MASTERARBEIT / MASTER’S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

„The Determinants of the Royalty Rate in Licensing Agreements“

verfasst von / submitted by

Barbara Vieider, B.Sc.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 913
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Volkswirtschaftslehre
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. Robert Kunst

Contents

ABSTRACT

This thesis analyses the determinants of the price for licensing intangible property, the royalty rate, for a sample of 1,834 agreements from the RoyaltyRange database, which has not been used in economic literature so far. For this, we use regression trees and their extension random forest as well as an OLS estimation. According to our findings, a global grant (instead of a restricted geographical scope), related parties, licensors who are universities, non-profit entities or individuals, the age of the agreement, the age of the licensor at the time of the agreement stipulation as well as the age difference between the licensor and the licensee are negatively associated with the royalty rate. We do not find any evidence for a significance influence of the exclusivity or the contract duration.

Diese Arbeit analysiert die Determinanten des Preises für die Lizenzierung von immateriellem Vermögenswerten, den Lizenzgebührensatz, für eine Stichprobe von 1.834 Lizenzverträgen aus der RoyaltyRange-Datenbank, welche bisher noch nicht in der Wirtschaftsliteratur verwendet wurde. Dazu verwenden wir Regression Trees und deren Erweiterung Random Forest sowie eine OLS Schätzung. Nach unseren Erkenntnissen sind ein globaler Geltungsbereich (anstelle eines begrenzten geografischen Geltungsbereichs), verbundene Unternehmen, Lizenzgeber, die Universitäten, Non-Profit Organisationen oder Einzelpersonen sind, das Vertragsalter, das Alter des Lizenzgebers zum Zeitpunkt der Vertragsunterzeichnung und der Altersunterschied zwischen Lizenzgeber und Lizenznehmer negativ mit dem Lizenzsatz verbunden. Wir finden keine Hinweise auf einen signifikanten Einfluss von Exklusivität oder der Vertragslaufzeit.

1. INTRODUCTION

The Organisation for Economic Co-operation and Development (OECD) reiterates in its compendium on intangible property (IP) that IP has grown in importance as a factor of production and taken on a central economic role while technical change and globalisation reshape the world economy (OECD, 2007). In fact, the OECD describes the growing significance of IP as one of the most important commercial developments in recent decades (OECD, 2006). Among the different ways of exploiting IP, one of the most common is licensing (Bakker and Verlinden, 2018). The licensing agreement constitutes the legal framework for licensing and specifies the respective terms and conditions as well as the price agreed on, which is either the royalty rate, a fixed fee or a combination of the two. In such an agreement, the royalty rate is usually expressed as a percentage of the revenue obtained using the IP and can add up to a significant part of a company's total revenues or expenses. According to the United States Patent and Trademark Office, revenue specific to the licensing of IP rights amounted to $115.2 billion in the US alone in 2012 (US Department of Commerce, 2016).

While the royalty rate agreed on for the licence of a specific IP results from the negotiations between the parties involved, several factors play an essential role in the determination of its level. The purpose of this analysis is to identify these factors and to examine their respective influence on the variation in royalty rates.

Many studies discuss the theory of licensing and the underlying principles that drive the behaviour of the contracting parties. However, empirical research on factors that determine the pricing in licensing agreements is relatively underdeveloped since the data necessary is rare. This work contributes to fill the gap in the literature by performing a data analysis with the aim to isolate some of the most relevant determinants of the royalty rates and to analyse how they correlate with the price. The dataset used for this analysis, to the best of our knowledge, has not been used in economic literature so far. It enables us to perform a large-scale study in order to analyse some factors that have already been studied before as well as additional determinants. Furthermore, we are the first to take a tree-based approach in addition to turning to economic theory to help in the variable selection in the context of analysing licensing data.

Gaining a better understanding of the factors that can explain differences in the royalty rate can inter alia help in the determination of "reasonable royalties" in patent-infringement lawsuits and to approximate arms-length prices for transfer pricing purposes. The latter has become increasingly important, because tax authorities have exhibited more concern about multinational firms transferring IP, and thus any profits that accrue to it, to subsidiaries in low-tax jurisdictions (Kankanhalli and Kwan, 2018).

This thesis is structured as follows. In section 2, we introduce the theoretical background on licensing while section 3 presents an overview of the related literature. Section 4 contains the hypotheses on how various factors could influence the royalty rate. In section 5, we describe and analyse the data, then we apply an off-the-shelf algorithm from the machine learning literature in section 6 to explore which factors available in the dataset best predict the royalty rates. In section 7, we apply an ordinary least squares (OLS) regression model. The final section concludes.

2. INTANGIBLE PROPERTY AND LICENSING

IP represents value, but lacks a physical embodiment able to be seen and touched, other than perhaps being represented by a document. Among others, such IP includes patents, copyrights and trademarks and may be used to generate significant economic value (OECD, 2017). IP is widely held to be the key profit and value driver for multinational enterprises, the source of their competitive advantage or perhaps dominant market position as well as the cause of barriers to entry into a market, which features might warrant an above-average remuneration for an entity holding the intangibles (Bakker and Verlinden, 2018).

Restricting imitation and duplication allows the holder to exercise a monopoly on the use of the IP. The degree to which this should be possible, determined by the Intellectual property rights (IPR) of the respective country, is highly controversial. Without or a low level of IPR protection, the social costs of monopoly power could be avoided. Furthermore, patents could hinder, slow down or even stop the innovation process, especially in highly innovative sectors, such as the software industry. However, the prospect of monopoly earning is often the reason for research and innovation in the first place. The resulting higher levels of research and innovation may offset the social costs of monopoly power and drive economic growth. A prominent example is the market launch of a new drug. Not protecting IPR would most probably make the drug (more) accessible and due to competition cheaper for those in need of it. Furthermore, other researchers could immediately build on the new research findings. However, it is possible that the drug would not have been invented in the first place, because the respective research costs could have been too high in comparison to the expected market price. There is some consensus in academic literature that there appears to be an optimal level of IPR regulation above which further enhancement reduces innovative activities (Qian, 2007). In line with this, several publications in academic literature suggest that a reform of IPR is far more attractive than abolition because it retains the good while minimizing the bad (Evans and Layne-Farrar, 2004).

Because IP is protected to some degree in most countries, one way of exploiting it is licensing. By definition, a licence is a "permission granted by and IP holder, the licensor, to another legal entity, the licensee, to make use of, sell or otherwise benefit from the underlying IP under certain restrictive conditions" (Grandstrand, 1999). There are several motivations for licensing, which are widely discussed in economic literature. The popularity and traditional use of licensing from the early age of modern industrial practices seem to be rooted in the fact that licensing ensures that the IP owner retains ownership of the IP (such as an innovative product, technology or copyright), but still receives an adequate amount of periodic royalties from the exploitation (Bakker and Verlinden, 2018). For example, granting a license is particularly important for companies that have limited resources to commercialize their own inventions (Sakakibara, 2010). Companies might choose licensing if the invention falls into areas that are not core to their business (Sakakibara, 2010). For licensees, licensing is an efficient way of accessing (pre-existing) technology or IP without the need to develop it from scratch. Thus, the reduction of R&D costs and the length of time of the innovation process make licensing an attractive solution (Bakker and Verlinden, 2018).

The legal framework for licensing is specified in a licensing agreement. A licensing agreement is a contract by which an owner of an IP, the licensor, permits another party, the licensee, to use the IP in question in accordance with the terms and conditions of the agreement. Due to the contractual freedom, the specific parameters under which the IP is licensed are negotiated freely by the parties involved. Non-compliance with the agreement results in breach, which is a serious consideration from both the legal and business perspectives (Bakker and Verlinden, 2018). Thus, setting the terms of the license, especially the various restriction of the exploitation in terms of geographical limitations or time limitations, is crucial (Bakker and Verlinden, 2018).

The royalty rate is the payment made by the licensee to the licensor, for the use of intangibles owned by the licensor as it is stated in the licensing agreement. Royalty rates are often expressed as a percentage of the revenues obtained using the owner's property. However, they can also be expressed in other terms (including a fixed value), depending on the specific characteristics of the licence agreement. Since past studies showed that the royalty rate represents the patent licensing price better than lump-sum payments (Sakakibara, 2010) and given the respective data availability, we choose to focus exclusively on the royalty rate as a form of remuneration for licensing in our analysis.

3. LITERATURE REVIEW

While many studies discuss theoretic models of licensing games (e.g. Gordanier and Miao, 2011) or offer insights into the economic and strategic motivations for licensing (Fosfuri, 2006), its advantages and disadvantages as well as the legal framework, few papers have studied what factors determine the price of licensing an IP based on empirical evidence.

One of these few is Sakakibara (2010), which claims to be the first empirical study to analyse pricing in patent licensing. Using a Three-stage-least-squares approach, Sakakibara (2010) examines the determinants of the price of patent licensing with data on Japanese patent licensing contracts, which were concluded between 1998 and 2003. The author finds that factors affecting the profitability of patents and bargaining power of licensors and licensees are good predictors of the royalty rate while proxies for the reservation price of licensors are less important. Furthermore, Sakakibara (2010) investigates the effect of the licensor and the licensee operating in the same industry and of the licence being exclusive. In this regard, the author finds no significant effect of neither exclusivity nor of the parties operating in the same industry.

Jayachandran et al. (2013) examine how country and contract characteristics influence the royalty rate in brand licensing using an ordinary least squares regression with data on brand licensing agreements. Although Jayachandran et al. (2013) are focusing on the risk of moral hazard, which is arguably more important in brand licensing compared to other licensing types, as the main influence on the price of brand licenses, they offer valuable insight on contract duration and contract exclusivity. According to their findings, contract duration and exclusivity had significant negative associations with royalty rate. They motivate this result by arguing that licensors depend to a greater extent on the licensee to protect the brand by using it appropriately with longer-term and exclusive contracts. This reasoning suggests that the results obtained in Jayachandran et al. (2013) could possibly be specific to brand licensing contracts.

Kankanhalli and Kwan (2018) provide the first large-scale, systematic empirical study of the economic determinants of IP licensing royalty rates under a bargaining power framework using a database of IP licensing agreements from RoyaltyStat LLC. In terms of contract-level factors, they find that exclusive contracts earn significantly higher rates. According to their results, fewer territorial restrictions are associated with lower rates, while the presence of know-how is associated with higher rates. Moreover, they find that licensors with limited bargaining power (or incentives to bargain) such as individuals, non-profits and universities command significantly lower rates. Additionally, they find that contracts in which the licensor's underlying technology is more substitutable exhibit lower rates.

4. HYPOTHESES

Several different aspects have a determinant influence on the agreed royalty rate of a licence. According to Sakakibara (2010), the most prominent factors are the profitability of the underlying patent and the relative bargaining power of the licensor and the licensee. In addition, various individual characteristics of the contracting parties as well as specific contractual features are linked to the price of a licence.

The relevant factors that have been identified in related literature and the respective hypotheses regarding their influence on the price are discussed in the following.

Exclusivity

Licensors negotiating exclusive contracts have a higher minimum willingness to accept, since they cannot relicense the same IP to another party (Kankanhalli and Kwan, 2018). Similarly, licensees have a higher maximum willingness to pay for exclusive rights to IP, given that this pre-empts the revenues for commercialization from being competed away by other parties (Kankanhalli and Kwan, 2018). Hence, exclusive contracts should command higher rates.

Hypothesis 1: The exclusivity is positively related to the royalty rate.

Geographical scope (Global Grant)

According to economic theory, contracts with greater geographical restrictions on the use of the licensed IP command lower rates (Kankanhalli and Kwan, 2018). This is because licensees would have, all else equal, a lower maximum willingness to pay for a more restrictive contract as their potential revenue base is limited (Kankanhalli and Kwan, 2018).

Hypothesis 2: A global grant is positively related to the royalty rate.

Related Parties

About 6 % of the agreements in the database were concluded between related parties. One could argue that prices set between related entities do not reflect market prices, but are relationship driven. In this case several additional factors could influence the price, such as compensations by other means, fewer risk factors, less information asymmetry, attempts to support related entities by offering lower royalty rates or to shift profits for tax evasion purposes. Although the direction of these effects is unclear, we conjecture that the overall effect of related parties on the royalty rate is negative.

Hypothesis 3: Related parties are negatively related to the royalty rate.

Identity of the counter-party

Kankanhalli and Kwan (2018) suppose that, when firms face counter-parties which are non-firm entities, for example non-profits and universities, the royalty rates should be lower. This is because such entities are, often by definition, not profit-seeking and may have other non-commercial objectives when entering into licensing agreements. Hence, as licensors, their minimum willingness to accept would be lower than profit-making entities.

Hypothesis 4: If the licensor is a university, a non-profit entity or an individual instead of a for-profit company, the royalty rate is, ceteris paribus, lower.

Agreement Duration

Jayachandran et al. (2013) argue that the contract duration is likely to influence royalty rates. However, the direction of this influence is not clear. On the one hand, a longer-term contract might correlate with a higher royalty rate to account for the risk that licensors depend to a greater extent on the licensee to use the licence appropriately while licensors exhibit a higher willingness to pay for a licence they can capitalize on for longer. On the other hand, a longer-term contract could signal a lower risk regarding the expected profitability in the presence of asymmetric information and thus correlate with a lower royalty rate. We conjecture that the latter effect is stronger and, consequently, that the overall effect is negative.

Hypothesis 5: The agreement duration is negatively related to the royalty rate.

Bargaining Power

Licensing agreements are the outcome of a negotiation between licensor and licensee. There is extensive literature providing clear evidence that bargaining power plays a major role in contract negotiations. The more bargaining power a party has the more favourable a contract will be for that party. Hence, in the case of licensing agreement, the royalty rate should be higher for a licensor with more bargaining power relative to the licensee and vice versa.

Given significant asymmetric information about the underlying IP between parties in licensing negotiations as well as several other differentiating factors, licensors and licensees are assumed to possess significantly unequal market power and bargaining strength (Dawson, 2013). Kankanhalli and Kwan (2018) and Sakakibara (2010) obtain a number of results consistent with this hypothesis.

However, since bargaining power cannot be measured, researchers have to rely on appropriate proxy variables in order to be able to draw some conclusions regarding its influence. In the related literature the most prominent proxies are the size of the respective parties measured by the number of employees or the financial situation.

We conjecture that information on the identity of the counter-party, the age of the licensor and the difference in the age of the licensor and the licensee can to some extent allow to control for bargaining power in licence agreement negotiations.

Age of the Licensor and Difference in the Age of the Licensor and the Licensee

We conjecture that licensors that are existing for longer relative to the respective date of the contract conclusion date have greater relative bargaining power and thus command higher rates. The assumption underlying this argument is that older entities have more market power, are in a better financial situation and are already more established in the respective market in contrast to start-up companies.

Arguably, a similar logic can be applied to the difference between a licensor and a licensee rather than the absolute amount of years. The hypothesis in this case would be that the larger the age difference between a licensor and a licensee, i.e. the older the licensor compared to the licensee, the bigger his relative bargaining power and the higher the royalty rate agreed between the two.

Hypothesis 6: The age of the licensor and difference in the age of the licensor and the licensee is positively related to the royalty rate.

5. DATA

Sakakibara (2010) highlights that actual licensing contract data is rare and has not been used often. One of the most important reasons is that firms tend to conceal information about licensing deals, which are typically considered to be strategic decisions that are rarely publicly disclosed (Fosfuri, 2006).

For the purpose of this thesis, we have been granted access to the RoyaltyRange database. RoyaltyRange is one of a handful of companies that collect data on royalty rates of IP licensing transactions. According to their website1, it is compiled by analysing approximately 150,000 agreements from a large number of publicly available sources, including securities and exchange commissions, and international stock exchanges and court databases. From those agreements, only the ones that comply with certain standards and requirements are entered into the database, such as disclosed remuneration mechanisms. The number of agreements with the information available for download in the database2 amounts to 3,576.

Generally, this database is used for benchmarking transfer pricing transaction, valuation of intangibles and purchase price allocations[3]. To the best of our knowledge and the knowledge of its operators, it has not been used in economic literature so far.

The unit of observation is a licensing contract entered into by two entities, a licensor and a licensee. In terms of the contract itself, we observe contract terms, payment terms and intangible information. Contract information consists of the effective date of the contract, the duration, the geographic scope of the licensing agreement, and the rights granted by the licensing firm to the licensee.

The information on each agreement is highly heterogeneous qualitative data, which had to be transformed to allow for a statistical analysis. One of the few numerical variables is the royalty rate, which is expressed in a percentage form. For some agreements, more than one royalty rate was stated. In this case, we took a simple average as price for the licence.

The information on the year the agreement was concluded, which coincided with the year of the effective date of the agreement in 98% of the cases, and the date of cessation was used to calculate the age and the duration of the agreement. However, there was no

[1] https://www.royaltyrange.com/home/royalty-rate-database/about-us
[2] Download: 26.10.2018 – 04.11.2018
[3] Source: https://www.royaltyrange.com/home/royalty-rate-database/about-us, accessed on 06 April 2019

date of cessation stated for about half of the agreements. Following the approach of Jayachandran et al. (2013), we considered the duration of the contract 99 years when it was indefinite.

The characteristics that could be meaningfully formalized by creating dummy variables are the exclusivity of a license, the geographical scope and the (NACE Code) sector to which it was assigned by the RoyaltyRange database operators. Since the territory for which the agreement is specified was highly heterogeneous, we decided to create a dummy variable only accounting for whether the agreement had a global scope or was limited in this respect.

To facilitate comparability while analysing the royalty rate, it is of central importance to consider the base it is applied on, since in absolute terms, the level of the royalty rate is not conclusive or comparable. For this reason, we excluded licence agreements that specified a base other than net sales. Following this reduction, the dataset contains 1,834 agreements.

6. SUMMARY STATISTICS

The royalty rate on the 1,834 agreements in our dataset ranges from 0-60% with a lower quartile of 2%, an upper quartile of 7% and a median of 4% (see Table 1).

Table 1: Distribution of the Royalty Rate

N	Royalty Rate (%)
Mean	5.52
St. Dev.	5.54
Variance	0.31
Min	0.00
Q25	2.00
Median	4.00
Q75	7.00
Max	60.00
N	1,834

Figure 1 visualizes the distribution of the royalty rates in the data. A royalty rate between 2-5% is observed in over 50% of the cases and a royalty rate of 1-6% in over 2/3 of the cases with 2% being the price agreed on most frequently.

The agreements were concluded between 1977 and 2018, while approximately 2/3 of the agreements were entered into between 2000 and 2011 (see Figure 2).

Figure 1: Distribution of the Royalty Rate

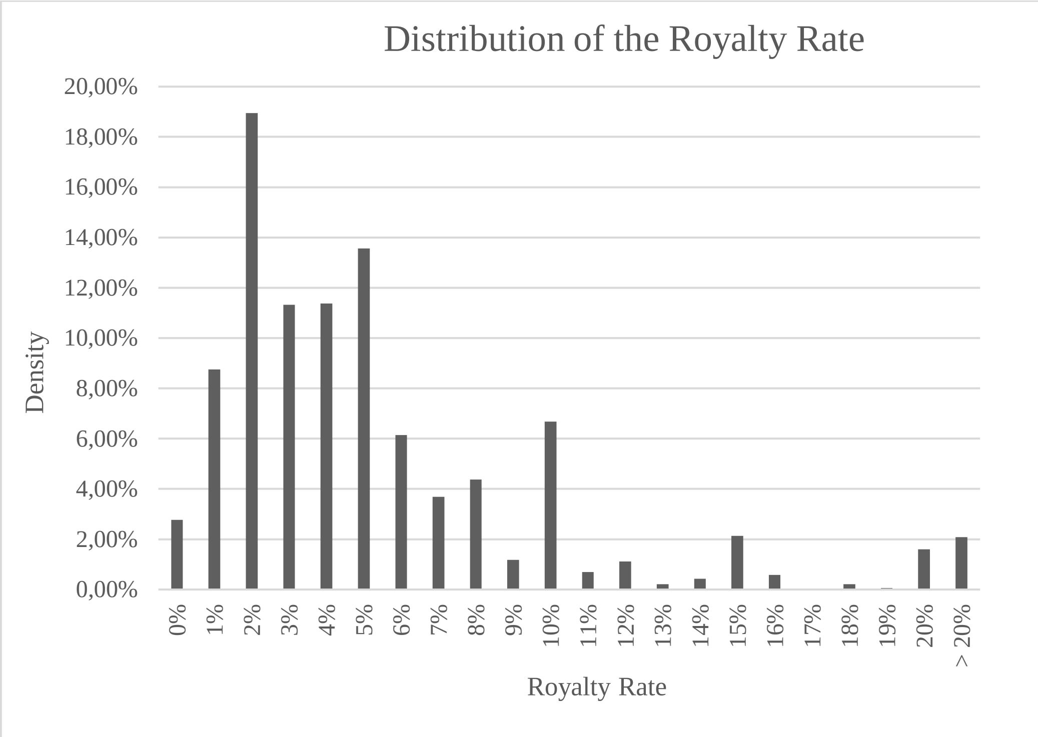
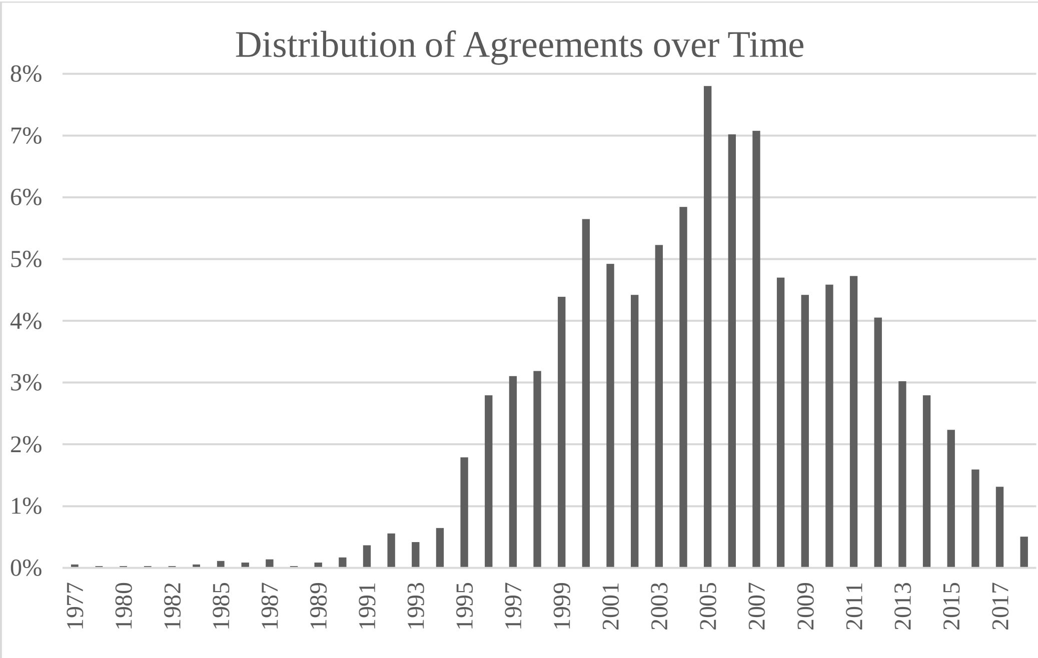


Figure 2: Distribution of Agreements over Time



When RoyaltyRange analysts enter the agreements into the database, each agreement is allocated to the industry (or the industries) that best fit the description of the underlying IP being licensed. Since many agreements were classified into more than one industry, the shares in the distribution of agreements by industry depicted in Figure 3 add up to more than 100%. The description of the NACE Statistical Classification of Economic Activities according to which the agreements are classified can be found in the Appendix.

The vast majority of agreements fall under the categories of sector C (Manufacturing) and G (Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles), followed by the sectors M (Professional, Scientific and Technical Activities) and Q (Human Health and Social Work Activities).

Figure 3: Distribution of Agreements by Industry

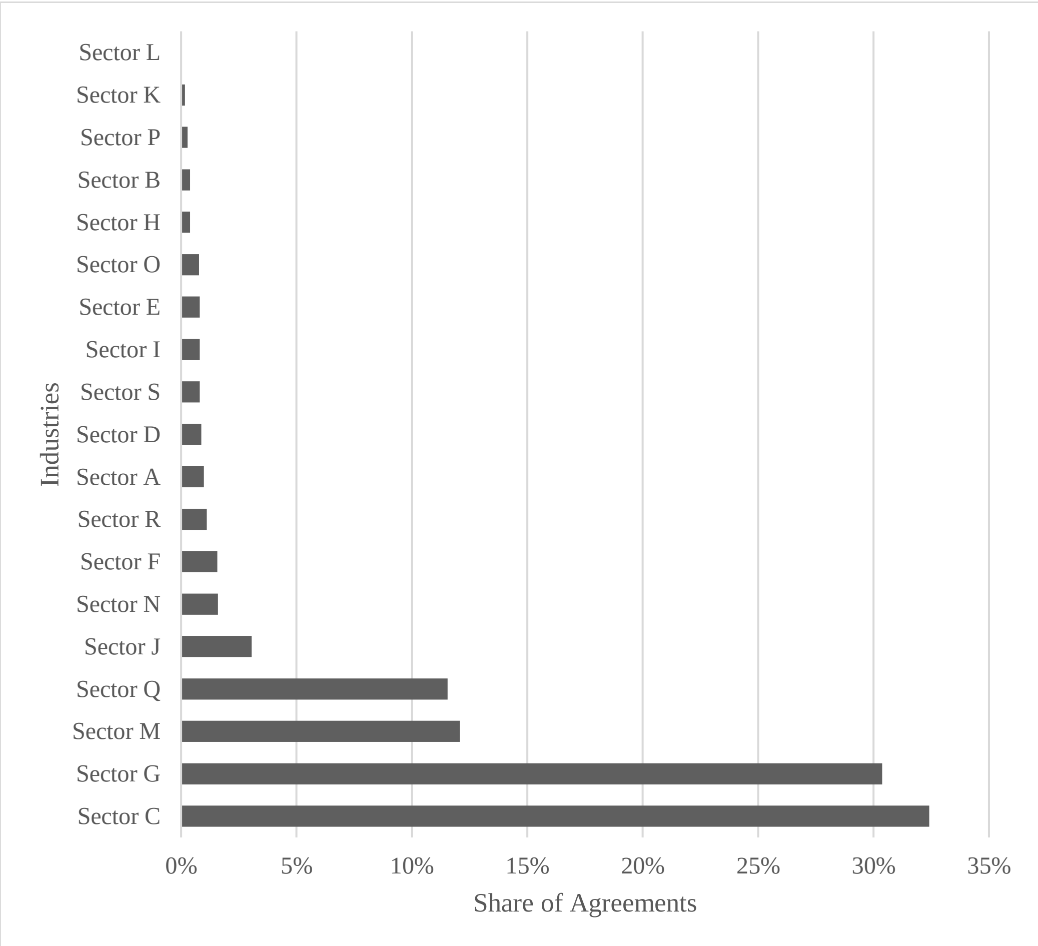
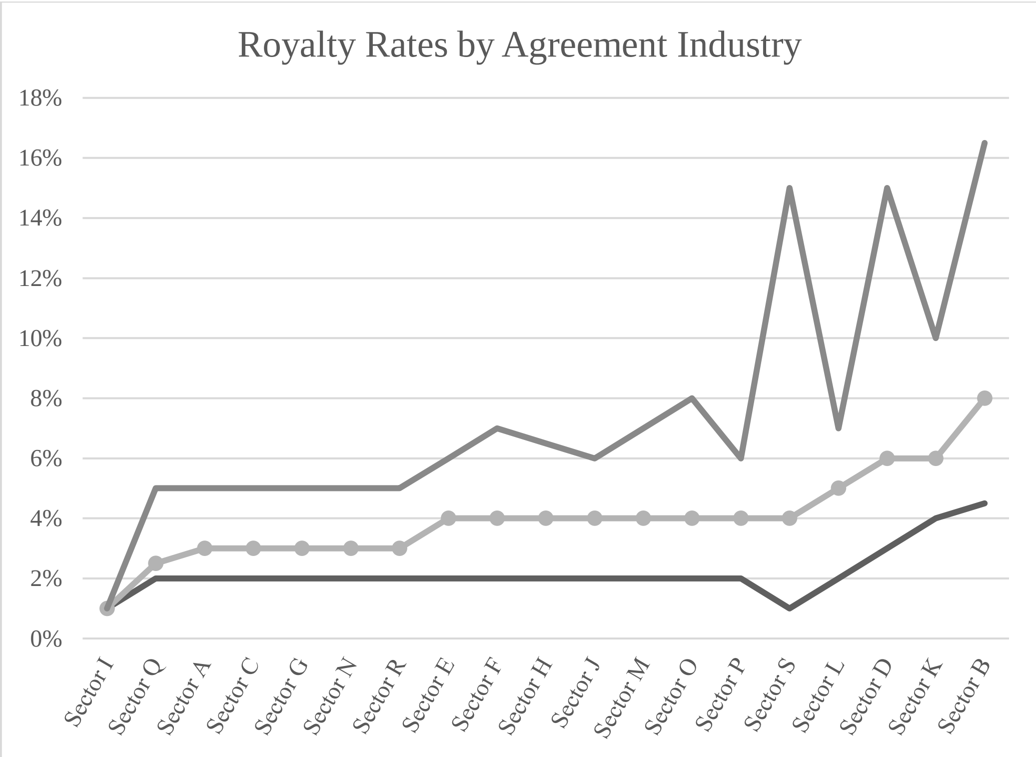


Figure 4 shows the median and interquartile range of royalty rates across industries[4] and thus gives an indication for its within- and across-industry variation. For example, the interquartile range of the licence fee rates for the 168 agreements in sector J (Information and Communication) is 2% - 6% with a median of 4%.

While sector I (Accommodation and Food Service Activities) exhibits a relatively low royalty rate with no within-industry variation, the within-industry variation seems to increase with the level of the royalty rate. Sector B (Mining and Quarrying), K (Financial and Insurance Activities) and D (Electricity, Gas, Steam and Air Conditioning Supply) show the highest median royalty rates along with substantial within-industry variation. The differences across industries can, inter alia, be explained by differences in patent protection and the profitability across industries.

Figure 4: Median and Interquartile Range of Royalty Rates across Industries



[4] Please note that Sectors P, L and K have less than 20 observations each.

7. TREE-BASED MODEL

To further analyse the data, we apply a tree-based model. Trees have been used widely for a number of years as one of the primary machine learning tools. In this specific case, we use regression trees (Breiman et al., 1984), and their extension random forest (Breiman, 2001a) to explore which factors available in the dataset best predict the royalty rate.

For a random forest, a dataset without missing observations is needed. One way to deal with missing observations is to simply discard observations with any missing values. This approach can be used if it can be assumed that the features are missing completely at random and if the relative amount of missing data is small. Discarding observations with missing values results in a dataset with 1699 observations (135 had to be deleted).

For the tree-based model, we use two versions of the dataset: one that includes the industries and one without industries.

Regression Trees

When building a regression tree one stratifies the predictor space, i.e. the set of possible values of royalty rates in the dataset, into a number of distinct and non-overlapping regions and then makes the same prediction given by the mean of the response values for the training observations in each region for every observation that falls into that region. The splits are sequential, based on a single covariate at a time exceeding a threshold that minimizes the average squared error. These splits can be summarized and visualized in a tree.

First, we randomly split the dataset into a training set (80% of the data) and a test set (20% of the data). Then, we take a *top-down*, *greedy* approach that is known as *recursive binary splitting* to grow a large tree, since it is computationally infeasible to consider every possible partition of the feature space. The approach is *top-down* because it begins at the top of the tree (at which point all observations belong to a single region) and then successively splits the predictor space; each split is indicated via two new branches further down on the tree. It is *greedy* because at each step of the tree-building process, the *best* split is made at that particular step, rather than looking ahead and picking a split that will lead to a better tree in some future step. We use cost complexity pruning, not to overfit the data and to account for the trade-off between a lower variance and less bias. In doing so, we prune the obtained large tree back to obtain a subtree.

The resulting big trees of the two versions of the dataset, which are shown in Figure 5 and Figure 6, are relatively incomprehensible and difficult to interpret, which makes any conclusions drawn from them void. However, the closer a branch is to the leaf (i.e. the further down in the tree), the less important the factor in determining the response, because the early splits which have caused a big decrease in the sum of squares are depicted by longer branches. The branches get smaller as the incremental improvements get smaller.

Figure 5: Large Tree with Industries

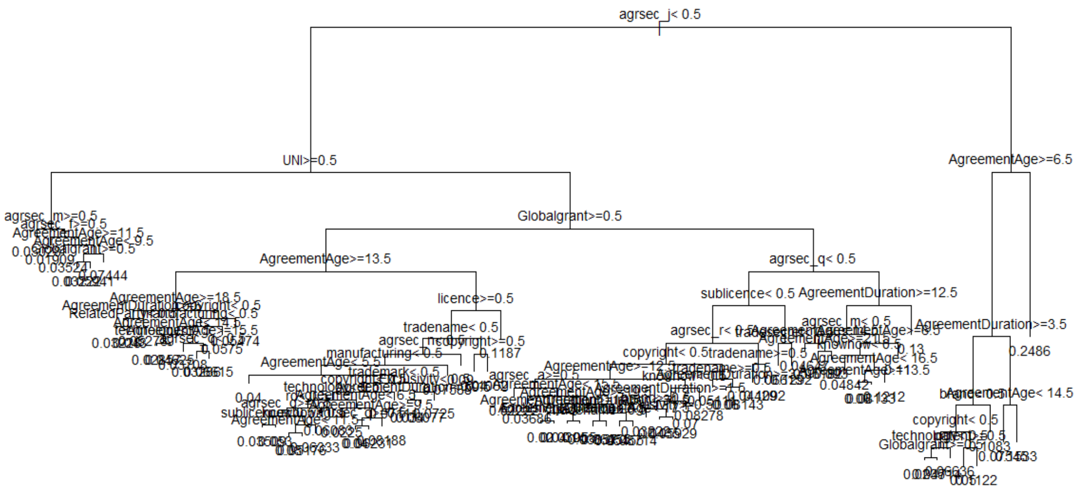


Figure 6: Large Tree without Industries

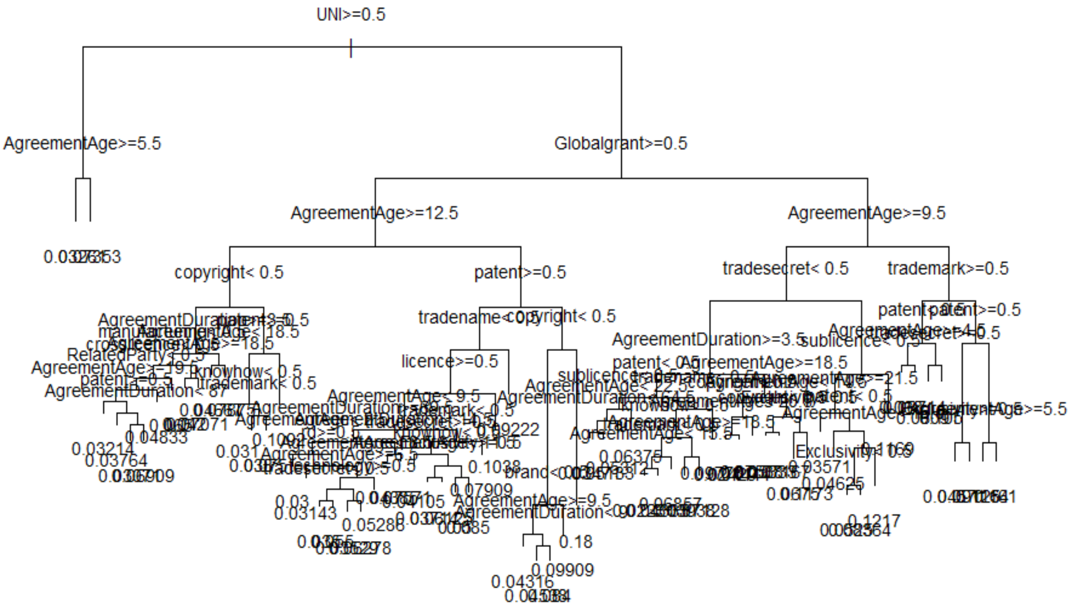
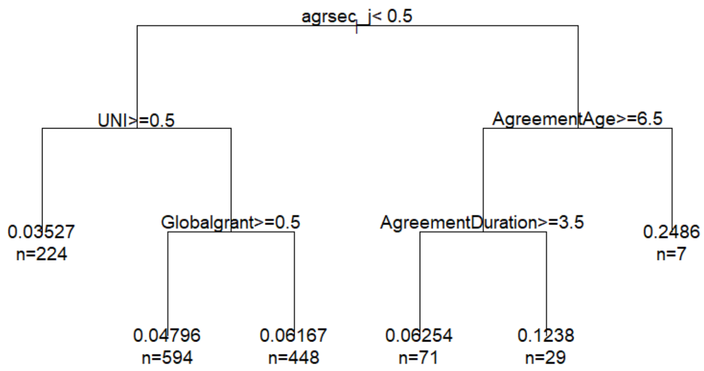


Figure 7: Pruned Tree with Industries



Pruning the obtained large trees results in the trees depicted in Figure 7 and Figure 8.

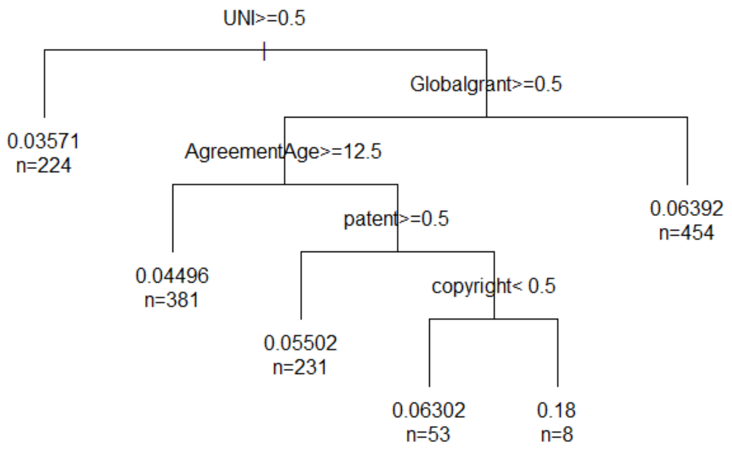
The tree in Figure 7 has 5 internal nodes and 6 terminal nodes or leaves. At a given node, the label indicates the left-hand branch emanating from that split, and the right-hand branch corresponds to the opposite of the label. The leaves indicate the predicted royalty rate for the respective region and the number of agreements in the training set that fall into this region.

For example, the predicted royalty rate for a licence agreement with an underlying IP not allocated to sector J (Information and Communication), where the licensor is a university, non-profit entity or individual is 3.53%. 224 agreements in the training set fall into this region. The predicted royalty rate for a licence agreement with an underlying IP allocated to sector J (Information and Communication), where the agreement has been stipulated after over (and including) 6.5 years ago and an agreement duration of over 3.5 years is 6.35%. 71 agreements in the training set fall into this region.

The pruned tree resulting from the dataset without industries is depicted in Figure 8.

.

Figure 8: Pruned Tree without Industries



The tree in Figure 8 has 5 internal nodes and 6 terminal nodes or leaves as well. However, since the dataset does not contain information on the industries anymore, it exhibits different splitting rules. For example, if the licensor is not a university, non-profit entity or individual, the geographical scope is not restricted and the agreement was entered into 12.5 years or longer ago, the predicted royalty rate is 4.5%.

Although trees are easy to interpret, it is important not to go too far in interpreting the structure of the tree, including the selection of variables used for the splits (Athey and Imbens, 2019). Particular covariates that have strong associations with the outcome may not show up in splits, because the tree splits on covariates highly correlated with those covariates.

Random Forest

In the following we apply a random forest approach to obtain more information on the importance of the respective factors. By aggregating many trees, using random forests, the predictive performance of trees can be substantially improved at the expense of some loss in interpretation.

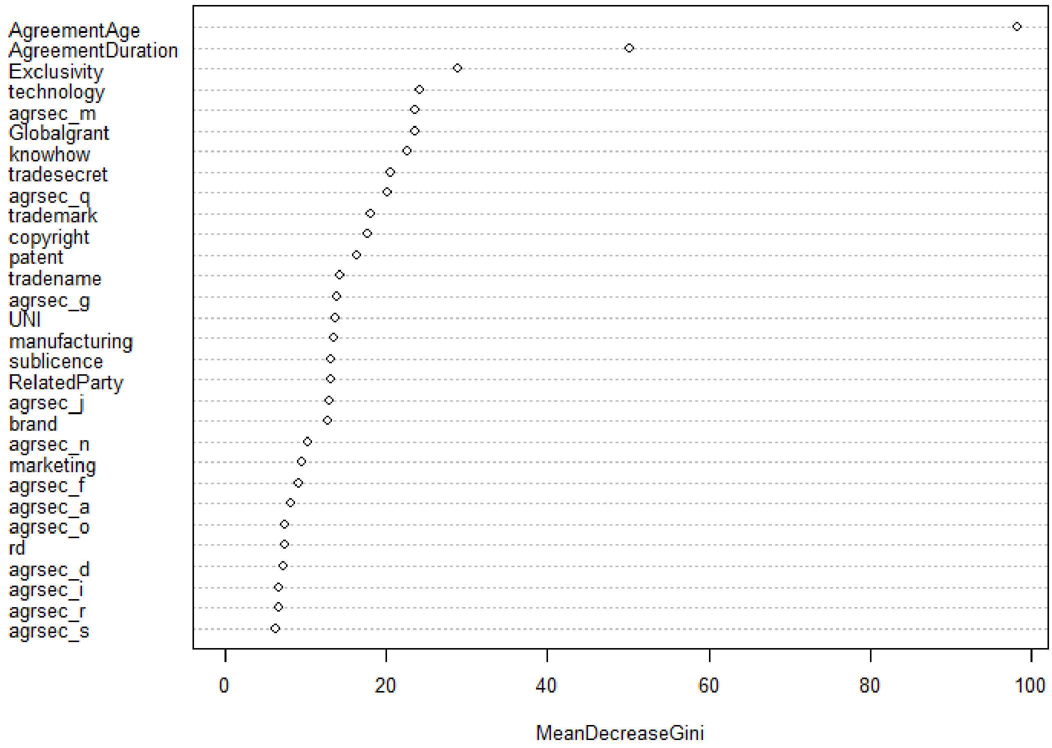
We build a 500 decision trees on the training sample of the smaller dataset.

To account for the fact that these trees could be highly correlated, meaning that they have very similar structures, and that averaging many highly correlated quantities (i.e. predictions from many trees from the same set) does not lead to as large a reduction in variance as averaging many uncorrelated quantities, random forests overcome this problem by forcing each split to consider only a subset of the predictors. One can think of this process as decorrelating the trees, thereby making the average of the resulting trees less variable and more reliable. This reduces the variance when averaging the trees.

Hence, when building the decision trees, each time a split in a tree is considered, a random sample of 6 predictors is chosen as split candidates from the full set of 41 predictors. A fresh sample of 6 predictors is taken at each split.

Figure 9 shows the result of the variable importance plot obtained with the random forest approach. The higher the mean decrease in Gini, the more important is a variable in predicting the royalty rate.

Figure 9: Variable Importance Plot with Industries



According to our results, how far the date of the conclusion of the agreement lies in the past and the agreement duration seem to be the most important factors needed to predict the royalty rate. In terms of contract features, the exclusivity and the geographical scope seem to be most decisive in the prediction. With regard to the types of licensing, technology licensing, know-how and trade secrets rank relatively high in the importance plot. The dummy variable for sector M (Professional, Scientific and Technical Activities) is the most important of the sector variables.

Discussion

It is important to note that trees generally do not have the same level of predictive accuracy as some other regression and classification approaches. Additionally, trees can be very non-robust. In other words, a small change in the data can cause a large change in the final estimated tree. Moreover, machine learning approaches generally give no direct indication on causal relationships.

However, regression trees also have a number of advantages over the more classical approaches. Traditionally in the empirical literature researchers limit the number of explanatory variables by hand, rather than choosing them in a data-dependent manner (Athey and Imbens, 2019) and rely on economic theory for this. However, one may not know ex ante which of the features matter, and which can be dropped from the analysis, especially if economic theory surrounding the research question is scarce or ambiguous.

Since economic theory and empirical research on the factors determining the royalty rate is scarce, we use the results obtained in the random forest to allow the data to play a bigger role in the variable selection process when applying a linear regression model in section 7. Hence, the variables ranking highest in the variable importance plot will be included in the linear regression model in section 7, independent of whether or not this is supported by economic theory or empirical studies on this topic.

8. OLS REGRESSION

Method

In order to quantify how the most relevant factors discussed in section 4 and available in the dataset correlate with the royalty rate, we estimate the following model using ordinary least squares (OLS) regression. Specifically, we estimate the following regression equation:

Equation 1: Baseline OLS Estimation

$$RR_i = \alpha + \beta_e E_i + \beta_g G_i + \beta_R R_i + \beta_u U_i + \beta_{aa} AA_i + \mu_i$$

In this specification, RR_i is the royalty rate for the licensing agreement i.

The explanatory variables chosen in this model on the basis of the hypotheses outlined above are a dummy for whether the agreement is exclusive (E_i), a dummy for whether the geographical scope is global and not restricted (G_i), a dummy for whether entities are related parties (R_i), and a dummy for the identity of the licensor, i.e. whether the licensor is a university, non-profit organization or a public entity, (U_i).

The variable importance plot obtained in section 6 indicates that, in addition to some of the variables already mentioned, the age of the agreement might be an important factor as well. Hence, we also include the agreement age (AA_i).

We include the agreement duration (AD_i) in a separate specification, because we lose many observations in doing so. The same holds true for the licensor's age at the time of the agreement (OA_i). In yet another specification, we replace this variable with the difference in age between the contracting parties instead, i.e. the subtracting the licensee's age from the licensor's age (D_i). We do not include them together as they might mitigate each other's influence since they are highly correlated (0.8125).

We do not include the industry dummies nor the country dummies, because we would have too many factors relative to the number of observations.

We use standard errors for the regression.

Results and Discussion

Table 2 presents the results. Column 1 shows the baseline estimation according to Equation 1. Standard errors are reported in parentheses below the coefficients while the stars indicate the level of significance.

Table 2: Results OLS Estimation

	(1) RoyaltyRate	(2) RoyaltyRate	(3) RoyaltyRate
Exclusivity	0.00328	0.00576	0.00971^{*}
	(0.00271)	(0.00396)	(0.00496)
Global Grant	-0.0135^{***}	-0.0190^{***}	-0.0154^{***}
	(0.00278)	(0.00419)	(0.00524)
Related Party	-0.0111^{***}	-0.0148^{**}	-0.0166^{***}
	(0.00419)	(0.00592)	(0.00617)
UNI	-0.0182^{***}	-0.0190^{***}	-0.0154^{**}
	(0.00325)	(0.00508)	(0.00607)
Agreement Age	-0.000921^{***}	-0.00154^{***}	-0.00151^{***}
	(0.000222)	(0.000269)	(0.000361)
Agreement Duration		0.0000618	0.0000176
		(0.0000443)	(0.0000571)
Licensor Age		-0.000124^{***}	
		(0.0000306)	
Age Difference			-0.0000856^{**}
			(0.0000362)
cons	0.0773^{***}	0.0879^{***}	0.0825^{***}
	(0.00449)	(0.00638)	(0.00860)
N	1699	769	434
R^2	0.045	0.093	0.085

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

First, the results show that there is no indication for a difference in royalty rates depending on whether the licence is exclusive or not. This finding is in line with the results of Sakakibara (2010). Although Kankanhalli and Kwan (2018) find a positive relationship between exclusivity and the royalty rate, this relationship is rather weak, because they find only marginal evidence at the 10% significance level. After investigating this puzzle further, they find that exclusivity is more valuable when a market is more competitive. Jayachandran et al. (2013) even find a negative association of exclusivity with the royalty rate for brand licensing. These mixed results are even more interesting given that exclusivity ranked relatively high in the variable importance plot (See Figure 9). Thus, we cannot confirm our hypothesis 1.

We find that a global grant correlates with significantly lower royalty rates, contradicting our hypothesis 2. In the baseline estimation reported in column 1, a global grant of a licence instead of a restricted geographical scope is associated with a decrease in the royalty rate of 1.35%. This amount even increases in the specification reported in column 2 and 3. This result is surprising since it contradicts the respective hypothesis made in section 4. However, it is in line with the findings of Kankanhalli and Kwan (2018) and Hegde (2014) who find that when a contract is worldwide (or similarly, when it has fewer territorial restrictions), its royalty rates are lower, contradicting their own hypothesis which corresponds to the one presented in this thesis. Jayachandran et al. (2013), on the other hand, do not observe any relationship between the use of contracts that cover multiple countries and the respective royalty rates.

Related parties command a lower royalty rate than third parties, confirming our hypothesis 3. The respective magnitude according to the first specification is 1.11%.

Furthermore, we find that licensors who are universities, non-profit entities or individuals ceteris paribus receive a 1.82% lower royalty rate relative to when the licensor is a firm. This is in line with our hypothesis 4 as well as the findings of Kankanhalli and Kwan (2018), who find results with the same direction although more pronounced in magnitude.

In contrast to the findings of Jayachandran et al. (2013), who report a negative effect of the agreement duration on the royalty rate, we find no significant effect of this factor. Thus, we could not confirm hypothesis 5.

Moreover, the hypothesis 6 that the age of the licensor at the time of the agreement stipulation and the difference in age between the licensor and the licensee have a positive effect on the royalty rate could not be confirmed. We find a small, but significant negative relationship between the two factors and the royalty rate. A possible explanation for this is that these two variables in fact do not reflect the bargaining power of the licensor.

Finally, our results show that agreements that have been stipulated further in the past exhibit lower royalty rates. However, the effect is very small. This variable was included based on the result of the random forest without support from economic literature.

Comparing the results between the two approaches, the tree-based approach and the linear regression model, would be misleading, because they are not measuring the same. However, it is interesting that the variables that ranked the highest in the variable importance plot either have a very small effect in the linear regression model or are not significant.

9. CONCLUSION

We analyse the determinants of the royalty rate for a sample of 1,834 agreements from the RoyaltyRange database with a special focus on contract features, covering several industries, countries and types of licensing. This thesis adds to the scarce empirical literature on royalty rates as the first study analysing this dataset and taking several variables into account that have not been studied before.

For this, we apply to approaches. First, we use regression trees and their extension random forest. This machine learning approach allows me to explore which factors available in the dataset best predict the royalty rates. The variable importance plot derived thereof suggests that the age of an agreement, the contract term and whether a licence is exclusive or not best predict the royalty rate in our sample.

Second, we apply an ordinary least squares (OLS) regression model. The variables selected are not only derived from economic theory, but also take the results of the machine learning approach into account. Hence, the variables ranking highest in the variable importance plot are included in the linear regression model, independent of whether or not this is supported by economic theory.

We find that a global grant (instead of a restricted geographical scope), related parties, licensors who are universities, non-profit entities or individuals and the age of the agreement, the age of the licensor at the time of the agreement stipulation and the age difference between the licensor and the licensee are associated negatively with the royalty rate. Our findings provide no evidence for a significant influence of exclusivity and contract duration. These results can help in the determination of “reasonable royalties” in patent-infringement lawsuits and to approximate arms-length prices for transfer pricing purposes.

The conclusions above are subject to a number of limitations. First, we only analyse the royalty rates as a price for the licence. Other payments in addition to an ongoing royalty rate, such as upfront payments, milestone payments, minimum fees or other fixed fees, are not taken into account, which might create a distortion. Second, we only observe ex-ante royalty rates, not ex-post royalty rates. Third, the agreements analysed are publicly available. However, one can conjecture that there is a systematic difference between agreements that are publicly available and the ones, which are not. In addition to this, the selection bias could be further increased in the selection of which of the publicly available agreements are accepted in the RoyaltyRate Database.

It would be interesting to investigate further why a global scope is associated negatively with the royalty rate, since it is in line with other empirical studies, but contradicts economic theory. In addition, further analysis on what additional factors can explain variation in the royalty rate should be encouraged.

APPENDIX

Description of the NACE Statistical Classification of Economic Activities (European Communities, 2008)

A	Agriculture, Forestry and Fishing
B	Mining and Quarrying
C	Manufacturing
D	Electricity, Gas, Steam and Air Conditioning Supply
E	Water Supply; Sewerage, Waste Management and Remediation Activities
F	Construction
G	Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles
H	Transportation and Storage
I	Accommodation and Food Service Activities
J	Information and Communication
K	Financial and Insurance Activities
L	Real Estate Activities
M	Professional, Scientific and Technical Activities
N	Administrative and Support Service Activities
O	Public Administration and Defence; Compulsory Social Security
P	Education
Q	Human Health and Social Work Activities
R	Arts, Entertainment and Recreation
S	Other Service Activities
T	Activities of Households as Employers; Undifferentiate Goods and Services Producing Activities of Households for Own Use
U	Activities of Extraterritorial Organisations and Bodies

Bibliography

Athey, S. and Imbens, G.W., 2019. Machine learning methods that economists should know about. Annual Review of Economics, 11.

Breiman, L., 2001. Random forests. Machine learning, 45(1), pp.5-32.

Breiman, L., Friedman, J., Stone, C.J., and A Olshen.R., 1984. Classication and Regression Trees. CRC press.

Dawson, P., 2013. Royalty Rate Determination. Journal of Business Valuation and Economic Loss Analysis, 8(1), pp.133-161.

European Communities, 2008. NACE Rev. 2. Statistical classification of economic activities in the European Community.

Evans, D.S. and Layne-Farrar, A., 2004. Software patents and open source: the battle over intellectual property rights. Va. JL & Tech., 9, p.1.

Fosfuri, A., 2006. The licensing dilemma: understanding the determinants of the rate of technology licensing. Strategic Management Journal, 27(12), pp.1141-1158.

Gordanier, J. and Miao, C.H., 2011. On the duration of technology licensing. International Journal of Industrial Organization, 29(6), pp.755-765.

Grandstrand, O., 1999. Patents and Intellectual Property: a general framework. The Economics and Management of Intellectual Property: Towards Intellectual Capitalism, pp.55-106.

Hegde, D., 2014. Tacit knowledge and the structure of license contracts: Evidence from the biomedical industry. Journal of Economics & Management Strategy 23(3), 568–600.

Jayachandran, S., Kaufman, P., Kumar, V. and Hewett, K., 2013. Brand licensing: what drives royalty rates?. Journal of Marketing, 77(5), pp.108-122.

Kankanhalli, G. and Kwan, A., 2018. An Empirical Analysis of Bargaining Power in Licensing Contract Terms.

OCED, 2007. Compendium of OECD work on intellectual property.

OECD, 2006. Report on the attribution of profits to permanent establishments, Parts I (general considerations), II (banks) and III (global trading). Cent. Tax Policy Adm.

OECD, 2017. Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations.

Qian, Y., 2007. Do national patent laws stimulate domestic innovation in a global patenting environment? A cross-country analysis of pharmaceutical patent protection, 1978–2002. The Review of Economics and Statistics, 89(3), pp.436-453.

Royaltyrange.com. (2019). About us | RoyaltyRange. [online] Available at: https://www.royaltyrange.com/home/royalty-rate-database/about-us [Accessed 06 April 2019].

Sakakibara, M., 2010. An empirical analysis of pricing in patent licensing contracts. Industrial and Corporate Change, 19(3), pp.927-945.

US Department of Commerce, 2016. Intellectual Property and the U.S. Economy: 2016 Update. Available at: https://www.uspto.gov/sites/default/files/documents/IPandtheUSEconomySept2016.pdf.

Verlinden, I. and Bakker, A., 2018. Mastering the IP life cycle from a legal, tax and accounting perspective.