2
Actual 1	51	6

Table A.10: Confusion Matrix - LPM - Dataset 4

Predicted	**0**	**1**
Actual 0	2709	5
Actual 1	45	12

Table A.11: Confusion Matrix - Logit - Dataset 4

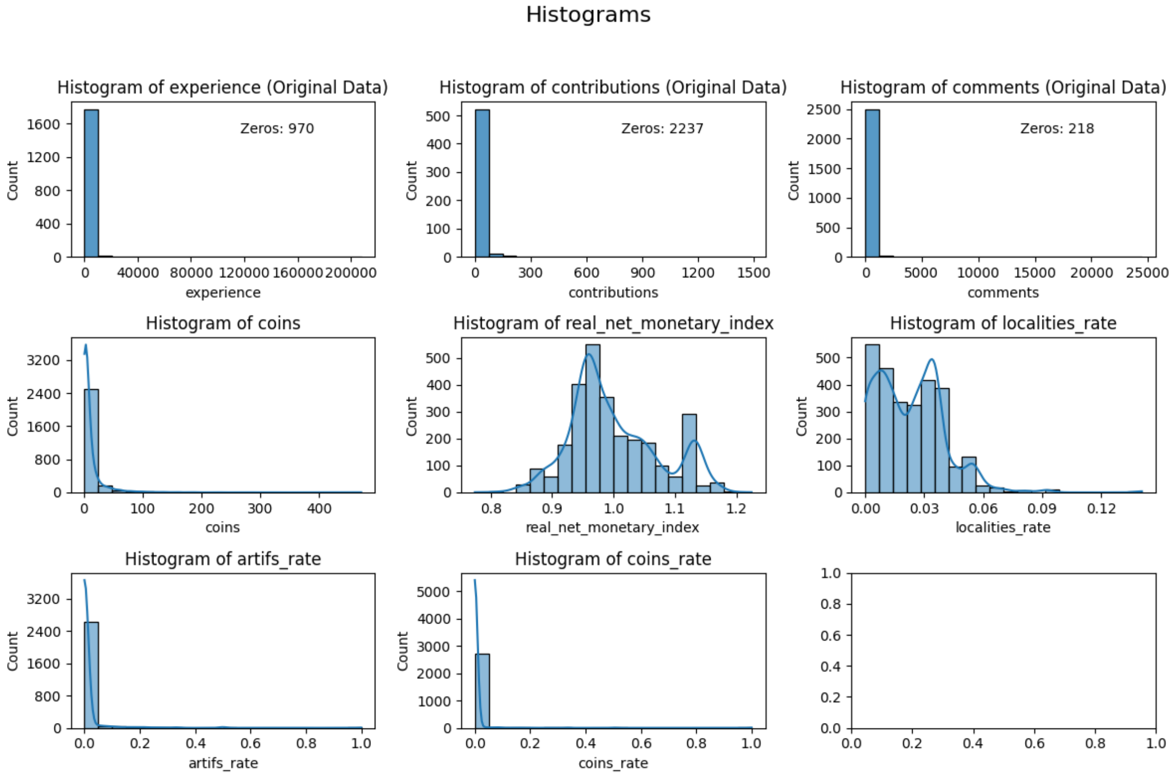

Figure A.20: Histograms - Dataset 4

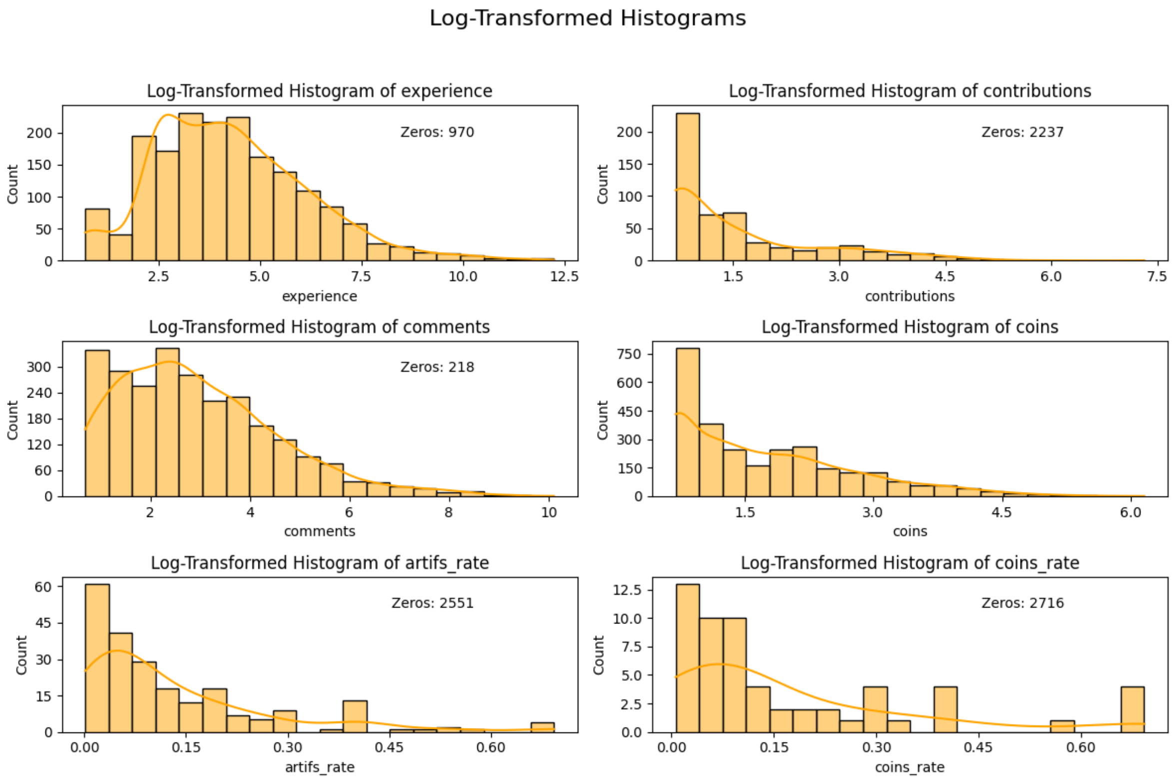

Figure A.21: Log-transformed Histograms - Dataset 4

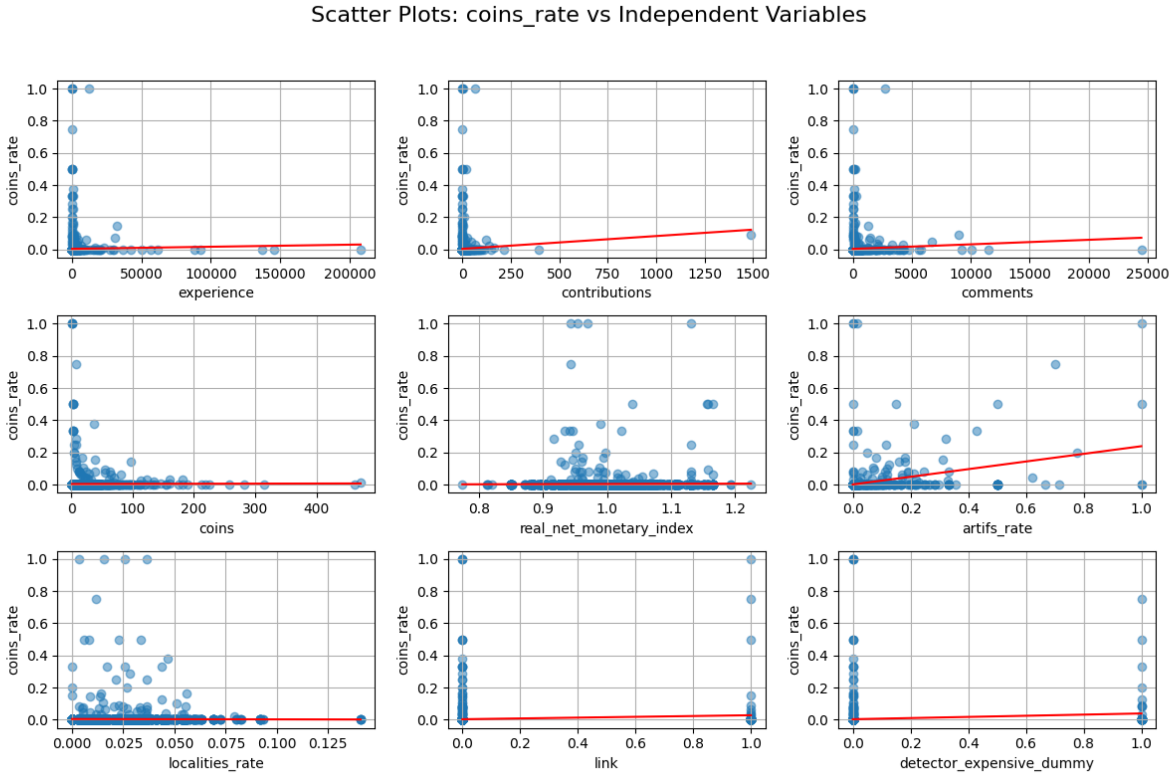

Figure A.22: Scatter Plots - Dataset 4

Predicted	0	1
Actual 0	2711	3
Actual 1	48	9

Table A.12: Confusion Matrix - Probit - Dataset 4

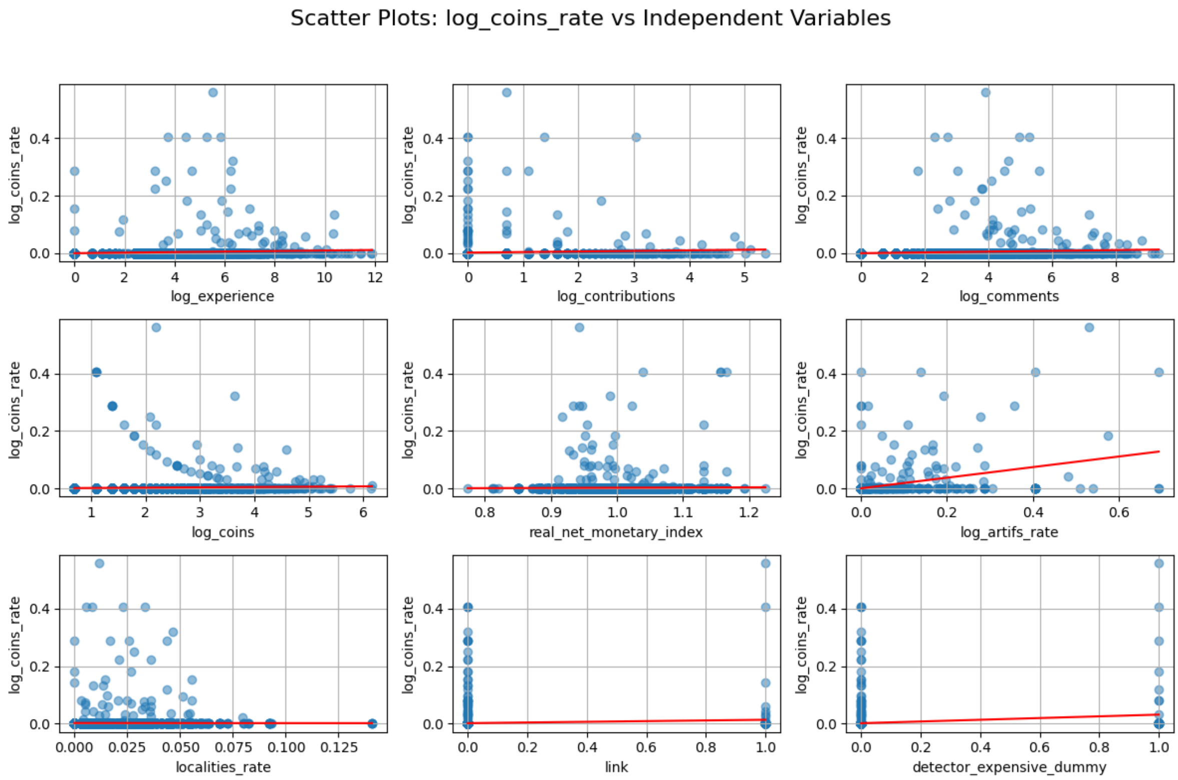

Figure A.23: Log-transformed Scatter Plots with deleted outliers - Dataset 4

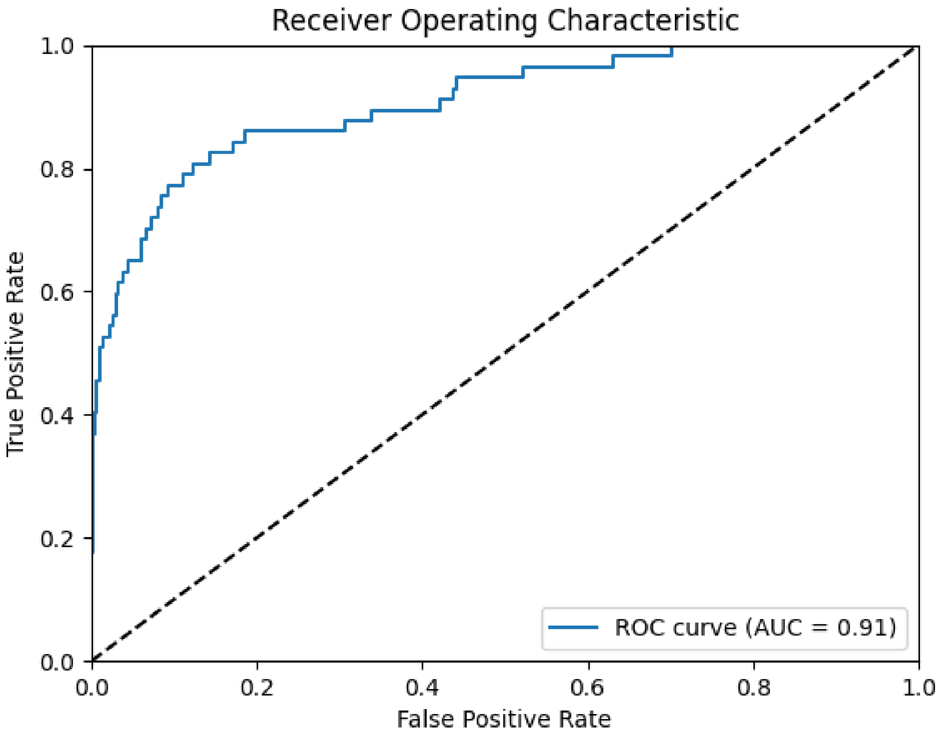

Figure A.24: Receiver Operating Characteristic - Logit - Dataset 4

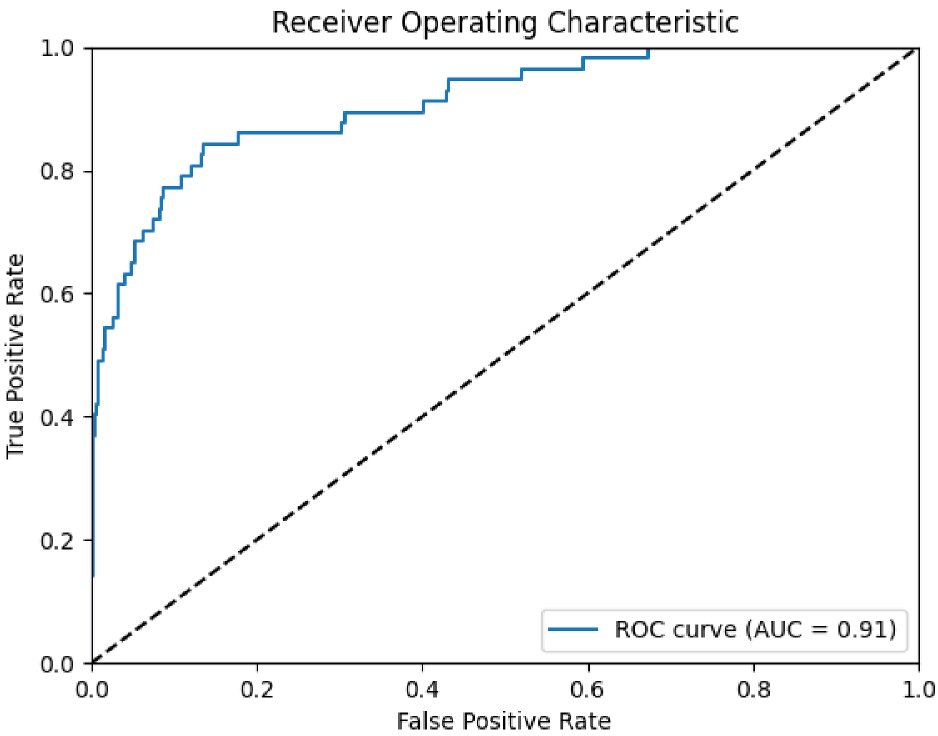

Figure A.25: Receiver Operating Characteristic - Probit - Dataset 4

A.5 Dataset 5

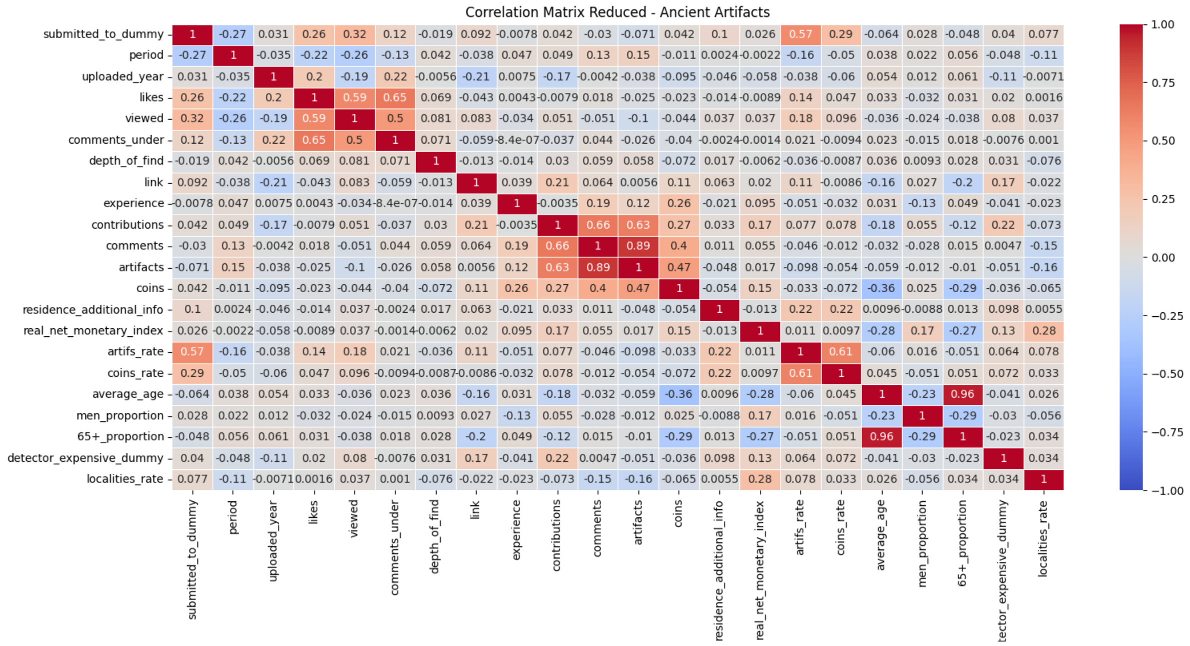

Figure A.26: Correlation Matrix - Dataset 5

Predicted	0	1
Actual 0	9970	233
Actual 1	1309	1097

Table A.13: Confusion Matrix - LPM - Dataset 5

Predicted	0	1
Actual 0	9799	404
Actual 1	935	1471

Table A.14: Confusion Matrix - Logit - Dataset 5

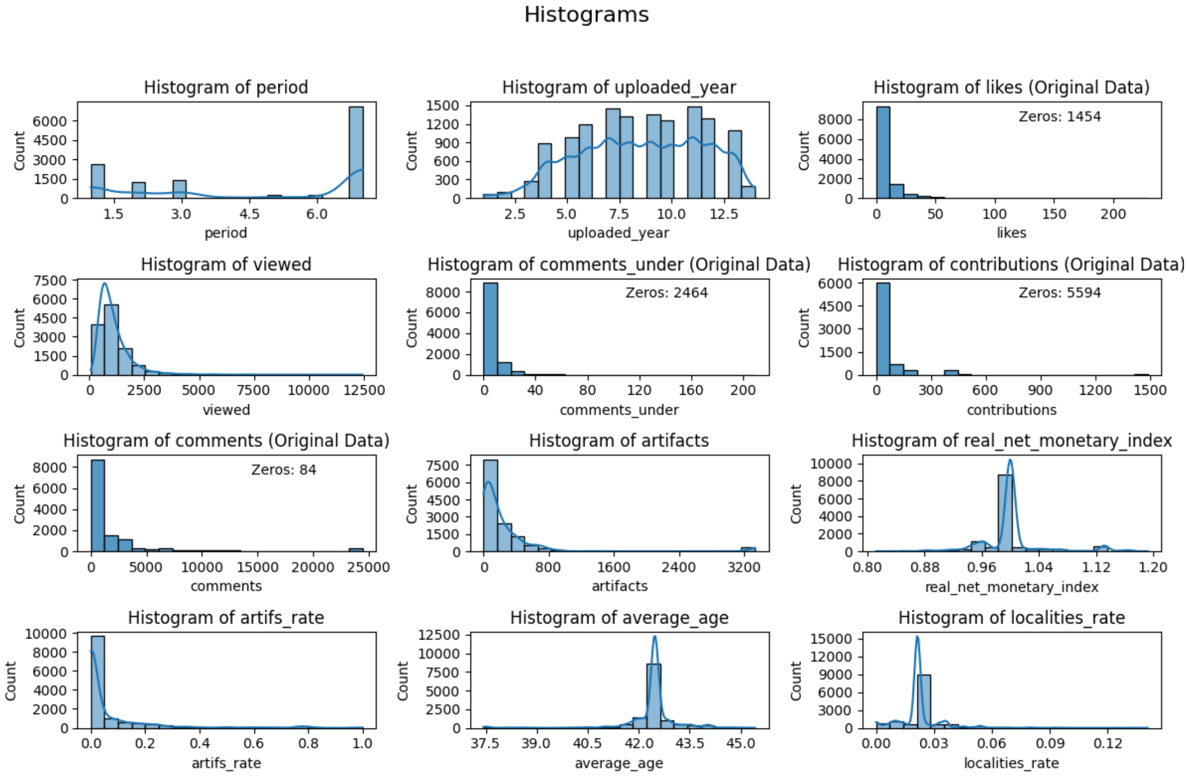

Figure A.27: Histograms - Dataset 5

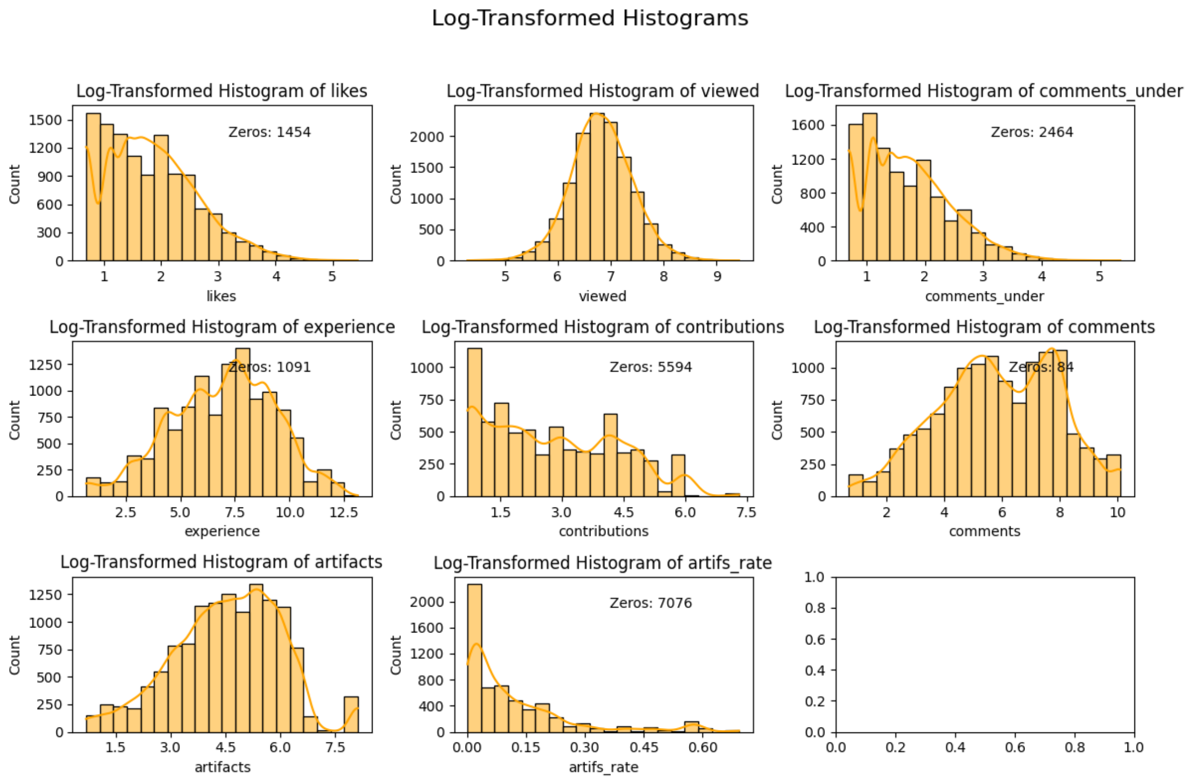

Figure A.28: Log-transformed Histograms - Dataset 5

Predicted	**0**	**1**
Actual 0	9838	365
Actual 1	1021	1385

Table A.15: Confusion Matrix - Probit - Dataset 5

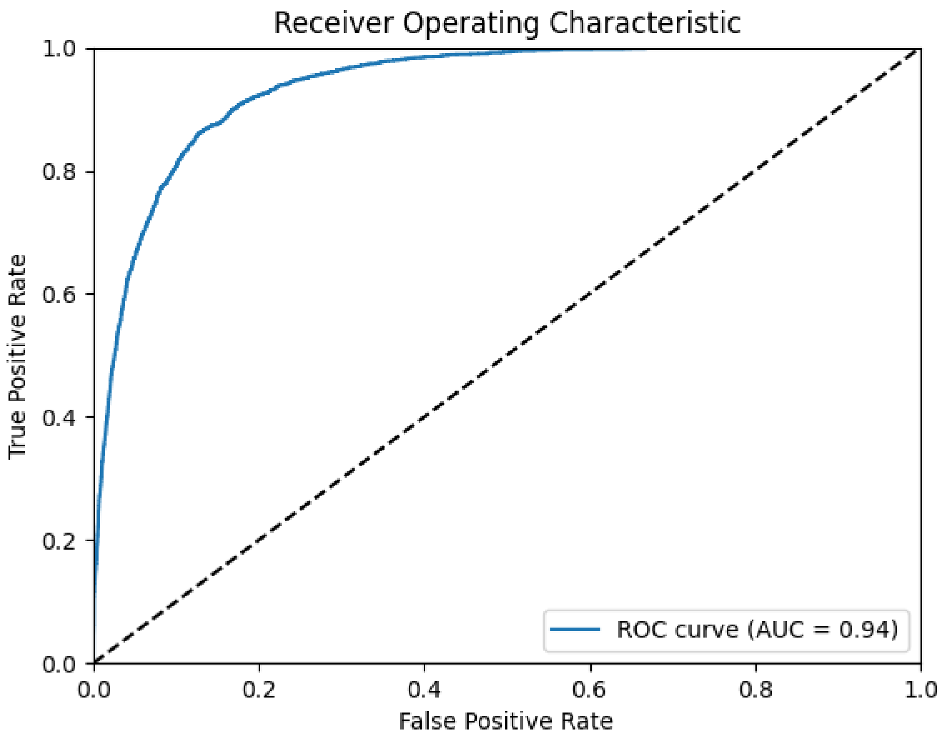

Figure A.29: Receiver Operating Characteristic - Logit - Dataset 5

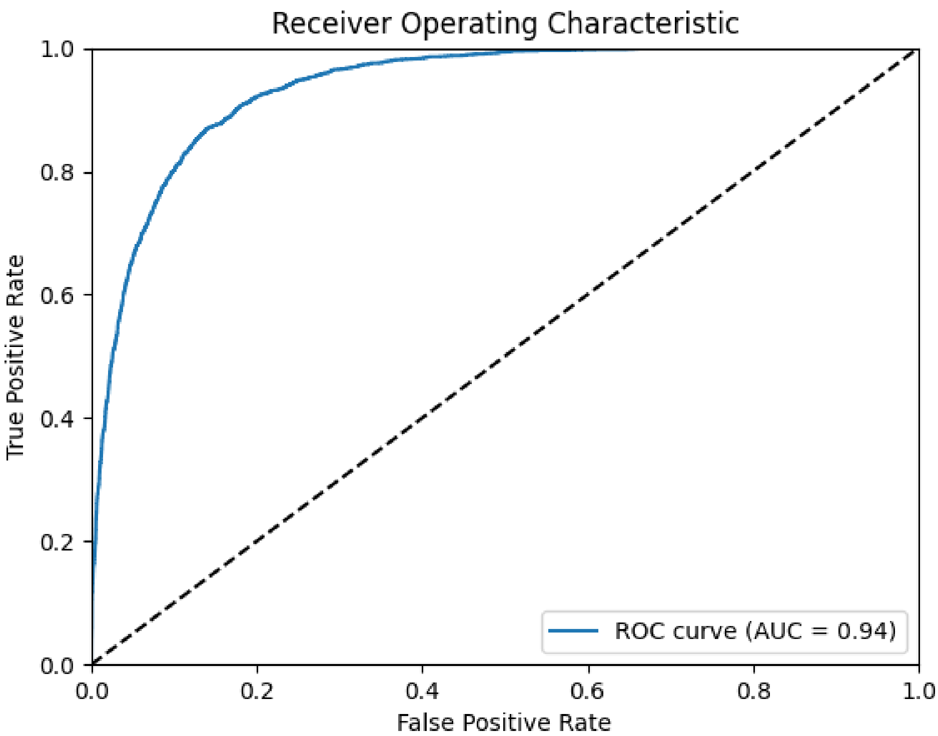

Figure A.30: Receiver Operating Characteristic - Probit - Dataset 5

Appendix B

Empirical Data and Source Codes

- Link to GitHub for empirical data and source codes:

  https://github.com/hawk-s/Metal_Detecting_Ownership_and_Non-Ownerhsip_Motives