Valuing Patents with Linear Regression

IDENTIFYING VALUE INDICATORS AND USING A LINEAR REGRESSION MODEL TO VALUE PATENTS

ADNAN AL-KHALAF, STEVE OSKAR GUSTAFSSON
REGRESSION ANALYSIS OF PRICING

Identifying value indicators and using a linear regression model to value patents

ADNAN AL-KHALAF
STEVE OSKAR GUSTAFSSON

Degree Project in Applied Mathematics and Industrial Economics (15 credits)
Degree Progr. in Industrial Engineering and Management (300 credits)
Royal Institute of Technology year 2015
Supervisors at KTH: Henrik Hult and Anna Jerbrant
Examiner: Boualem Djehiche

TRITA-MAT-K 2015:16
ISRN-KTH/MAT/K--15/16--SE
ACKNOWLEDGMENTS

Special thanks to Dimitris Giannoccaro & Falah Hosini at IAMIP Sverige AB for the tips and guidance. Thanks to Victor Rudolf at IAMIP Sverige AB & Trent Smith for taking the time and providing us with much appreciated help and information. Many thanks to Henrik Hult & Anna Jerbrant for much needed feedback. A big thank you to the person who provided us with data-set who wish to be anonymous.
This thesis consists of two parts. The first part of the thesis will conduct a multiple regression on a data-set obtained from the Ocean Tomo’s auction results between 2006 to 2008 with the purpose to identify key value indicators and investigate to what extent it is possible to predict the value of a patent. The final regression model consists of the following covariates: Average number of citations per year, share of active family members, age of the patent, average invested USD per year, and nine CPC’s as dummy variables. The second part of the thesis will investigate why it is difficult to value a patent and the different factors and changes that have contributed to a growing importance of patent valuation by applying theories from knowledge-based economy and industrial change. This is done by conducting a literature review and interviews. The results of this thesis state that it is only possible to construct a model that has an explanation degree of 50.21%. The complexity of a patent's value derives from uncertainties about future context of the patent and non-quantifiable parameters of the patent. Furthermore, we find evidence of a shift from tangible assets to intangible assets in industrial nations which motivates the growing importance of patent valuation.
## Contents

List of Tables vi

List of Figures vii

1 **Introduction** 1
   1.1 Background .................................................. 1
   1.2 Purpose and Aim ............................................. 1
   1.3 Problem Definition ......................................... 2
   1.4 Structure of thesis ......................................... 2

2 **Theoretical Framework** 3
   2.1 Regression Analysis ........................................ 3
      2.1.1 Definition .............................................. 3
      2.1.2 Ordinary Least Square - *OLS* .......................... 5
   2.2 Errors ...................................................... 7
      2.2.1 Multicollinearity ..................................... 7
      2.2.2 Homoscedasticity & Heteroscedasticity ..................... 9
   2.3 Improvement modification methods ........................... 10
      2.3.1 *R*\(^2\) and Adjusted *R*\(^2\) ....................... 10
      2.3.2 Hypothesis Testing .................................... 11
      2.3.3 *F*-statistics & *p*-value ............................. 12
      2.3.4 Akaike Information Criterion - AIC .................... 14
      2.3.5 Residual Analysis ..................................... 21
      2.3.6 Cross-Validation ...................................... 16

3 **Method** 17
   3.1 Empirical Approach ......................................... 17
      3.1.1 Ocean Tomo ............................................. 17
      3.1.2 Dataset ................................................ 17
   3.2 Variable Selection .......................................... 19
      3.2.1 Excluded Variables .................................... 19
   3.3 Final Model ................................................. 21
   3.4 Model Verification .......................................... 22
      3.4.1 Linearity between Response Variable & Covariates .... 22
      3.4.2 Expected value of residual is zero ................. 24
      3.4.3 Homoscedasticity ...................................... 25
      3.4.4 Measurement errors .................................... 26
      3.4.5 Multicollinearity ..................................... 27

4 **Results** 28
   4.1 Final Model Validation ...................................... 28
      4.1.1 Cross validation ....................................... 28
      4.1.2 Residual analysis ...................................... 30
      4.1.3 *R*\(^2\), *F*-statistic & *p*-value ..................... 30
   4.2 Prediction equation ........................................ 31
**List of Tables**

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Step by step presentation of excluded variables</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>Result outputs of VIF-test</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>Regression results of the finale model with White's robust estimates</td>
<td>28</td>
</tr>
<tr>
<td>4</td>
<td>Comparison between observed and predicted values of 10 randomly selected patents</td>
<td>39</td>
</tr>
<tr>
<td>5</td>
<td>White's robust coefficient covariance matrix estimate</td>
<td>46</td>
</tr>
<tr>
<td>6</td>
<td>Coefficient covariance matrix estimate</td>
<td>46</td>
</tr>
</tbody>
</table>
**List of Figures**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Scatter plot illustrating positive multicollinearity between two covariates.</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Scatter plot illustrating homoscedasticity and heteroscedasticity</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>F-distributions where $df_1$ is degrees of freedom in numerator and $df_2$ is degrees of freedom in denominator.</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>Scatter plot of residuals vs. fitted values for residual analysis</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>Histogram of the residuals</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>Scatter plot showing relationship between <em>Price &amp; Nr. of citings</em>.</td>
<td>22</td>
</tr>
<tr>
<td>7</td>
<td>Scatter plot showing relationship between <em>Price &amp; Average invested USD/year</em>.</td>
<td>23</td>
</tr>
<tr>
<td>8</td>
<td>Scatter plot showing relationship between <em>Price &amp; share of active family members</em></td>
<td>23</td>
</tr>
<tr>
<td>9</td>
<td>Scatter plot showing relationship between <em>Price &amp; share of active family members</em></td>
<td>24</td>
</tr>
<tr>
<td>10</td>
<td>Histogram of residuals</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>Number of citings plotted against regression residuals.</td>
<td>26</td>
</tr>
<tr>
<td>12</td>
<td>Average USD invested per year plotted against regression residuals.</td>
<td>26</td>
</tr>
<tr>
<td>13</td>
<td>Age of patent plotted against regressions residuals</td>
<td>26</td>
</tr>
<tr>
<td>14</td>
<td>Share of active family members plotted against regression residuals</td>
<td>26</td>
</tr>
<tr>
<td>15</td>
<td>Scatter plot illustrating cross-validation of the model</td>
<td>29</td>
</tr>
<tr>
<td>16</td>
<td>Normal QQ plot - Showing residual normality</td>
<td>30</td>
</tr>
<tr>
<td>17</td>
<td>Scatter plot showing relationship between log<em>Price &amp; fitted price</em></td>
<td>31</td>
</tr>
<tr>
<td>18</td>
<td>Mean logged price for every age</td>
<td>34</td>
</tr>
<tr>
<td>19</td>
<td>Patent filing activity over time</td>
<td>41</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Background

Between January 2013 and December 2013 [PatBase] patents exchanged ownership in the US alone, yet there exists no definite monetary valuation method. This can partly be explained by the lack of transparency in the patent market, which led Schankerman & Pakes to state that patents private value "is in general unobserved" [M. Schankerman & A. Pakes]. In the article New sources of growth: intangible assets conducted by OECD it is stated that investments in intangible assets are growing, and even surpassing investments in traditional capital. Bearing this in mind, and the number of patents that exchanged ownership in the US in 2013, one quickly realizes the importance of patent valuation in today’s technical society.

There have been many studies conducted throughout the years trying to find methods in valuating patents. In 1998 Schankerman looked at patent value by technology field and nationality, however this type of study leaves out a lot of variables that could further explain the true value of a patent. Furthermore, studies including surveys have also been conducted in order to value patents, Harhoff & Verspagen [D. Harhoff & B. Verspagen 2005] conducted such a study considering various characteristics of a patent. However such studies retrieve subjective patent values [J. Bessen].

In the paper Protecting their Intellectual Assets: Appropriability Conditions and why U.S. Manufacturing Firms Patent (or not) conducted by Cohen, Nelson & Walsh published in 2000 it is suggested that "firms can profit from patents in ways other than protecting the profits that may directly accrue to the commercialization or sale of a patented innovation". Furthermore their findings suggest that strategic reasons are the basis for patenting when considering US manufacturing firms.

1.2 Purpose and Aim

Industrial R&D is widely seen as a key driver in productivity and economic growth. Firms can and do use various methods to protect their innovations, including patents, secrecy and licensing agreements [W. Cohen & R. Nelson & J. Walsh]. To what extent firms lay value in these methods has over years shifted.

In this thesis we will investigate the importance of patent valuation by applying theories from knowledge-based economy and industrial change. This is done by conducting a literature review and interviews with Dimitris Giannoccaro CEO of IAMIP Sverige AB and Trent Smith, Chief IP Officer at Tobii AB.

Furthermore, a major part of intellectual properties (IP) is the protection of innovation where many believe that patent rights are essential to the return to invention and are consequently a key inducement to R&D. However, no precise
or commonly agreed approach on monetary patent valuation exist [T. Fisher & J. Leidinger], as mentioned in the introduction. Moreover a mathematical approach will be used to analyze and predict the value of a patent based on the data set at hand. A multiple regression model will be used with the following covariates:

a. Average citings per year
b. Share of active family members
c. Patent age
d. Average invested per year
e. CPC

1.3 Problem Definition
The problem investigated in this thesis can be divided into two parts:

1. Which covariats are important when valuing a patent and to what extent is it possible to predict the true value of a patent?

2. What factors and changes can be identified that contributes to a growing importance of patent valuation and why is it difficult to value patents?

1.4 Structure of thesis
The thesis is divided into three parts:

1. The first part will explain the theory behind the multiple regression model.

2. The second part will apply the theory explained in the first part in order to construct a mathematical based prediction model for patent valuation. Regression on patent sale prices will be performed.

3. Finally interviews and a literature review will be conducted in order to answer the second problem definition.
2 Theoretical Framework

In this section we will introduce the necessary theory that is used throughout this thesis.

2.1 Regression Analysis

Linear regression is a commonly used method when one want to relate a response variable to values of one or more explanatory variables, also called covariates [H. Lang]. However the covariates do not explain the true value of dependent variable, the difference between the actual observation and the explained value is captured by the residual.

The covariates used in this thesis will be observational, meaning that the data gathered on the covariates are outcomes that are not controlled. Furthermore this thesis will be using the multiple regression model as well as a structural interpretation. In other words more than a single covariate will be used in order to explain the dependent variable where the covariates will be considered to influence the dependent variable.

The multiple regression model is valid when five assumptions are met, which in turn will motivate the use of the ordinary least square (OLS) estimator.

2.1.1 Definition

The linear regression model is defined as following:

\[
y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + ... + x_{ik} + e_i, \quad i = 1, 2, ..., n \quad (2.1)
\]

Where \(y_i\) is the response variable whose values depends on the covariates \(x_{i1}, x_{i2}, ..., x_{ik}\) plus the normally distributed residual \(e_i\). The parameters \(\beta_j\) are unknown and are to be estimated from data [H. Lang].

For convenience, matrix notation is introduced and equation (2.1) is transformed into

\[
Y = X\beta + e \quad (2.2)
\]

Where
\[ Y_{n,1} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \]

\[ X_{n,(k+1)} = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{pmatrix} \]

\[ \beta_{(k+1),1} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \]

\[ e_{n,1} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \]

The covariate \( x_{i,0} \) is usually assumed to be the constant 1, and \( \beta_0 \) is the intercept [H. Lang].

### 2.1.1.1 Dummy variables

A dummy variable is used to measure a categorical effect on the response variable that either takes the value one or zero. When the categorical phenomenon occurs the dummy variable takes the value one, and zero otherwise. In regression analysis a dummy variable is used just as any other covariate [P. Kennedy].
2.1.1.2 Logarithmic transformations of variables

Considering the linear model \( y_i = x_i \beta_i + e_i \), there are four possible combinations of transformation involving logarithms:

1. The linear case with no transformation.
2. The linear-log model.
3. The log-linear model.
4. The log-log model.

In this thesis, the log-linear transformation will be used, i.e., \( \log(y_i) = x_i \beta_i + e_i \).

The literal interpretation of the coefficients \( \hat{\beta} \) is that one-unit increase in \( x \) will produce an expected increase in \( \log(y_i) \) of \( \hat{\beta} \) units. In terms of \( y_i \) itself, this means that the expected value of \( y_i \) is multiplied by \( e^{\hat{\beta}} \) [K. Benoit].

Logarithmic transformation of variables in a regression model is often a warranted [H. Lang] operation to use in situations where non-linear relationship exists between the dependent and independent variables. Moreover, logarithmic transformation are also a convenient means of transforming a highly skewed variable into one that is more approximately normal[K. Benoit].

2.1.2 Ordinary Least Square - OLS

Using the linear regression model when solving statistical problems an assumed optimal method for estimating the unknown parameters \( \beta \) is the use of the Ordinary Least Square (OLS) estimator [P. Kennedy].

The estimator uses the values of \( \beta \) that minimizes the sum of the corresponding squared residuals in order to get the estimated value, denoted by \( \hat{\beta} \). This is achieved by solving the normal equations (2.3) for \( \hat{\beta} \) [H. Lang].

\[
\begin{align*}
X^T \hat{e} &= 0 \\
X^T (Y - X \hat{\beta}) &= 0 \\
X^T Y - X^T X \hat{\beta} &= 0 \\
X^T X \hat{\beta} &= X^T Y \\
\end{align*}
\]

Furthermore, the assumptions of the linear regression model (Section 2.1.2.1) can be used to show that the OLS estimator \( \hat{\beta} \) is an unbiased estimator of \( \beta \) [H. Lang]

\[
\begin{align*}
\hat{\beta} &= (X^T X)^{-1} X^T (X \beta + e) = \beta + (X^T X)^{-1} X^T e \\
\hat{\beta} - \beta &= (X^T X)^{-1} X^T e \\
\implies E[\hat{\beta}] &= \beta.
\end{align*}
\]

The covariance matrix for \( \hat{\beta} \) is then computed as

\[
\text{here the natural logarithm is used, where the base is } e \approx 2.71828
\]
\[
\text{Cov} (\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^\top] \\
= (X^\top X)^{-1} X^\top (\sigma^2 I) X (X^\top X)^{-1} \\
= (X^\top X)^{-1} \sigma^2.
\]

(2.5)

2.1.2.1 Assumptions

There are many situations statistical problems can be characterized in and in many of which the OLS estimator is not optimal. The multiple regression model consists of five basic assumptions concerning the way in which the data are generated (P. Kennedy):

1. The model consist of a dependent variable that can be calculated as a linear function of a specific set of independent variables, plus a disturbance term (residual). The unknown parameters of this linear function are assumed to be constants and form the vector \( \hat{\beta} \).

   Violations of this assumption:
   - **Wrong regressors** - the absence of relevant variables or the inclusion of irrelevant variables
   - **Nonlinearity** - when the relationship between the dependent and the independent variables is not linear
   - **Changing parameters** - when the parameters \( \beta \) do not remain constant

2. The expected value of the disturbance term in the model is zero i.e., \( E[e_i] = 0 \) for all \( i \).

   Violation of this assumption leads to a biased intercept [P. Kennedy]

3. The model is such that all of the disturbance terms have the same variance and are not correlated with one another.

   Violations of this assumption:
   - **Heteroscedasticity** - when the disturbance terms do not all have the same variance
   - **Autocorrelated errors** - when the disturbance terms are correlated with one another

4. The observations on the independent variables can be considered fixed in repeated samples, i.e., it is possible to repeat the sample with the same independent variable values.

   Violations of this assumption:
   - **Errors in variables** - errors in measuring the independent variables
   - **Autoregression** - using a lagged value of the dependent variables as an independent variable
5. The number of observation is greater than the number of independent
variables and there is no exact linear relationships between the inde-
pendent variables.
Violation of this assumption associates to the problem of multicolli-
nearity, i.e., two or more independent variables being approximately linearly
correlated.

2.2 Errors
When performing a regression analysis one has to bear in mind potential errors
that could arise which needs to be evaluated in order to ensure that the results
obtained by the defined regression model are valid. This section of the thesis
will explain such errors a multiple regression model could be subject to.

2.2.1 Multicollinearity
Multicollinearity is a phenomenon that arises when there is an approximate
linear relationship between two or more variables in the sample data such
that the variables are linearly dependent. The consequence of this phenom-
ena is large variance in the OLS estimates of the collinear variables. Large
variance makes hypothesis testing obsolete due to inaccurate estimates of the
parameters. Furthermore problems in identifying what parameters effects one
another could lead to specification errors. However multicollinearity is not an
issue when OLS is used for prediction purposes [P. Kennedy].

2.2.1.1 Detecting Multicollinearity
When the estimated standard deviation of the estimated parameters are very
large one is to suspect multicollinearity [H. Lang]. Below three commonly used
methods for detecting multicollinearity will be presented.

Correlation Matrix
The first method for detecting multicollinearity is through the correlation ma-
trix which mathematically is expressed as below

\[ R(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Cov}(X_i, X_i)\text{Cov}(X_j, X_j)}} , \quad i \neq j \] (2.6)

The diagonal entries will all be ones since each variable is perfectly correlated
with itself, and the non-diagonal entries in the matrix will represent the simple
correlation coefficient for the data set. Any correlation coefficient above the
value 0.8 is an indication of high correlation between the variables in question
[P. Kennedy].

Scatter Plot
The second method for detecting multicollinearity is by plotting the measure-
ments of two variables against each other; if there exists multicollinearity be-
tween the two variables then the measurements should follow a straight line as
illustrated by Figure 1 below.

Figure 1: Scatter plot illustrating positive multicollinearity between two co-
variates.

**Variance Information Factor - (VIF)**

Lastly one can calculate the VIF-value of each covariate in order to detect
multicollinearity. A value above 10 is an indication of multicollinearity. The
VIF- value is calculated according to equation (2.7) below [P. Kennedy].

\[
VIF = \frac{1}{1 - R^2} 
\]  

(2.7)

Here \(R^2\) is obtained by doing a regression for each covariate while the rest of
the covariates are independent variables.

### 2.2.1.2 Remedy for Multicollinearity

There is no commonly agreed upon approach for solving multicollinearity since
the reason behind multicollinearity varies from case to case. However the
main aim is to reduce the variances of the estimated coefficients. This can be
achieved in multiple ways, namely;

- *Adding new variables*

- *Obtain more data*

- *Exclude one of the collinear variables* - this choice of solution derives
  another problem, namely biased estimation of the remaining variables.

However there are guidelines on how to treat multicollinearity, namely

a. "If \(R^2\) from the regression exceeds the \(R^2\) of any of the independent
   variable regressed on the other independent variables", then don’t mind
   multicollinearity [P. Kennedy].

b. If the \(t\)-statistics exceeds 2, don’t mind multicollinearity [P. Kennedy].
2.2.2 Homoscedasticity & Heteroscedasticity

For the given data each covariate in the multiple regression model contains an unobserved error term. If all these error terms are the same, the disturbance of the model are said to have a uniform variance or to be *homoscedastic*, i.e., $E[e_i^2] = \sigma^2$ for all $i$ [P. Kennedy].

If the error terms are not all the same, then the disturbances are said to be *heteroscedastic* and the $e_i$ are independent between observations. This causes coefficient estimates $\beta_i$ to be inconsistent [H. Lang].

![Figure 2](image)
Figure 2: Scatter plot illustrating homoscedasticity and heteroscedasticity

Figure 2.b illustrates how heteroscedasticity affects the properties of the OLS estimator. The higher absolute values of the residuals to the right in the graph indicate a positive relationship between the error variance and the covariate. Additional large errors in this graph would tilt the OLS regression line and increase the variation of the line, i.e., the variance of $\hat{\beta}$ will increase [P. Kennedy] making the model more inaccurate.

2.2.2.1 Detecting Heteroscedasticity

In order to attack the problem of heteroscedasticity the first step is to determine whether or not heteroscedasticity actually exists. Two test for this is presented below.

**Visual inspection of residuals**

The residuals are plotted on a graph against the covariate to which the disturbance variance is suspected. If it appears that the absolute magnitude of the residuals is related to the covariate as the value of the covariate increases, then a more formal check for heteroscedasticity is in order [P. Kennedy]. Figure 2 illustrates the difference in graph character between homoscedastic and heteroscedastic disturbance terms. Figure 2.a shows the homoscedasticity type where you can see on average a consistent variance in the disturbance term regardless of the value of the covariate while Figure 2.b shows heteroscedastic disturbances.
The White test

If it appears from the visual inspection of residuals that there exists het-
eroscedasticity in the error terms a comparison of the OLS covariance matrix
with the covariance matrix computed by White’s Consistent Variance Estima-
tor could determine the existence. White’s estimator is a consistent estimator
for the covariance matrix for the heteroscedastic regression model [H. Lang].
It is described as

\[
\text{Cov}(\hat{\beta}) = (X^TX)^{-1}X^T D(\hat{e}^2) X (X^TX)^{-1}
= (X^TX)^{-1}X^T \left( \sum_{i=1}^{n} x_i^t x_i \right) X (X^TX)^{-1}
\]

(2.8)

where \(D(\hat{e}^2)\) is the \(n \times n\) diagonal matrix whose \(i\):th diagonal element is \(\hat{e}_i^2\). To
make the covariance matrix estimator (2.8) more robust it is a good practise
to scale it by \(N/(N - K - 1)\) [H. Lang]

\[
\text{Cov}(\hat{\beta})_{\text{scaled}} = \text{Cov}(\hat{\beta}) \cdot \frac{N}{N - K - 1}.
\]

(2.9)

If there is no difference between the OLS and White’s covariances matrices
there does not exist heteroscedasticity.

2.2.2.2 Remedy for Heteroscedasticity

A reformulation of the model by redefining, include or exclude relevant co-
variates may help correct for heteroscedasticity. If heteroscedasticity is still
present the employment of White’s Consistent Variance Estimator in the re-
gression model will improve the estimates without altering the coefficients.

2.3 Improvement modification methods

In this Section statistical tools will be presented that will be used in chapter
3. The purpose of the tools is mainly to examine to what extent the regres-
sion model represents the data. This is especially important when building a
predictive model.

2.3.1 \(R^2\) and Adjusted \(R^2\)

\(R^2\) is used as a measure of regression fit, i.e., measures how well the observed
outcomes are replicated by the regression model. Mathematically described as
[B. Hansen]:

\[
R^2 = 1 - \frac{\sum_{j=1}^{N} \hat{e}_j^2}{\sum_{j=1}^{N} (y_j - \hat{y})^2} = 1 - \frac{\hat{\sigma}^2}{\sigma_y^2}
\]

(2.10)

\(^2\)Inclusion of new relevant covariates does not always render in a more explanatory model.
An AIC test should be conducted when adding more covariates to the model.
A high $R^2$ value indicates a good model where the coefficients explain the true value $y$ to a large extent, which in turn implies a minimized error term $\hat{e}$.

However $\hat{\sigma}_y^2$ and $\hat{\sigma}_y^2$ are biased estimates [B. Hansen] and is solved by adjusting the degrees of freedom such that the adjusted $R^2$, also denoted $\bar{R}^2$ is slightly lower than $R^2$ [H. Lang].

$$\bar{R}^2 = 1 - \frac{(N - 1) \sum_{j=1}^{N} \hat{e}_j^2}{(N - K) \sum_{j=1}^{N} (y_j - \bar{y})^2}$$

Adding an extra covariate to the regression model means a change in degrees of freedom. If the extra covariate accounts for very little of the unexplained variation in the response variable $\bar{R}^2$ will drop. Therefore one should only consider adding covariates if $\bar{R}^2$ raises [P. Kennedy].

### 2.3.2 Hypothesis Testing

Hypothesis testing is a method used in statistics in order to test claims or hypothesis of an unknown parameter. This is done by testing the likelihood (also referred to as p-value) that a parameter is true [H. Lang].

The process of hypothesis testing can be divided into four steps:

1. Firstly identify a null-hypothesis ($H_0$) and an alternative hypothesis ($H_\alpha$). The null-hypothesis suggests that the sample observation results purely from chance. The alternative hypothesis suggests that the sample observations are derived from some cause.

2. Identify a test statistics that will assess whether $H_0$ is true or not.

3. Calculate the p-value. A small p-value is evidence against the null-hypothesis.

4. Finally compare the obtained p-value in step 3 to the predefined acceptable significance level $\alpha$. If $p$-value $\leq \alpha$ one rejects $H_0$ and $H_\alpha$ is valid.
2.3.3 F-statistics & p-value

In regression analysis F-statistics is a popular method to test the significance of a covariate to the response variable. The key here is to compute the p-value by finding the F-statistics. The F-statistic follows the F-distribution and is calculated according to the equation below [R. Williams]:

\[
F = \frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2}{K} \left( \frac{\sum_{i=1}^{N} (y_i - \hat{y})^2}{N-K-1} \right)
\]  

(2.12)

Where

\( \hat{y}_j \) is the estimated value  
\( y_j \) is the observed value  
\( \bar{y} \) is the mean value of the response variable  
\( N \) is the number of observations  
\( K \) is the number of covariates

In order to obtain the p-value one compares the F-statistic with the area under the F-distribution which is illustrated in the figure below.

---

\(^3\)Under the assumption that the error terms are normally distributed.
This implies that a large F-statistic gives a small p-value. When using the p-value for hypothesis testing, the null hypothesis is defined such that the coefficient equals to zero and the alternative hypothesis suggest the opposite, namely that the coefficient is not zero. Mathematically:

\[ H_0 : \beta_j = 0 \]

\[ H_\alpha : \beta_j \neq 0 \]

There is also the possibility to test whether \( R \) number of covariates are equal to zero and hence have no significance on the dependent variable:

\[ H_0 : \beta_j = \beta_{j+1} = \beta_{j+2} = \cdots = 0 \]

This can be achieved by the F-test according to the following test\(^4\):

\[ F = \frac{1}{R} \frac{\hat{e}_*^2 - |\hat{e}|^2}{s^2} = \frac{N - K - 1}{R} \left( \frac{|\hat{e}_*|^2}{|\hat{e}|^2} - 1 \right) \]  
\[ (2.13) \]

\(^4\)Again, under the assumption that the error terms are normally distributed.
where $\hat{e}$ are the residuals for the unrestricted regression and $\hat{e}^*$ are the residuals for the restricted regression, i.e., the corresponding covariates are left out of the regression. $F$ has an $F(R, N - K - 1)$ distribution under the null hypothesis and $H_0$ is rejected if $F$ is large [H. Lang].

Another useful test that can be used when $H_0 = 0$ for all $\beta$’s except possibly the intercept is presented in equation (2.14) [H. Lang]. The value is treated and interpreted just as described above for testing a single covariate.

$$F = \frac{N - K - 1}{K} \frac{R^2}{1 - R^2}$$  \hfill (2.14)

### 2.3.4 Akaike Information Criterion - AIC

With the goal of improving the model it is common to add more significant covariates in order to get a more explanatory model. Unfortunately, a covariate already included in the model may have its usefulness negated by inclusion of new covariates [P. Kennedy].

A common test to check if a covariates should enter the model or not is the use of the Akaike Information Criterion (AIC). The AIC value of the model are calculated as

$$AIC = n \cdot \ln(|\hat{e}^2|) + 2k$$

where $k$ is the number of covariates and $n$ the number of observations [H. Lang].

The test suggests that one should use the model that minimizes the AIC value.

### 2.3.5 Residual Analysis

A residual analysis are conducted on the regression model to investigate violation of the second assumption stating that the expected value of the disturbance term is zero. In practical problems this assumption is commonly not possible, but the analysis is used as a further check on the adequacy of the regression model. What type of violation and to what extent it has can indicate on how to improve the model.

A good way to check for this is the use of the residual vs fit plot. It is a scatter plot with residuals on the y-axis and fitted values (estimated responses) on the x-axis, and is used to detect non-linearity and unequal error variances [H. Seltman]. An example of this plot is illustrated in Figure 4

---

5 Including the intercept, $\beta_0$. 

---

14
In Figure 4 the points appear randomly scattered around zero allowing us to assume it is reasonable that the error terms have a mean of zero, i.e., $E[e^2] = 0$. Furthermore, the vertical width of the scatter does not appear to increase or decrease across the fitted values allowing us to assume it is reasonable that the variance in the error terms is constant.

A further check is the use of the histogram of the residuals. Figure 5 shows an example of such table, the distribution of the residuals is illustrated and if it is normally distributed around zero the assumption can be regarded as valid.

![Histogram of residuals](image)

Figure 5: Histogram of the residuals
2.3.6 Cross-Validation

A major tenet of conventional statistics is that a regression model provides a much more optimistic explanation on the set of data that was used in its derivation than it does of other data provided in similar fashion [H. Sauerbrei]. A often recommended methodology for the assessment of the predictive ability of a regression models is the use of cross-validation of a model [P. Cook]. In cross-validation, the procedure can be done in different ways but with the common purpose of testing the model on a set of data that has not been used when deriving the model.

One way of using cross-validation is to separate the data at hand into two sets, called the training set and the testing set. The model estimates the parameters $\beta$ using the training set only. Then the model uses these estimates and is asked to predict the output values for the data in the testing set. If the model is able to predicts the output of the testing set, the model are said to be cross validated.
3 Method

This chapter will describe the quantitative methodological approach of this thesis. First a description of how the data is collected, treated and complemented is presented. Secondly the prediction model for pricing patents is determined and examined. Lastly the assumptions described in section 2.1.2.1 will be examined.

It is also worth mentioning that the multiple regression model is the method that will be used in order to determine the results of this thesis.

3.1 Empirical Approach

3.1.1 Ocean Tomo

This thesis will make use of Ocean Tomo’s patent auction results from spring 2006 to fall 2008. Between that period Ocean Tomo held a total of 8 auctions and a total of 554 patents were sold in a total of 154 deals. Every auction follows the same structure: First date and location are set and announced, secondly sellers and patents are registered, followed by registration of potential bidders. Finally the auction itself takes place.

Every patent deal consist of at least one U.S. patent, and each deal is offered by one seller. The sellers have the opportunity to set a minimum price, if nothing is specified by the seller, Ocean Tomo applies a reserve price of $10,000. If the highest bid does not meet the reserve price, no transaction will take place and the deal will remain unsold.

In 2009 Ocean Tomo sold its auction business to ICAP for $10,000,000 [J. Wild] and ICAP has chosen not to disclose any auction results, including the results published by Ocean Tomo prior to the acquisition. Therefore this thesis has been limited to the first eight auction results previously published by Ocean Tomo.

3.1.2 Dataset

The original dataset received consisted of five variables: deal number, sale price, expected value/asking price, assignee and finally U.S. patent number. U.S. patent number and sale price are the two variables that where required for the purpose of this thesis. By identifying the U.S. patent number more data and variables could be identified and retrieved for each patent. This was possible by the use of three patent tools that will be described in the next section.

From this dataset one variable will be included in the regression:

---

6 Only sold deals will be examined in this thesis.
7 Based on the dataset at hand.
1. **Price**, the auction result of the patent in question. Also the response variable.

3.1.2.1 **P\(^2\)ALS, PatBase & Espacenet**

**P\(^2\)ALS**

P\(^2\)ALS is a business tool offered by IAMIP Sverige AB that was used to create multiple patent portfolios for each auction result. P\(^2\)ALS allowed for the retrieval of 10 variables:

2. **Citing**, the number of references the patent in question has received from other patents.

3. **Number of family members**, a patent family could consist of multiple family members. Each family member represents a country and protects the exact same innovation in that country.

4. **Number of active family members**, here only members that where active at the time of the auction are regarded as active.

5. **Dummy for granted in Europe when the auction took place**, 0 if patent is not granted in Europe, 1 if patent is granted in Europe.

6. **Dummy for granted in Europe before auction**.

7. **Years before granted in EP**, the number of years before a patent family member is granted in Europe.

8. **Number of Cooperative Patent Classification (CPC)**, a classification system that has been jointly developed by the U.S. and European patent offices. Only the first four letters are considered.

9. **Dummy for each CPC class**, up to three letters.

10. **Age**, the age of the patent from priority date to the date of the auction.

11. **Number of inventors**

12. **Average USD invested per year**.

**Espacenet**

The European Patent Office (EPO) offers a search tool through their website that enabled for the retrieval of the following variable:

13. **Number of independent claims**, are claims that directs to the essentials of the innovation.

**PatBase**

PatBase is another business tool that was used in order to retrieve information regarding whether a patent was subject to a lawsuit or not.

14. **Patent subject to a lawsuit**
3.2 Variable Selection

The dataset at hand was limited and consisted of three useful variables which was not enough for building a regression model. However the dataset in question allowed for further gathering of variables through various tools as described on page 18.

The variable selection was discussed and agreed upon along with IAMIP’s patent experts after meetings and initially all 13 variables were included in the model. Some variables where included due to curiosity and intuitive relation and some due to expected impact on the price based on years of knowledge from IAMIP’s experts. The variables that where included due to curiosity allowed for experimentation and also helped minimize the risk that one or more variables of importance to the model where overlooked.

A step by step process was applied when choosing the optimal model (the initial model consisted of all 14 variables) and the following rules of thumb where applied:

1. Any covariate or dummy variable with a $p$ – value greater than 10% is excluded from the model.

2. If the exclusion of a covariate or a dummy variable from the tested model means a lower AIC-value, the new model is preferred.

3. If the excluded covariates lead to a higher $\bar{R}^2$ they are to remain excluded.

Furthermore interaction effect between covariates where examined, and the above rule of thumbs where applied here as well.

3.2.1 Excluded Variables

In Table 1 below is a summary of the excluded variables:

---

8 As described in section 3.1.2
<table>
<thead>
<tr>
<th>Excluded Variables</th>
<th>Variable</th>
<th>Unit</th>
<th>p-value</th>
<th>$R^2$</th>
<th>AIC</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CPC F16*</td>
<td>Dummy</td>
<td>0.949816</td>
<td>0.4896</td>
<td>74.89148</td>
<td>Excluded first due to highest $p$-value</td>
</tr>
<tr>
<td></td>
<td>CPC B29*</td>
<td>Dummy</td>
<td>0.866105</td>
<td>0.491</td>
<td>72.89583</td>
<td>Excluded next due to highest $p$-value, lower AIC &amp; higher $R^2$</td>
</tr>
<tr>
<td></td>
<td>Independent Claims</td>
<td>Numeric</td>
<td>0.750398</td>
<td>0.4924</td>
<td>70.92696</td>
<td>Excluded next due to highest $p$-value, lower AIC &amp; higher $R^2$</td>
</tr>
<tr>
<td></td>
<td>CPC H02*</td>
<td>Dummy</td>
<td>0.75832</td>
<td>0.4937</td>
<td>69.03746</td>
<td>Excluded next due to highest $p$-value, lower AIC, &amp; higher $R^2$</td>
</tr>
<tr>
<td></td>
<td>CPC B64*</td>
<td>Dummy</td>
<td>0.703826</td>
<td>0.495</td>
<td>67.14055</td>
<td>Excluded next due to highest $p$-value, lower AIC, &amp; higher $R^2$</td>
</tr>
<tr>
<td></td>
<td>CPC G07*</td>
<td>Dummy</td>
<td>0.637459</td>
<td>0.4962</td>
<td>65.29747</td>
<td>Excluded next due to highest $p$-value, lower AIC, &amp; higher $R^2$</td>
</tr>
<tr>
<td></td>
<td>Granted in Europe at time of auction</td>
<td>Dummy</td>
<td>0.624617</td>
<td>0.4973</td>
<td>63.53793</td>
<td>Excluded next due to highest $p$-value, lower AIC, &amp; higher $R^2$</td>
</tr>
<tr>
<td></td>
<td>CPC H03*</td>
<td>Dummy</td>
<td>0.646509</td>
<td>0.4983</td>
<td>61.79644</td>
<td>Excluded next due to highest $p$-value, lower AIC, &amp; higher $R^2$</td>
</tr>
<tr>
<td></td>
<td>CPC H01*</td>
<td>Dummy</td>
<td>0.591976</td>
<td>0.4994</td>
<td>60.02291</td>
<td>Excluded next due to highest $p$-value, lower AIC, &amp; higher $R^2$</td>
</tr>
<tr>
<td></td>
<td>CPC G02*</td>
<td>Dummy</td>
<td>0.553874</td>
<td>0.5004</td>
<td>58.33135</td>
<td>Excluded next due to highest $p$-value, lower AIC, &amp; higher $R^2$</td>
</tr>
<tr>
<td></td>
<td>CPC B41*</td>
<td>Dummy</td>
<td>0.613429</td>
<td>0.5013</td>
<td>56.70656</td>
<td>Excluded next due to highest $p$-value, lower AIC, &amp; higher $R^2$</td>
</tr>
<tr>
<td></td>
<td>CPC Y10*</td>
<td>Dummy</td>
<td>0.429052</td>
<td>0.5023</td>
<td>54.97907</td>
<td>Excluded next due to highest $p$-value, lower AIC, &amp; higher $R^2$</td>
</tr>
<tr>
<td>Years before granted in Europe</td>
<td>Years</td>
<td>0.444344</td>
<td>0.5028</td>
<td>53.645</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Granted in Europe</td>
<td>Dummy</td>
<td>0.594538</td>
<td>0.5034</td>
<td>52.26625</td>
<td>Excluded next due to highest $p$-value, lower AIC, &amp; higher $R^2$</td>
</tr>
</tbody>
</table>
### Table 1: Step by step presentation of excluded variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>p-value</th>
<th>$\bar{R}^2$</th>
<th>AIC</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPC G10* Dummy</td>
<td></td>
<td>0.402088</td>
<td>0.5044</td>
<td>50.56627</td>
<td>Excluded next due to highest $p$-value, lower AIC, &amp; higher $\bar{R}^2$</td>
</tr>
<tr>
<td>Nr. of Numeric</td>
<td>CPC value</td>
<td>0.350635</td>
<td>0.5048</td>
<td>49.30772</td>
<td>Excluded next due to highest $p$-value, lower AIC, &amp; higher $\bar{R}^2$</td>
</tr>
<tr>
<td>CPC G08* Dummy</td>
<td></td>
<td>0.254229</td>
<td>0.505</td>
<td>48.22523</td>
<td>Excluded next due to highest $p$-value, lower AIC, &amp; higher $\bar{R}^2$</td>
</tr>
<tr>
<td>Age²</td>
<td>Years²</td>
<td>0.250338</td>
<td>0.5045</td>
<td>47.59059</td>
<td>Excluded next due to highest $p$-value &amp; lower AIC</td>
</tr>
<tr>
<td>CPC G11* Dummy</td>
<td></td>
<td>0.206183</td>
<td>0.5041</td>
<td>46.97475</td>
<td>Excluded next due to highest $p$-value &amp; lower AIC</td>
</tr>
<tr>
<td>Nr. of inventors</td>
<td>value</td>
<td>0.166843</td>
<td>0.5033</td>
<td>46.64414</td>
<td>Excluded next due to highest $p$-value &amp; lower AIC</td>
</tr>
</tbody>
</table>

#### 3.3 Final Model

After testing multiple interaction effects through multiple regressions the final model is established and will consist of the following covariates and dummy variables:

1. Average number of citings per year
2. Average USD invested per year
3. Age of patent (at the time of the auction)
4. Share of active family members
5. Dummy for CPC B62*
6. Dummy for CPC C08*
7. Dummy for CPC C11*
8. Dummy for CPC F02*
9. Dummy for CPC G06*
10. Dummy for CPC G09*
11. Dummy for CPC G11*
Furthermore the response variable Price is logged mainly to capture outliers.

3.4 Model Verification

In Section 2 five assumptions where presented that are of essence for the OLS-estimator to be optimal. This section will examine whether the final model violates any of these assumptions.

3.4.1 Linearity between Response Variable & Covariates

The first assumption is that there is linearity between the response variable (price) and the covariates. This can be examined graphically through four scatter plots, the first between price & average number of citings per year, the second between price & average USD invested per year, the third between price & share of active family members and lastly between price & age of patent. If there exists a linear relationship between the dependent variable and the covariates then the first assumption is not violated. Below are four scatter plots that presents the relationship between price and the mentioned covariates.

Figure 6: Scatter plot showing relationship between Price & Nr. of citings.

As illustrated in Figure 6, there exists a linear relationship between price and citings. No violation of this assumption is observed.
As illustrated in Figure 7 there is a linear relationship between the independent variable and the covariate in question, therefore no violation of the assumption is observed.

In Figure 8 it is clear that there exists a linear relationship between Price and Share of active family members based on the data-set used in this thesis therefore the assumptions is regarded as valid.
Lastly, uncertain linearity is observed in Figure 9 based on the data-set used. However, we argue that there is a linearity and this assumption is regarded as valid for this model.

### 3.4.2 Expected value of residual is zero

To evaluate the assumption whether the residual of the model has an expected value of zero, the use a histogram of the residuals are plotted and examined.
In Figure 10 we can see that the residuals are normally distributed around zero. Hence the expected value is zero and the assumption can be regarded as valid.

### 3.4.3 Homoskedasticity

The third assumption states that the residuals should have the same variance and should not be correlated. This can be examined by generating scatter plots of the covariates in the model and the residuals of the regression. Four scatter plots are generated together with the residuals: *Average number of citings per year*, *Average USD invested per year*, *Age of patent*, *Share of active family members*. 

![Histogram of residuals](image)

Figure 10: Histogram of residuals
When examining the four figures it can not be concluded that the residuals are homoscedastic. It is thus necessary to examine a comparison between the covariance matrix and the covariance matrix computed with the White’s Consistent Variance Estimator. After a comparison of the matrices\textsuperscript{9} we can conclude that they are not equal and that there exist heteroscedasticity, hence an employment of the White’s Consistent Variance Estimator is necessary.

With the employment of White’s estimator the assumption is argued to have been considered and thus can be validated.

3.4.4 Measurement errors

The fourth assumption states that the independent variables can be considered fixed in repeated samples. When there are errors in measuring the independent variables the assumption is violated since these measurement error makes the independent variable stochastic.

The data that are used in this thesis is the former disclosed auction prices of the patents sold by Ocean Tomo and is official actual values. The complemented data is firstly collected from P\textsuperscript{2}ALS, a tool used by IAMIP a company that has its core business in Intellectual Properties and rely on accurate information. PatBase is also a professional business tool offered by Minesoft & RWS Group to patent searchers. Furthermore, Espacenet is an international organization handling patent information that are used worldwide and the data collected from these organizations are regarded valid.

\textsuperscript{9}The two matrices can be found in the appendices.
3.4.5 Multicollinearity

Multicollinearity is an assumption that is violated once there exists a linear relationship between the covariates in the regression model as described in Section 2. By using the VIF-test the following results are obtained:

<table>
<thead>
<tr>
<th>VIF-test</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CitingYear</td>
<td>1.051415</td>
</tr>
<tr>
<td>ShareActivFamMem</td>
<td>1.279378</td>
</tr>
<tr>
<td>AgeAuction</td>
<td>1.126346</td>
</tr>
<tr>
<td>InvestedPYear</td>
<td>1.380128</td>
</tr>
<tr>
<td>CPC*</td>
<td>all under 1.279609</td>
</tr>
</tbody>
</table>

Table 2: Result outputs of VIF-test

The value of the VIF test on the different covariates are to be of 10 or greater if there exists multicollinearity in the model. All the covariates shows a value well below this limit and therefore there exists no multicollinearity in the regression model.
4 Results

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | 10.2247  | 2.760e-01  | 37.052  | < 2e-16  **|
| CitingYear     | 0.0798   | 1.265e-02  | 6.307   | 8.07e-10 **|
| ActivFamMem    | 0.5974   | 2.085e-01  | 2.866   | 0.004399 **|
| AgeAuction     | 0.0292   | 1.563e-02  | 1.866   | 0.062871 . |
| InvestedPYear  | 0.0001   | 1.967e-05  | 3.613   | 0.000345 ***|
| B62*           | -1.7449  | 1.438e-01  | -12.136 | < 2e-16  ***|
| C08*           | -2.5793  | 1.505e-01  | -17.140 | < 2e-16  ***|
| C11*           | -3.6578  | 2.004e-01  | -18.255 | < 2e-16  ***|
| F02*           | -2.9646  | 1.857e-01  | -15.969 | < 2e-16  ***|
| G01*           | -0.6299  | 2.723e-01  | -2.313  | 0.021262 * |
| G06*           | 0.3581   | 1.226e-01  | 2.921   | 0.003702 **|
| G09*           | 0.9093   | 5.640e-01  | 1.612   | 0.107763 |
| H04*           | 0.4343   | 1.209e-01  | 3.592   | 0.000373 ***|
| Y02*           | -3.9813  | 1.899e-01  | -20.962 | < 2e-16  ***|

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.044 on 372 degrees of freedom
Multiple R-squared: 0.5189, Adjusted R-squared: 0.5021
F-statistic: 521.8 on 13 and 372 DF, p-value: < 2.2e-16

Table 3: Regression results of the finale model with White’s robust estimates

4.1 Final Model Validation

In this section the above presented regression model will be validated by first a cross validation, secondly a residual analysis will be performed, and lastly R², F-statistic & p-value will be evaluated.

4.1.1 Cross validation

The data set is separated into two different sets, one training set and one testing set. Using the model a regression on the training set is conducted and used to predict the price on the testing set.
Figure 15: Scatter plot illustrating cross-validation of the model

Figure 15 shows that much of the observed prices does not fall into the 90% interval. The model can not entirely be cross-validated.
4.1.2 Residual analysis

![Normal QQ plot](image)

Figure 16: Normal QQ plot - Showing residual normality

As the figure above illustrates the residuals are normally distributed.

4.1.3 $R^2$, F-statistic & p-value

As can be observed in Table 3 all t-values are less than $-2$ or greater than $2$ except $AgeAuction$ and $G09*$ which confirms the alternative hypothesis ($H_a$) that the covariates are of significance on the price with a confidence interval of 95%. However $AgeAuction$ and $G09*$ are also considered to have a significance on the price since on a 90% significance level as defined in section 3.2. Also these two covariates contribute to a higher $\bar{R}^2$ which is desired in a prediction model. $R^2$ of the final model has a value of 0.5021 which means that the model explains 50.21\% of the variation in the data, as illustrated in Figure 17.
Figure 17: Scatter plot showing relationship between logPrice & fitted price

Furthermore all $p-values$ are less than 5% except AgeAuction and G09* that are kept due to the above mentioned fact; they both contribute to a higher $R^2$.

4.2 Prediction equation

From Table 3 we can define the prediction equation as following:

$$E[Price] = e^{\beta X + 0.5\sigma^2}[H.Lang]$$ (4.1)

where

$$\beta X = 10.2247 + 0.0798 *\text{CitingYear} + 0.5974 *\text{ActivFamMem}$$

$$+ 0.0292 *\text{AgeAuction} + 0.0001 *\text{InvestedPYear}$$

$$- 1.7449 * B62_{0,1} - 2.5793 * C08_{0,1}$$

$$- 3.6578 * C11_{0,1} - 2.9646 * F02_{0,1}$$

$$- 0.6299 * G01_{0,1} - 0.3581 * G06_{0,1}$$

$$- 0.9093 * G09_{0,1} + 0.4343 * H04_{0,1}$$

$$- 3.9813 * Y02_{0,1}$$
Each CPC dummy variable have an index (0,1) showing that these variables only can take on the values 0 if a patent does not belong to a specific CPC or 1 if the patent in question belongs to the specific CPC.

1. CitingYear - are the number of citings a patent receives on average per year.

2. ActivFamMem - is the number of active family members divided by number of family members.

3. AgeAuction - is the age of the patent at the time of the auction. When applying this prediction equation on a patent one takes the patent age.

4. InvestedPYear - is the average amount of USD invested on a patent per year.

5. CPC* are dummies for each CPC class a patent belongs to. CPC class defining the application field for the patent. A patent can belong to more than one CPC class.
   a. B62 – Transporting - Land vehicles for traveling otherwise than on rail
   b. C08 – Chemistry - Organic macromolecular compounds; their preparation or chemical working-up; compositions based thereon
   c. C11 – Chemistry - Animal and vegetable oils, fats, fatty substances and waxes; fatty acids therefrom; detergents; candles
   d. F02 - Engines or pumps - Combustion engines
   e. G01 – Instruments - Measuring; testing
   f. G06 – Instruments - Computing; calculating; counting
   g. G09 - Instruments - Education; cryptography; display; advertising; seals
   h. H04 - Electricity - Electric communication technique
   i. Y02 - Technologies or application for mitigation or adaptation against climate change

The CPC definitions are taken from the Cooperative Patent Classification website [CPC].
5 Discussion

When building a prediction model one wish to replicate observed data to as high degree as possible. A perfect model would replicate each observation such that all data points in Figure 15 would be along the blue line. However such model is impossible to achieve for patent valuation. Although patents follow a generalized template where one can observe and fetch different attributes, the true quality of a patent needs to be analyzed within the text and needs to be put in the right context. Moreover the value of a patent is subjective and therefore very difficult to predict. However we have managed to identify key value indicators and build a model that explains 50.21% of the variation in the observed data. Due to the remaining unexplained variation that is not captured in the defined prediction model, one is to use this model with prudence.

5.1 Indication of Covariates

In this section a discussion regarding the covariates of the final model and its indication from the regression will be presented. When evaluating a covariate from the model, the rest is assumed to be fixed.

Citings per Year

In the examination process of a patent it is documented all the prior art of the patent. This is a search to discover if the patent has a close related content of another patent. The more citings a patent receives, the more inventions are build upon the underlying invention of the focal patent. Thus, citings can be viewed as the technological quality of the patent, which explains the high significance of the covariate. Although the significance is high the model suggests that the number citings does not affect the price as much.

Share of active family members

The covariate share of active family members is also of significance and indicates that the more family members that are active the higher value of a patent. This is expected since each family member protects the innovation in question in a different country meaning that the patent operates in a larger market.

Age of the patent

The age of the patent has surprisingly a lower significance than expected. When plotting the mean price of the patent for each age we observe that the logged price is decreasing after 13 years as can be observed in Figure 18 which motivates the inclusion of the covariate $Age^2$ at the beginning of the regression. However $Age^2$ had no significance on the final model and was removed. Therefore the final model misses to capture decreasing affect on the price as a patent gets older and only accounts for the increasing affect.
Invested USD on the patent
This covariate had a significance level of 99.9% on the final model and indicates that the more one invest in a patent, the more the patent will be worth. However it is worth mentioning that this will not always be the case since the data-set analyzed are biast since only sold patents are part of the regression. Furthermore the data obtained for this covariate are costs of the patents up until 2015 and divided by the number of years the patent had been granted. Costs vary differently depending on the family member and a more accurate measure would have been to look at the costs up until the auction date of each patent family. Unfortunately such data could not be obtained.

CPC*
Since patents concern inventions in different application fields there will be a spread of significance depending on what field it resides in. The model have identified nine fields of significance. Two of these fields concerns patents in the Chemistry industry indicating lower value of patents in corporations operating there. The rest have a different affects within the same industry as can be observed in Table 3.

5.2 Conclusion & Future research
To conclude we are able to predict a patent value using the covariates in Table 3 with an explanation degree of 50.21%. The most important covariate being CPC C11* with an explanation degree of 26.93%. The second most important covariate is the number of citing per year which has an explanation degree of 10.33%. The third most important covariate is another CPC, namely CPC G06 with an explanation degree of 8.491%. The remaining covariates influence the price of a patent to different extent, all below 4 except CPC H04 that has
an explanation degree of 6.174%.

The reason behind the relatively low $R^2$ is the fact that the model misses to capture important covariates that are hard to quantify. Two different patents can have almost identical attributes, however the quality of one patent could be far higher which means a higher patent value. Such covariate that could have a high explanation degree that are hard to quantify are the claims. The claims of a patent are very important and needs to be of high quality in order to consider a patent being valuable. There are numerous of factors that determines a patents value which can not be quantified in the same way as the variables in the model does. The value of a patent is ultimately determined by how much the buying end is prepared to pay. Here subjective aspects of valuation are difficult to capture in an objective model based on deterministic values. Factors that here cannot be quantified as easy may be financial situations of the buying end, technological trends and how a patent aligns with the patentportfolio of a company.

When examining the cross-validation test the model does not succeed to capture the majority of the output in the testing set. Hence, indicating that the predictability of the model is not optimal. This coincides with the low $R^2$ value of 0.5012 obtained from the regression analysis. This results highlights a complexity in patent valuation and that even though there is correlation between the covariates and the price in the model the variance is substantial making the explanatory part of the prediction unreliable.

For future research we suggest the use of up-to-date data since the data is very sensible to changes. From a technology perspective seven years is a long time. Moreover the data collected for this thesis is biased since it origins from an auction and therefore a larger and more random data is required in order to obtain representative results.
6 Growing Importance of Patent Valuation

Up until now, this thesis has focused on finding value indicators of patents and constructing a mathematical based prediction model. As the results illustrates, such a model is very difficult to build due to large unexplained variation of patent values. However we argue that patents are an integral part of every technical corporation due to the technical society we currently live in and will therefore in this section examine what factors and changes are behind the growing importance of patent valuation. Moreover we will discuss the complexity of patent valuation.

6.1 Method

The qualitative method of this section is divided into two part. The first part will consist of two interviews conducted by us where the first interviewee is Dimitris Giannoccaro, CEO of IAMIP Sverige AB. This interview was semi-structured and was based on seven questions that can be found in the appendices. The interview lasted for approximately 20 minutes and was recorded. The second interviewee is Trent Smith who is Chief IP Officer at Tobii AB. This interview was unstructured and discussion led that went on for approximately an hour. Notes were taken during the whole hour.

Secondly, a literature review was conducted where we looked at the importance of patent valuation by identifying key factors that lie behind the growing role of patents. This is done with theories about knowledge-based economy and industrial change. Google scholar and KTHB Primo search tools where used to find relevant articles and literature.

The objective of the literary approach is to gain further understanding and perspective of patent usage and the complexity that derives from them in order to identify why it is important to value patents. Furthermore, a valuation approach of single patents have been derived and considered earlier in this thesis, in this part of the thesis we will look at a broader context and consider patents as a growing industrial function and part of a transition in the economy. Therefore relevant articles in the subjects of patent as part of a knowledge-based economy and industrial and technical transformation have been analyzed.

6.2 Theoretical framework

The way this problem definition have been viewed in this thesis is with theories concerning industrial dynamical transformation. Here aspects regarding Large Technical Systems (LTS) [R. Mayntz & T. Hughes, P. Blomqvist & J. Larsson] and the different functions and mechanisms that plays part in a transformation in industrial functions [F. Geels] are conceptualized. A mutual viewpoint in the discussion of dynamical transformation is the different processes of change that takes part in order for a transformation to persist. Furthermore arguments suggests that these processes has to take part in different areas in society to
cope with the complexity and the broad field that larger technical- and industrial functions affects.

A multi-level perspective for observing and understanding technological transitions in societal functions is brought up in an article written by Geels [F. Geels]. By using this perspective he identifies that there are certain patterns and mechanisms that triggers a technological transformation. These patterns and mechanisms can be observed in three different levels. First level is of the technological niches and refers to a technology of itself as a driver of a transition. In the second level he suggests that sociotechnical (ST) regimes also has a major part in a transition. These regimes can be viewed as different social groups other than the engineering community that recognizes an ongoing transition. Here activities from groups such as users, policy makers, societal groups, suppliers, scientists, capital banks etc. has to contribute. Lastly, the use of a sociotechnical landscape level describes how external factors such as oil prices, economic growth, wars, emigration, broad political coalitions, cultural and normative values and environmental problems, has to fall in place to create the right "environment" for a transition to occur.

The relation between the three concepts can be understood as a multi-level perspective where the meso-level ST-regimes accounts for the stability of technological development, the macro-level of landscapes consists of slow changing external factors and the micro-level niches accounts for the development of new innovations and technology. Furthermore, the characters of the concepts are nested as a hierarchy where niches are embedded within regimes and regimes are embedded within landscapes. All the different factors and mechanisms of the multi-level perspective follow certain patterns and are part of a long process in a technological transition. New technology emerges in niches where they have room for errors and learning and are protected from "normal" market selection or criteria. To gain further success technologies has to align its developments at the different levels. Processes of development within the niche level has to be reinforced by changes at regime level and at the level of ST-landscapes. These alignments determines if a shift will occur.

6.3 Empirical results

We will split the results in three parts, first we will present the results obtained based from Trent Smith’s interview, secondly we will present the results obtained from Dimitris Giannoccaro’s interview, and lastly the results from the mathematical model obtained in Section 4.2 will be presented.

Trent Smith

This interview was conducted based on two purposes; first, at the time of the interview we were still looking for data for this thesis thus we asked for data, and secondly we wanted to build an understanding on how companies valued patents. Smith pointed us towards the right direction regarding the data and a data-set was successfully obtained. As for patent valuation the results are not
as clear. First of all the value of a patent is not of significant importance since the value of the patent is equal to the amount spent on the patent for accounting purposes. Another problematic aspect is the valuation of a single patent. When protecting an invention or a method it is rarely protected by a single patent but rather of multiple patents. However when patents are bought there are a few key factors that are considered, namely the claims of the patent, the age of the patent, and most importantly prior experience. From prior experience most patents tends to be worth everything from $0 - $100’000 while an important patent with at least 10 years before expiry is worth approximately $500’000.

Furthermore patents are not always acquired due to the innovation it protects, sometimes patents are acquired because they are made available and suits the patent portfolio the company in question hold. This is due to the fact that every enterprise wish to build a strong patent portfolio for strategic and marketing purposes. A company with a strong patent portfolio signals that the company is well established in the market and reliable.

Moreover patent valuation is tricky due to the dynamical characteristic technology holds. As technology evolves more rapidly the more rapid the value of a patent can shift in both directions.

**Dimitris Giannoccaro**

The first thing one looks at when valuing a patent is the legal status of the patent, secondly one looks at the family members in order to see in which countries the patent operate. This is done since one cannot add additional countries after 12 months from the filing date. Key markets are USA, China, Germany, France and Great Britain. At the point when a patent is to be sold there are a lot of uncertainties that cannot be predicted which forces one to make assumptions regarding the legal status. "A patent that is granted in key markets could be overturned which in turn would leave us with a patent with no market value." This is a factor that the model defined in Section 4.2 is not able to predict, in fact no mathematical model can predict. Just as one cannot predict an oil crisis or recession. Another factor that makes a patent worthless is the simple fact that one forgets to pay the annual fee. This is quite common when mergers takes place. However a lot of resources are invested for maintaining each patent. In fact most enterprises have patent assistants whose job is to maintain all the patents the enterprise possess. The most important value indicator that no mathematical model can capture is the quality of a patent which is very hard to quantify. This can only be captured by a patent expert within the field the patent operates.

Global enterprises patent the most and this is done in order for them to keep an competitive edge and make sure that nobody copies their innovations. However summing together all patenting done by smaller enterprises would surpass the patenting of all the Global enterprises. No matter size of the enterprise they all face the same problem, namely valuing the patents at hand. They are
regarded as costs in accounting.

The increased interest for patent valuation can be pined down to the mere fact that intangible assets have emerged to the surface and are considered to hold a value. In the accounting books patents are seen as costs which is problematic when one is to hold talks with investors or banks since the first question an investor or a bank manager asks is if you hold any assets. Presenting non valued patents as assets will not be considered sufficient.

The Prediction Model

The results obtained in Chapter 4 has an explanation degree of just over 50% indicating the absence of key variables that could have a high explanation degree. Such variable that is also hard to quantify is the quality of the patent. Below ten randomly selected patents are tested with the obtained prediction model in Section 4.2 and are compared to the observed data. As the results indicates the model does not manage to correctly predict the true value of the patent in question. By looking at patent US5438536A where the predicted value is $966'781.09 below the auction result one immediately realize the absence of key explanation variables. This illustrates the difficulty in building a prediction model based on quantitative variables. Key factors such as quality and circumstances cannot be quantified.

<table>
<thead>
<tr>
<th>Patent Number</th>
<th>Predicted Value</th>
<th>Observed Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>US5155847A</td>
<td>$423'206.20</td>
<td>$598'345.00</td>
</tr>
<tr>
<td>US6801987B2</td>
<td>$64'781.60</td>
<td>$275'898.00</td>
</tr>
<tr>
<td>US5987454A</td>
<td>$445'810.83</td>
<td>$1'207'052.00</td>
</tr>
<tr>
<td>US7197640B2</td>
<td>$89'648.69</td>
<td>$130'548.00</td>
</tr>
<tr>
<td>US5699527A</td>
<td>$143'519.33</td>
<td>$352'000.00</td>
</tr>
<tr>
<td>US6070798A</td>
<td>$88'415.49</td>
<td>$220'000.00</td>
</tr>
<tr>
<td>US5564731A</td>
<td>$164'484.23</td>
<td>$44'000.00</td>
</tr>
<tr>
<td>US5438536A</td>
<td>$78'218.91</td>
<td>$1'045'000.00</td>
</tr>
<tr>
<td>WO0188664A3</td>
<td>$184'635.70</td>
<td>$110'000.00</td>
</tr>
</tbody>
</table>

Table 4: Comparison between observed and predicted values of 10 randomly selected patents.

6.4 Discussion

In the last few decades we have experienced great pace in the growth of economies in the western world. Many have suggested this as effect of a shift to what is to be called the "information age" and the knowledge-based economy.
Even though this is hyped labels of today’s society it is a reflection of a real transformation where you can see “knowledge” as opposed to labor, machines, land or natural resources, as the key economic asset that drive long-run economic performance [A. Jaffe & M. Trajtenberg].

The Organisation for Economic Co-operation and Development [OECD] published a paper in 1996 on the knowledge-based economy where they defined the concept knowledge-based economy as following: "the role of knowledge (as compared with natural resources, physical capital and low skill labour) has taken on greater importance. Although the pace may differ all OECD economies are moving towards a knowledge-based economy." This definition recognizes the transition that has taken place where industrial countries have gone from manufacturing based to service driven economies. Thus raises the the importance of patent valuation given the recognized transition.

In the paper The Knowledge Economy published by [W.Powel & K.Snellman] they compare the transition of the automobile. They argue that today the car is less the product of metal fabrication and more a product of a smart computer to illustrate the economic transformation. Moreover, these changes in production are regarded as part of a shift from tangible assets to intangible assets [C. Shapiro & Hal R. Varian]. Previously manufacturing was key for industrial nations economy while today service-driven companies are key to a sustainable economy. This is further illustrated by the activity in the market where Truecaller whose core business lie in providing their customers with information by building a phone-book tool customized for mobile platforms. Truecaller recently received investments worth 430'000'000 SEK [VA] although their accounting show negative annual results. Clearly the investors highly value their IP’s and their 100’000’000 users. One can argue that these type of investments would not be possible by applying traditional valuation methods that make use of key performance indicators and is further evidence of a transition where investors value other aspects of a business.

In 1987, 65’590 [PatBase] patents exchanged ownership while in 2013 207’043 patents exchanged ownership which is an increase of 215% over a period of 26 years. Looking further back, around 1983 to the late 1990s the number of issued patents increase from 47’642 to more than 168’040 which suggest acceleration in the production of new knowledge [W.Powel & K.Snellman]. Furthermore the European Patent Office show that in 1990 8.9% of all Finnish patent applications were high technology patents while in 2000 an increase to 51.6% of all patent applications were high technology [W.Powel & K.Snellman]. The overall patent activity has increased over time as can be seen in Figure 19. On the contrary, another study suggest that the increase in patenting is down to the fact that international treaties and institutions have integrated patent systems over time [D. Somaya]. However we argue that such changes done by Large Technical Systems (LTS) are key indicators that society is adjusting to the observed transition, namely that knowledge-based information is a growing part of society.
A major contributor to the growth of patenting in the U.S. has been universities. University patenting has grown much more rapidly than patenting firms as been confirmed by Owen-Smith saying that there have been more than eightfold increase in university patenting over the period 1976-1998. Furthermore, the increase in patenting has led to changes in the legal and regulatory environments. These changes have further motivated organization to patent inventions. The establishment of Court of Appeals of the Federal Circuit (CAFS) in America 1982 can here be seen as one of the landmarks in the shifting institutional environment within the patent sphere. CAFS works as a specialized appeals court for patent cases and has helped foster an environment where more inventors opt to protect their inventions by patenting, and has a general view that are pro patent. These factors can be seen as different mechanisms in a transition process of patenting as an industrial function. This somewhat aligns with the theories described by Geels where he discuss technological transition with a multi-level perspective and argues that a transition do not only involve technological changes, but also a process of changes in elements such as users practices, regulations and industrial networks. Instead of a technological transition, these factors described above can be seen as part of a industrial transition where the growth of patenting represents a new technology. The growth of patenting within universities and the establishments of institutions corresponds to sociotechnical-regimes that provides stability for the development of innovation and patenting. This is happening while the society we live in are becoming more and more surrounded by technology and companies are becoming more dependent on new and top of edge technology in order to keep a competitive advantage. This new environment for innovation and patenting in the economy can be seen as a factor in the ST-landscape perspective described in Geels theories. While Geels discusses how technology together with the other factors brought up
in his article, causes a societal transition, we argue that we can see how the growth of innovation and patenting are causing an industrial transition in the same way. A transition towards an economy that is, to a greater degree based on intangible assets.

In today’s competitive market patents are used by companies as a key factor in order to reach economic success and to keep a competitive edge [D. Giannoccaro]. However the main function of patents can vary depending on the size of the enterprise. Reason for patenting for smaller or middle sized enterprises is to make sure that they are protected, say for instance you are the best at making screw drivers. You want to keep that competitive edge by patenting since it allows you to operate solely in the market and avoid being copied by competitors, especially larger enterprises with larger resources and marketing channels. A patent is the only protection that will allow such enterprises to keep a competitive edge. For global enterprises patents have multiple functions, the first is to make sure that others do not copy their innovation, secondly patents allows larger enterprises to have greater room for negotiations where patents are used for trading purposes which aligns with Trent Smith’s view.

The given dynamical use of patents further complicates the valuation of patents as the final model in Section 4.2 illustrates, it is very complex to value patents. The value of patents is subjective and the monetary value of patents is in general very hard to observe. The value is actually unobserved until a inquiry of a patent arises [F. Hosini] which leads to the question "what is the market value of a patent?". The answer to this question is once again subjective and depends on different factors such as who the buyer is, the timing of the inquiry and which technology the patent concerns. Furthermore, Mr. Smith points out that one have to consider the aspect of the valuation of multiple patents rather than one. Since products and technologies that are commercialized into the market often are protected by more than one patent it is important to look at patent valuation from a broader perspective, namely by looking at part of or entire portfolio. It is also important to consider the timing, since the economic potential of a patent can vary over time. This is due to the fact that the associated benefits increase or decrease over time [F. Munuri & R. Oriani].

6.5 Conclusion

To revisit the issue of what factors and changes that contribute to a growing importance of patent valuation we can conclude that three key factors can be identified, namely; 1. investments in intangible assets have surpassed traditional investments, 2. the establishment of institutions and organizations that signals a shift towards an information society based on intangible assets, and 3. that knowledge is a key economic asset that drive long-term economic performance. Regarding the complexity of patent valuation we argue that the dynamical use of patents combined with uncertainties regarding the future context of the patent and the many non-quantifiable parameters such as who the
buyer is and the quality of a patent makes patent valuation subjective which in turn makes an objective valuation very difficult to obtain.

Based on the findings of the economical transition and the complexity of patent valuation we suggest that further research needs to be conducted on this field.
References


[D. Giannoccaro] Dimitris Giannoccaro Interview conducted by us May 2015


[F. Geels] Frank W. Geels Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study December 2001

[F. Hosini] Falah Hosini Meeting at IAMIP Sverige AB, Senior strategic IPR adviser and Innovation consultant April 2015


[J. Owen-Smith] Jason Owen-Smith From separate systems to a hybrid order: accumulative advantage across public and private science at Research One universities September 2002

[J. Wild] Jeff Wild Ocean Tomo sells IP transactions group June 2009

http://www.iam-media.com/blog/Detail.aspx?g=b35b6475-adfb-43bf-bf1c-ada95be7d6b0


[PatBase] Developed by Minesoft & RWS Gr A professional search tool for patent information


[T. Smith] Trent Smith Interview conducted by us March 2015


## Appendixes

### Table 5: White’s robust coefficient covariance matrix estimate

<table>
<thead>
<tr>
<th></th>
<th>Int.</th>
<th>Citi.</th>
<th>ActFa.</th>
<th>Age</th>
<th>Inv.</th>
<th>B62</th>
<th>C08</th>
<th>C11</th>
<th>F02</th>
<th>G01</th>
<th>G06</th>
<th>G09</th>
<th>H04</th>
<th>Y02</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int.</td>
<td>0.08</td>
<td>-0.00</td>
<td>-0.04</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.03</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>Citi.</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>ActFa.</td>
<td>-0.04</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>Age</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Inv.</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>B62</td>
<td>-0.03</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>C08</td>
<td>-0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>C11</td>
<td>-0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.33</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>F02</td>
<td>-0.02</td>
<td>0.00</td>
<td>0.02</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.33</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>G01</td>
<td>-0.03</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.07</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>G06</td>
<td>-0.01</td>
<td>-0.00</td>
<td>0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>G09</td>
<td>-0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.32</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>H04</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Y02</td>
<td>-0.01</td>
<td>-0.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.10</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

### Table 6: Coefficient covariance matrix estimate

<table>
<thead>
<tr>
<th></th>
<th>Int.</th>
<th>Citi.</th>
<th>ActFa.</th>
<th>Age</th>
<th>Inv.</th>
<th>B62</th>
<th>C08</th>
<th>C11</th>
<th>F02</th>
<th>G01</th>
<th>G06</th>
<th>G09</th>
<th>H04</th>
<th>Y02</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int.</td>
<td>0.07</td>
<td>-0.00</td>
<td>-0.04</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.00</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.00</td>
</tr>
<tr>
<td>Citi.</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>ActFa.</td>
<td>-0.04</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>Age</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Inv.</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>B62</td>
<td>-0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.56</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>C08</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.56</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>C11</td>
<td>-0.00</td>
<td>0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.10</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>F02</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.39</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>G01</td>
<td>-0.01</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.05</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>G06</td>
<td>-0.01</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>G09</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.28</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>H04</td>
<td>-0.01</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Y02</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.12</td>
</tr>
</tbody>
</table>
A Interview questions

Trent Smith
- What is important to consider when valuing patents?

- How important is a patent valuation method for a business like Tobii?

- Is it important to consider an entire patent portfolio or is it usually single patents that are of great value?

- How are patents used at Tobii?

Dimitris Giannoccaro
- Vad anser du är det svåraste med att värdera patent? (What are the difficulties in valuing patents?)

- Varför är det viktigt att värdera patent? (Why is it important to value patents)

- Vilka förändringar kan du identifiera till ett ökat intresse för värdering av patent? (What changes and transformations can you identify that has contributed to an increased interest in patent valuation?)

- Hur är din erfarenhet av hur företag hanterar, underhåller och utnyttjar patent? (How is your experience of how business manages, maintains and utilizes patents?)


- Hur viktigt är patent för tekniska företag och hur stor vikt lägger man på patent när man skall fatta strategiska beslut inom sin verksamhet? (How important are patents for businesses and are they a part of the strategic making?)

- Hur är eran ekonomi modul tänkt att stödja era kunder? (How is your economic module going to support your customers?)