A Multiple Linear Regression Model To Assess The Effects of Macroeconomic Factors On Small and Medium-Sized Enterprises

EMIL BOOK
LINUS EKELÖF
A Multiple Linear Regression Model To Assess The Effects of Macroeconomic Factors On Small and Medium-Sized Enterprises

EMIL BOOK
LINUS EKELÖF

Degree Projects in Applied Mathematics and Industrial Economics (15 hp)
Degree Programme in Industrial Engineering and Management (300 hp)
KTH Royal Institute of Technology year 2019
Supervisors at KTH: Jörgen Säve-Söderbergh, Julia Liljegren
Examiner at KTH: Jörgen Säve-Söderbergh
Abstract

Small and medium-sized enterprises (SMEs) have long been considered the backbone in any country’s economy for their contribution to growth and prosperity. It is therefore of great importance that the government and legislators adopt policies that optimise the success of SMEs. Recent concerns of an impending recession has made this topic even more relevant since small companies will have greater difficulty withstanding such an event.

This thesis will focus on the effects of macroeconomic factors on SMEs in Sweden, with the usage of multiple linear regression. Data was collected for a 10 year period, from 2009 to 2019, at a monthly interval. The end result was a five variable model with a coefficient of determination of 98%.

Keywords: Statistics, Applied Mathematics, Regression Analysis
Sammanfattning

Små- och medelstora företag (SMEs) har länge varit ansedda som en av de viktigaste komponenterna i ett lands ekonomi, främst för deras bidrag till tillväxt och framgång. Det är därför mycket viktigt att regeringar och lagstiftare för en politik som främjar SMEs optimala tillväxt. Flera år av högkonjunktur och oro över kommande lågkonjunktur har gjort detta ämne ytterst relevant då små företag är de som kommer att drabbas värst av en svårare ekonomisk tillvaro.


Nyckelord: Statistik, Tillämpad matematik, Regressionsanalys
Acknowledgement

We want to express great gratitude to our supervisor during this project Per Jörgen Säve-Söderbergh for advising and giving us the necessary feedback through the process.
Contents

1 Introduction ............................................. 1
   1.1 Background ........................................ 1
   1.2 Purpose and significance .......................... 2
   1.3 Research questions ................................ 2
   1.4 Limitations and Scope ............................ 3

2 Literature review ..................................... 4

3 Economic theory ..................................... 6
   3.1 Macroeconomic theory ............................ 6
      3.1.1 Fiscal policy ................................ 6
      3.1.2 Monetary policy ............................ 7

4 Mathematical theory ................................ 9
   4.1 Multiple Linear regression ...................... 9
      4.1.1 Assumptions ................................ 9
      4.1.2 Ordinary Least Squares .................... 10
   4.2 Possible errors .................................. 11
      4.2.1 Multicollinearity .......................... 11
      4.2.2 Heteroscedasticity ........................ 11
      4.2.3 Autocorrelation ........................... 12
   4.3 Measures of influence ............................ 12
   4.4 Model evaluation ................................ 13
      4.4.1 Test of significance of Regression ....... 13
      4.4.2 Residual Analysis .......................... 13
      4.4.3 R^2 ........................................... 14

5 Model selection ..................................... 15
   5.1 Criteria for model selection ..................... 15
      5.1.1 Adjusted R^2 ................................ 15
<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1 Appendix</td>
<td>41</td>
</tr>
<tr>
<td>A.2 Appendix</td>
<td>43</td>
</tr>
<tr>
<td>A.3 Appendix</td>
<td>45</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Background

The last decade has been characterized by economic expansion and many companies around the globe have experienced steady growth. However, recent years of low economic growth has led many experts to express concerns of a coming recession. To avoid the repercussions of a recession, Swedish government and Sweden’s monetary institution Riksbanken have different sets of tools to combat high volatility in the business cycle. Since companies are bound by their economic environment, they will be affected by changes in the business cycle. The impact will differ greatly depending on the size of the company, with small companies having greater difficulty withstanding the downward trend of the economy. The focus of this thesis will be limited to small and medium-sized enterprises (SMEs).

SMEs are vital to any country’s economy. Their growth and success play an important role in economic growth and prosperity. Due to their significance it is of great importance that the optimal conditions are in place. As reported by Ekonomifakta, SMEs represent 99.9% of the companies in Sweden and employ 64.7% of the labour force [1]. Their importance is further established by a contribution of 61% of GDP in the year of 2015 [2]. As defined by the European Commision, SMEs are grouped in the following three categories, micro, small and medium-sized. The determining factors include staff headcount and either of turnover or balance sheet total. As of SMEs for 2019, the maximum staff headcount is 250 with a turnover of less than 50 million Euros or a balance sheet total of less than 43 million Euros [3].

The significant relationship between the stock market and the economy presents an opportunity to examine stock market indices to evaluate the per-
formance of most segments of an economy. The relationship between stock indices and macroeconomic variables have been confirmed in previous literature studies [4]. However, further studies of the specific relationship between the success of SMEs and macroeconomic variables in Sweden are needed. This paper will attempt to establish the significance of this relationship using multiple linear regression. The index of MSCI Sweden Small Cap will be used by collection of monthly data during the period from 2009 to 2019. MSCI indices are made by Morgan Stanley Capital International (MSCI) using the global investable indexes (GIMI) methodology [5]. Their purpose is to allow cross-regional comparisons across all market capitalisations sizes.

1.2 Purpose and significance

The influence that macroeconomic variables have on small and medium-sized enterprises is important to understand for policy makers, investors and SMEs themselves. The focus will be on the influential policy makers i.e. the Swedish government and its monetary institution Riksbanken.

The purpose of this paper is to complement existing literature on the relationship between macroeconomic variables and the success of SMEs, studying the situation in Sweden. This paper may provide the Swedish government and Riksbanken an indication of how their monetary and fiscal policies affect the growth of SMEs. This may provide guidance on how to effectively manage these policies to optimise the success of SMEs.

The significance of this paper largely depends on its results, and it is not expected to actually greatly influence the Swedish government and Riksbanken. If the results are surprising, it may spark interest and further examination. However, there is certainly an expectation to at least complement existing literature and indicate where future research is needed. This paper may also be of interest for other decision-making actors such as investors and SMEs themselves.

1.3 Research questions

This thesis will examine following questions:

- How does certain macroeconomic variables affect the success of SMEs?
- Which of these macroeconomic variables have the greatest influence?
• What should the Swedish government and Riksbanken have in consideration when using fiscal and monetary policy, to facilitate the success of SMEs?

1.4 Limitations and Scope

In every thesis there are limitations to which results are achievable and which conclusions that can be made from these. One major problem with MSCI Sweden Small Cap is that it is not a great measurement for the success of all types of small and medium-sized enterprises. Firstly, most SMEs are not public limited companies, it is usually the larger companies of SMEs that are. Secondly, MSCI Sweden Small Cap tracks the small capitalisation segment of the investable equity market using different criterias for which companies are included in the small cap than the ones that are used when defining SMEs. Practically this means that there is some overlap between SMEs and large companies in the small capitalisation segment, i.e. some large companies may be included in the small capitalisation segment.

Another limiting factor concerns the 10 year time span that is used for the regression. It is a relatively short time span and as often the case, interpolating and prediction of the future under these circumstances is usually quite poor and inadequate. This is because the regressor variables might change in the future, which severely limits the predictive ability of the regression model. Especially since this time period is characterized by economic expansion and the regression model would be more suitable for other expansion periods. This regression model will be used to identify which variables that are most likely to affect the performance of SMEs, it is not recommended to be used as a prediction model for the stock index.

There are many different factors that affect SMEs, from region specific to internal and external. The scope of this thesis will however be limited to the external macroeconomic factors, especially those that the Swedish government and Riksbanken may influence with their monetary and fiscal policies.
Chapter 2

Literature review

In this section related works are examined, including studies not focusing on small and medium-sized enterprises but also studies on general relationships between macroeconomic variables and the equity market.

Muhammad Irfan Javaid Attari thesis “The relationship between Macroeconomic Volatility and the Stock Market Volatility: Empirical Evidence from Pakistan, 2013”. In his thesis Attari, used the method Exponential Generalized Autoregressive Conditional Heteroskedasticity. The variables that was assumed to influence the stock market was interest rates, inflation and GDP. The results concluded that there were no relationship between GDP and stock return [6].

John H. Boyd, Jian Hu, Ravi Jagannathan wrote the thesis “The Stock Market’s Reaction to Unemployment News: Why Bad News Is Usually Good for Stocks” (2005), which concludes that rising unemployment were positive news during expansions and negative during recessions [7].

Donatas Pilinkus has in his thesis “Macroeconomic Indicators and Their Impact on Stock Market Performance in the Short and Long Run: The Case of the Baltic States.” (2010), modeled the impact that macroeconomic indicators had on stocks in the Baltic market indices. He used vector autoregression to determine the short time relationships between variables like GDP, inflation, interest rate, money supply, industrial production index with stock market indices. He concluded that GDP, importation and state debt had no significant influence [4].

Ahmet Ozcan’s thesis “The Relationship Between Macroeconomic Variables and ISE Industry Index” (2012), used Johansen’s cointegration tests to determine how macroeconomic variables impact the ISE industry index. He concluded that the variables gold price, oil price, exchange rate, interest rate,
money supply and other variables had a long equilibrium relationship with the ISE industry index [8].
Chapter 3

Economic theory

This section will define the theoretical economic framework that is used to examine this thesis’ research questions. The focus will be limited to macroeconomic theory since that is what will be needed to understand variable selection and discussion of results, including the purpose and effect of governmental policies.

3.1 Macroeconomic theory

Macroeconomics is the study of an economy as a whole. It is often used by governments to aid them in forecasting, policy-making and an overall understanding of how their respective economy functions. Macroeconomics may include both the study of long-term economic growth as well as the short-term alternation between periods of economic expansion and contraction, i.e. the business cycle. Both of these are of great importance for the government, however the scope of this thesis is limited to the short-term which is why the business cycle and economic policies to mitigate its effects are in focus. There are namely two different types of economic policies that can be used, monetary and fiscal policies. Generally, the fiscal policies are considered more effective than the monetary ones for stimulating an economy. However, they are not used as often due to time constraints and bureaucracy. Both of these stabilization policies are used by most countries’ governments today [9].

3.1.1 Fiscal policy

Fiscal policy has the intention to change the aggregate demand by adjusting government spending, subsidies or taxes. This will result in a change in con-
sumers disposable income, which should cause a change in consumption and therefore the demand. The fiscal policy can be either expansionary, contractionary or neutral. The expansionary fiscal policy is generally applied when there is a contraction phase in the business cycle and the goal is to mitigate its effects. The government increases spending and subsidies, and lowers taxes for an increase in aggregate demand. The contractionary fiscal policy is the opposite; it is generally applied in the expansion phase of the business cycle when there is risk for excessive inflation. The government decreases spending and subsidies, and increases taxes. Finally, the neutral fiscal policy is applied when the government does not want to affect the aggregate demand. The impact of these fiscal policies will depend on the multiplier effect which makes the usage more difficult since their effectiveness is not guaranteed [10].

3.1.2 Monetary policy

Monetary policy is applied by a country’s central bank with the goal of managing inflation. The central bank achieve their monetary policy by adjusting the money supply which is possible with buying or selling of government bonds, regulating the amount that banks are required to have as reserve or adjusting the repo rate [10]. The monetary policy in Sweden is performed by Riksbanken and their goal is to keep a steady inflation at 2% and avoid high unemployment [11].

Their ambition is that the monetary policy should be predictable to facilitate for actors in the market, to help their planning and decision making. Riksbanken’s main method for monetary policy is the adjustment of the repo rate. “The repo rate is the rate of interest at which banks can borrow or deposit funds at the Riksbank for a period of seven days.” [12].

The repo rate influences the economy in many different ways, and it is said to function through three intertwined channels of transmission mechanism, i.e. interest channel, credit channel and exchange rate channel. The interest channel describes how the overall consumption decreases when the interest rates increase. This is partly due to the fact that it becomes more profitable to save money with increasing interest rates. Since many of the actors are households, a higher interest rate implies higher interest expenses which means lower disposable income, leading to less consumption and aggregate demand. Lastly, an increase in interest rate makes investments less profitable, which can have a substantial impact on many companies. The credit channel illustrates the repo rate’s effect on financial institutions, e.g. banks. Higher interest rates give banks less incentive to lend money which also will lead to less
consumption. The exchange rate channel describes how the repo rate affects the exchange rate, with a higher repo rate usually leading to a higher valuation of the currency. The reasons for this is because a higher repo rate makes Swedish products and assets more attractive for foreign investors, which will lead to an influx of capital and a higher demand for the currency. A stronger currency affects an economy in mainly two different ways. Firstly, foreign products become cheaper compared to domestic ones, which will lead to a decrease in demand for domestic products and consequently lower domestic consumption. Secondly, companies that import products to Sweden will pay less and therefore increase their revenues. However, this also means that domestic production companies will earn less which may be devastating if much of the economy relies on these companies [13] [14].
Chapter 4
Mathematical theory

This section will define the multiple linear regression methodology that will be used for this thesis.

4.1 Multiple Linear regression

The purpose with multiple linear regression is to model the relationship between a dependent variable, the response variable $y$, against two or more independent explanatory variables, the regressor variables $x_k$ [15]. The general form for the multiple linear regression equation is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$$  \hspace{1cm} (4.1)

Where $\beta$ are the regression coefficients which will be estimated with ordinary least-squares. The $k$ is the number of regressor variables and $\varepsilon$ is the error term. It can be written in matrix form with $n$ number of observations as:

$$y = X\beta + \varepsilon$$  \hspace{1cm} (4.2)

Where: $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix}$, $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

4.1.1 Assumptions

In order to accurately estimate the regression coefficients with the use of ordinary least-squares (OLS), five different assumptions must be fulfilled. Viola-
tion of these assumptions may yield a $\beta$ that will differ significantly if applied to different data sets [15].

1. The relationship between the response variable $y$ and the regressors $x$ are approximately linear.
2. The error term epsilon has a mean of zero: $E(\epsilon) = 0$
3. The error term epsilon has constant variance: $V(\epsilon) = \sigma^2$
4. The errors are uncorrelated
5. The errors are normally distributed.

### 4.1.2 Ordinary Least Squares

For estimating the regression coefficients the method of ordinary least squares will be used. The goal is to calculate the vector of least squares estimates $\hat{\beta}$ by minimizing the sum of squares of residuals $SS_{Res}$ [15]:

$$ S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta) $$  \hspace{1cm} (4.3)

The minimized sum of squares is obtained by deriving and setting equal to zero:

$$ \frac{\partial S}{\partial \beta} |_{\hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0 $$  \hspace{1cm} (4.4)

Which is simplified to the least-squares normal equations:

$$ X'X\hat{\beta} = X'y $$  \hspace{1cm} (4.5)

Multiplying by the inverse of $(X'X)^{-1}$ gives the least-squares estimator of $\beta$.

$$ \hat{\beta} = (X'X)^{-1}X'y $$  \hspace{1cm} (4.6)

Provided that the inverse exist, i.e. the regressors are linearly independent, (assumption 1). The fitted regression model corresponding to the observed values are:

$$ \hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y $$  \hspace{1cm} (4.7)

The *gauss-markov theorem* states that if the errors have expected value of 0 (assumption 2), have equal variance (assumption 3) and are uncorrelated (assumption 4). The OLS estimator $\hat{\beta}$ is then the best linear unbiased estimator [15].
CHAPTER 4. MATHEMATICAL THEORY 11

4.2 Possible errors

This section describes some of the possible errors that might happen with multiple linear regression.

4.2.1 Multicollinearity

Multicollinearity is said to occur when there is high correlation among the covariates, i.e. when two or more regressor variables have near linear dependencies between them. The effects of severe multicollinearity are many and significant, including poor estimates of the $\beta$s, large standard errors and susceptibility to influential observations. There is however multiple ways to manage the presence of multicollinearity, with the easiest method being the removal of one or more of the affected regressor variables. To detect multicollinearity, the variance inflation factors (VIF) are often used. A VIF-value greater than 5, is generally considered to indicate that the multicollinearity needs to be managed [15].

$$ VIF_j = C_{jj} = (1 - R_j^2)^{-1} $$

(4.8)

4.2.2 Heteroscedasticity

The third assumption for ordinary least-squares is that the residuals should have approximately a constant variance, there should therefore be no patterns in the scale location-plot, which plots the fitted values against the square root of the standardized residuals. Furthermore the problem of heteroscedastic is said to exist if the constant variance assumption is not met. If heteroscedastic is present, it may cause the standard deviation to be inconsistent and the significance of the coefficients to be invalid. The OLS is no longer the best linear unbiased estimator [16].

Figure 4.1: Heteroscedasticity plot [17]
The three scale location-plots in figure 1, illustrates two types of heteroscedasticity and one homoscedasticity, the latter being what is desired. The two other plots show obvious patterns which mean that there is not constant variance. Heteroscedasticity can be dealt with the use of transformation of certain regressor variables [16].

### 4.2.3 Autocorrelation

This thesis will evaluate the performance of SMEs for a ten-year period which means that both the response and the regressor variables will be time-oriented. When this is the case, it is important to consider how autocorrelation can affect the regression analysis. Autocorrelation happens when the errors are correlated with themselves at different time periods which violates assumption 4. It often occurs when the model is missing regressor variables or the time dependence is not taken in consideration. One of the methods to detect autocorrelation is by the usage of a Durbin Watson test [15].

A model with autocorrelation would be:

\[
y_t = \beta_0 + \beta_1 x_t + \varepsilon_t = \phi \varepsilon_{t-1} + a_t
\]  (4.9)

Durbin Watson test:

\[
H_0 : \phi = 0, \quad H_1 : \phi > 0
\]

\[
d = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}
\]  (4.10)

If d is between \(d_l\) and \(d_u\), the null-hypothesis can be rejected and autocorrelation is present, see Appendix A.3.

### 4.3 Measures of influence

Models often contain observations that are outliers, i.e observations that depart significantly from the rest of the observations. It is important to detect these observations since they might have great influence over the ordinary least squares estimators. Outliers are either leverage points or influential points or both, where leverage points deviates significantly from the regressor value and they may not necessarily influence OLS estimators. Contrarily, the influential point is not consistent with the predicted value and it may significantly affect the fit of the model. It is necessary to examine these observations and they might even need to be removed. One of the methods to measure influence is
the Cook’s Distance value which measures the effects on the OLS when deleting observation $i$. A large value of $D_i$, generally a value above one, indicates that the $i$th observation has significant influence on the OLS and it might need to be removed [15].

Cooks Distance:

$$D_i(X'X, pMS_{Res}) = D_i = \frac{(\hat{\beta}(i) - \hat{\beta})'X'X(\hat{\beta}(i) - \hat{\beta})}{pMS_{Res}}, i = 1, 2, ..., n$$ (4.11)

4.4 Model evaluation

This section will present different evaluation measures for the OLS model.

4.4.1 Test of significance of Regression

Once estimation of $\beta$'s values has been completed, it is important to know how significant the results are and whether a linear relationship between the response and the regressors can be statistically determined. The $F$-statistics and its corresponding $p$-values will be used to see if the null hypothesis can be rejected, and a linear relationship can be confirmed. The null hypothesis $H_0$ is defined as:

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0, \quad H_1 : \beta_j \neq 0, \text{ for at least one } j.$$ 

By definition of $F$-statistics can be calculated as:

$$F_0 = \frac{SS_R/k}{SS_{Res}/(n - k - 1)} = \frac{MS_R}{MS_{Res}}$$ (4.12)

And we can reject our null-hypothesis if $H_0$:

$$F_0 > F_{a,k,n-k-1}$$ (4.13)

$P$-values are measures of probability, if $p$ is larger than alpha its an indication that the null-hypothesis cannot be rejected at that level. If a significance level of 95% is chosen, a $p$-value larger than 5% suggest it should be excluded from the model. In other words, the null hypothesis cannot be rejected at the chosen significance level [15].

4.4.2 Residual Analysis

In order to check for violations of the assumptions in section 4.1.1, a residual analysis needs to be performed. Plotting the residuals is an effective way to
investigate if some violations of the above assumptions occur. Residuals are defined as the observed value or data points, deviation from the fitted value \[ e_i = y_i - \hat{y}_i, \ i = 1, 2, ..., n \] (4.14)

An important property of residuals is that they have a mean of zero and that the average variance of the residuals are defined as:

\[
\frac{\sum_{i=1}^{n}(e_i - \bar{e})^2}{n - p} = \frac{\sum_{i=1}^{n}e_i^2}{n - p} = SS_{Res} = MS_{Res} \tag{4.15}
\]

**Standardized residuals**: Scaling residuals is a way that makes it easier to find outliers or extreme values. Values of scaled residuals \( d_i > 3 \) indicates outlier or extreme values that need to evaluated \[15\].

\[
d_i = \frac{e_i}{\sqrt{MS_{Res}}}, \ i = 1, 2, ..., n \tag{4.16}
\]

The normal probability plot will be used to check whether normality assumption holds. Small deviations in the tails does not affect the model greatly. Residuals vs fitted values plot to check for non-linear connections between variables and to test for heteroscedastic. Scale-location plot illustrates whether the third assumption is fulfilled i.e. does the errors have the same variance. Residuals vs leverage plot illustrates if there is an occurrence of outliers in the data set that needs to be handled \[15\].

### 4.4.3 \( R^2 \)

\( R^2 \) is the value that indicates how good the fit is to the data points in the regression model. A \( R^2 \)-value is a number between 0 and 1, corresponding to the percentages of variability in the model. It is defined as \[15\]:

\[
R^2 = \frac{SS_B}{SS_T} = 1 - \frac{SS_{Res}}{SS_T} \tag{4.17}
\]

A major problem that occurs when evaluating on the basis of \( R^2 \) is that when adding another explanatory variable, the \( R^2 \)-value will always increase. This may lead to the belief that the model with more variables is the better one, which may lead to overfitting. Overfitting variables for a limited set of data points results in a overly complex model. An overly complex model causes the beta coefficients from OLS to be misleading and it would make the model inadequate for prediction of future outcomes. *Occam’s razor*-principle states that a simple model is preferable over a more complex model \[15\].
Chapter 5

Model selection

This section will describe the model selection process.

5.1 Criteria for model selection

There are several different criteria that can be used to compare different models and there is not one specific criteria that is considered the best. The criteria to be used depends on the model and what the model is meant to achieve.

5.1.1 Adjusted \( R^2 \)

Contrarily to \( R^2 \), \( R^2_{\text{Adj}} \) accounts for the number of variables in the model, to avoid overfitting [15]:

\[
R^2_{\text{Adj}} = 1 - \frac{SS_{\text{Res}}/(n - p)}{SS_T/(n - 1)}
\] (5.1)

Where \( n \) is the number of observations and \( p \) is the number of variables.

5.1.2 Bayesian information criterion (BIC)

BIC is a function that rewards models with fewer variables over models with higher number of variables. It penalizes with the amount of variables to discourage overfitting. BIC is a greater penilizer than other information criterium, due to the logarithmic penalty term. The function of the Bayesian information criterion for OLS is:

\[
BIC_{\text{Sch}} = n \ln \frac{SS_{\text{Res}}}{n} + p \ln(n)
\] (5.2)
Where $p$ is the number of variables and $n$ the number of observations. A small value of BIC is desirable [15].

### 5.1.3 Mallows’s $C_p$

Mallows’s $C_p$ criterion is a useful measure when comparing models since it relates the number of variables $p$ with the average error of prediction as well as bias in the model. Generally, small values of $C_p$ are desirable. If $C_p$ is smaller than or equal to $p$, the model have no bias. However, it is sometimes more suitable to select a model with a higher bias, if it means that the average error is much lower [15].

$$C_p = \frac{SS_{Res}(p)}{\hat{\sigma}^2} - n + 2p$$  \hspace{1cm} (5.3)

Where $p$ is the number of variables and $n$ the number of observations.

### 5.2 Best subset regression

The variables that the regression analysis will evaluate are derived from previous studies and certain variables that seem appropriate in accordance with economic theory. It is most likely that not all variables will have a significant relationship with the response and it is important to select a model with relevant variables. It is often preferable with few but significant variables. There are many different techniques for model selection and the one this project will be using is best subsets regression, also referred to as all possible regressions. This method fits all regression models for 1 regressor, 2 regressors and so forth. This way it will calculate $2^k$ number of regression models, where $k$ is the number of regressor variables. This is a computer intensive procedure. However, the number of regressor variables in this thesis is 8, which would yield 256 different regression models, easily calculated using modern computers [15]. The “best” model for $k$ number of regressor variables is selected by the residual sum of squares $SS_{Res}$ criteria. The $k$ number of different best models can now be compared using criteria such as $R^2_{Adj}$, AIC, BIC, $C_p$. 

Chapter 6

Methodology

This thesis will evaluate how different macroeconomic variables affect small and medium-sized enterprises using regression analysis. The macroeconomic variables are chosen on the basis of previous studies and variables that the authors find relevant. There are certainly variables that have an impact on the performance of SMEs but are not included in this thesis. This occurred due to lack of sufficient data, not being able to identify or quantify relevant variables and certain variables that have a constant value, e.g. payroll tax which influence SMEs but have not changed for the last 10 years. Another example is subsidies for SMEs, which directly impacts their success. However, no quantifiable nor suitable data has been found and subsidies will therefore not be included in the variable selection.

6.1 Response variable / MSCI Sweden Small Cap Index

This thesis examines the relationship between the success of SMEs and various macroeconomic variables. Success will be defined as financial success and will be measured by the performance of a stock index for SMEs. The majority of SMEs in Sweden are not publicly limited companies which means that a stock index might not represent their success adequately. It is therefore of great importance that the stock index tracks the micro- or small capitalisation segment of the stock market, for the best representation a stock index can produce for SMEs. The index of MSCI Sweden Small Cap will be used and it is produced by Morgan Stanley Capital International (MSCI) which is one of the largest global index providers. The index currently has 102 constituents,
and represents approximately 14% of the free float-adjusted market capitalisation in the Swedish equity market [18]. The price index level will be collected at a monthly interval from 2009 to 2019 using the domestic currency, SEK. (See appendix A.1).

### 6.2 Regressor variables

The following variables will be used in the multiple linear regression analysis. Due to their complexity, many of the variables will have significant correlation or near linear dependencies, which should lead to high multicollinearity in the regression model.

#### 6.2.1 Inflation

Inflation refers to a rate of which general price level of goods and services increases. Inflation rate is the percent increase in overall price level per year. It is measured by a consumer price index, CPI. Inflation has no major costs for society in the long run since prices and wages rise simultaneously, short-time wages are considered sticky and do not raise as fast as prices. This could lead to lower consumption which decreases returns. Inflation is near-linear dependent with both unemployment and GDP, according to the Phillips curve and the quantity theory of money [10]. Previous studies by Muhammad Irfan Javaid Attari and Donatas Pilinkus concludes that there is a positive correlation between stock market and inflation [6] [4].

#### 6.2.2 Repo rate

As stated in section 3.1.2, the repo rate is the main tool for Riksbanken to perform its monetary policy. According to macroeconomic theory, adjusting the repo rate will stimulate consumption and therefore have a positive influence for SMEs. Previous studies by Muhammad Irfan Javaid Attari and Donatas Pilinkus also establishes a negative correlation between the market interest rate and the outcome of the stock market [4] [6]. The repo rate should have a direct impact on stock returns for SMEs. Measured in percentage.

#### 6.2.3 GDP

Gross Domestic Product is a measurement of how well the overall economy is doing. Its measured by adding the values of goods and services produced
in a country during a specific time period. GDP is a practical tool to measure the size of the economy and it facilitates comparing countries with previous years to measure their growth. The measurement also facilitates to compare other countries against each other. Should have a positive effect on SMEs due to increase in spending in a country should also mean increase in profitability for SMEs. GDP is calculated at a quarterly basis, measured in billions SEK \[10\]. Formula for GDP:

\[
G = C + I + G + (X - M) \\
\]

(6.1)

\(C\) = Consumer spending, \(I\) = Investments, \(G\) = Government spending, \(X\) = Exports, \(M\) = Imports.

6.2.4 Company taxation

The company taxation rate is a direct cost for SMEs and it is imposed by the government on companies. Company taxation is rarely adjusted, mainly because of the bureaucratic limitations that legislators encounter, more can be read under section 3.1. Since taxation is a direct cost, raising taxes should have a negative effect on SMEs. It is measured in percentage.

6.2.5 Exchange rate

The exchange rate is the market value of EUR/SEK. This variable affects SMEs by exportation and importation of goods and services, which have partly increased due to globalisation. It is however unknown if the correlation with the response is positive or negative, depending on whether SMEs are more export or import dependant. Exchange rate is affected by the repo rate, as stated in section 3.1.2.

6.2.6 Economic Tendency Survey (Konjunkturbarometern)

Konjunkturbarometern is a measurement of companies and consumers outlook on the economy, both present and future, and is often used as an indicator for the business cycle. When there is a strong belief that the market will enter a contraction period, consumers should spend less due to concerns of lower disposable income in the future which would amplify the negative market trend. Therefore, Konjunkturbarometern should have a positive correlation with the response variable and GDP.
6.2.7 **Oil spot price (Europe Brent spot price)**

Oil spot price is the price of a barrel of crude oil. Price is set by the aggregate supply and demand of oil in the global market and therefore has a high correlation with the exchange rate. Ahmet Ozcan concluded in his thesis that industry stock indexes correlated with the spot price of oil [8].

6.2.8 **Unemployment rate**

The employment is the total number of people employed in a country, both full and part time employed. The unemployed is defined as the amount of people that are actively seeking for employment, but are currently unemployed. The labor force is the sum of employment and unemployment, dividing unemployed by labour force results in the unemployment rate. When unemployment rate is low, households should have more money to spend, which in turn creates economic growth and a positive impact on SMEs. The unemployment rate is strongly related to the inflation, in accordance with the Phillips curve [10]. Previous study from John H. Boyd, Jian Hu and Ravi Jagannathan shows the effects unemployment has on the stock market. They concluded that unemployment had a positive correlation during expansions and a negative correlation during contractions [7]. It is measured in percentage.

6.3 **Collection of data**

The data has been collected from a 10 year window with datapoint for every month, giving the study a total of 120 data points. GDP that is released every quarter, is giving an average for every month during that quarter. Here is the list of where the data is collected from:

- Unemployment. Ekonomifakta 2019-04-18
- GDP. Ekonomifakta 2019-04-15
- Company taxation. Ekonomifakta 2019-04-18
- Europe brent spot price. U.S Energy Information Administration. 2019-04-21
- KPI - Inflation. Ekonomifakta. 2019-04-14
6.4 Regression analysis

6.4.1 Fitting full model

Implementation of ordinary least squares to fit the full model, i.e. the model where all the regressor variables are used.

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \epsilon \]  

\( y = \) MSCI Sweden Small Cap price index level, \( x_1 = \) Unemployment rate, \( x_2 = \) Company taxation, \( x_3 = \) GDP, \( x_4 = \) Exchange rate, \( x_5 = \) KPI (Inflation), \( x_6 = \) Konjunkturbarometern, \( x_7 = \) Repo rate, \( x_8 = \) Europe brent oil spot price

| Estimate  | Std. Error | t value | Pr(>|t|) |
|-----------|------------|---------|----------|
| (Intercept) | -5031.9904 | 868.9822 | -5.791 | 6.62e-08 |
| Unemployment | 14.4287 | 14.1294 | 1.021 | 0.309387 |
| Company taxation | -38.3209 | 9.6915 | -3.954 | 0.000136 |
| GDP | 6.4643 | 0.4713 | 13.716 | <2e-16 |
| Exchange rate | 46.5381 | 25.4829 | 1.826 | 0.070501 |
| KPI | 4.5228 | 18.6862 | 0.242 | 0.809195 |
| Konjunkturbaro. | 1.9557 | 1.5378 | 1.272 | 0.206121 |
| Repo rate | -144.8866 | 44.3069 | -3.270 | 0.001432 |
| Europe brent spot price | 0.4502 | 0.6476 | 0.695 | 0.488437 |

Table 6.1: Coefficients for the full model

F-statistic: 728.5 on 8 and 111 DF, p-value: < 2.2e - 16

6.4.2 Full model analysis

The full model fit is overall good, the \( R^2_{Adj} \) is 0.98 and the p-values for Company taxation, GDP and Repo rate is satisfactory. It should be noted that some
of the regressor variables have low significance which means that a linear relationship for each variable cannot be confirmed. A residual analysis is performed and the regression assumptions are evaluated.

The normal Q-Q plot illustrates some deviations for the tails, however they are not significant and the normality assumption holds. The residuals vs fitted values plot illustrates that the expected residual is approximately equal to zero and the small deviation is expected, therefore assumption 2 holds. Observing the plot there might be a pattern in the errors suggesting that autocorrelation is present and a Durbin-watson test will be added in further analysis. In the scale-location plot the reader can observe that the variance is constant and there is no pattern, which suggests that heteroscedasticity is not present, therefore assumption three holds. In the residual vs leverage plot there are almost no occurrences of outliers or high leverage points. The observations 1, 18 and 120 are influential, however not influential enough for the need of intervention, since their $D_i$-values are less than one. For the fourth assumption, a test for multicollinearity is needed. Performing a VIF-test produces the table below, where it is seen that strong multicollinearity is present.
6.4.3 Best subsets regression

The full model residual analysis concluded that the OLS-assumptions are met and no transformations is needed. However, the succeeding analysis reveals weak significance for certain regressor variables and strong multicollinearity in the model. It is therefore suitable to proceed with the best subsets regression to select the most important regressor variables and pick the most appropriate model. The best subsets regression will be calculated using the leaps-package in R, with a method called regsubsets. It will return the two “best” regression models for each number of regressor variables, and the criteria for “best” model will be having the lowest SS_{Res}. This will produce 15 different regression models that subsequently can be examined and compared using $R^2_{Adj}$, BIC and Mallows’s $C_p$. The following plots illustrates the result. Values for the two models with one variable have not been included since they are outliers and makes the other values indistinguishable.

<table>
<thead>
<tr>
<th>Unemployment</th>
<th>Company taxation</th>
<th>GDP</th>
<th>Exchange rate</th>
<th>KPI</th>
<th>Konjunktur</th>
<th>Repo rate</th>
<th>Oil spot price</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.679</td>
<td>5.917</td>
<td>17.582</td>
<td>3.649</td>
<td>6.571</td>
<td>3.613</td>
<td>18.803</td>
<td>4.171</td>
</tr>
</tbody>
</table>

Table 6.2: VIF table for the full model
The above plot illustrates how $R^2_{Adj}$ changes depending on the number of variables for the different models. There is a maximum of 0.9802315 for the 11th model, that have 6 variables. Most of the $R^2_{Adj}$-values are very similar. See appendix A.2.
Figure 6.3: BIC vs number of variables plot

The above plot illustrates the different BIC-values. The minimum BIC-value is -449.94 for the fifth model with 3 variables. See appendix A.2.
The above plot shows the different $C_p$-values with a minimum of 4.782224 for the ninth model with 5 variables. See appendix A.2.

Two different models seems promising, the fifth and ninth. They have the smallest BIC and $C_p$-value respectively and very similar $R^2_{Adj}$-values. Both models are selected for further examination.

|                | Estimate  | Std. Error | t value | $\text{Pr}(>|t|)$ |
|----------------|-----------|------------|---------|-------------------|
| (Intercept)    | -4744.9649| 344.3142   | -13.781 | $<$2e-16          |
| Company taxation | -39.5634 | 8.1923     | -4.829  | 4.30e-06          |
| Exchange rate  | 43.3347   | 20.7085    | 2.093   | 0.0386            |
| Repo rate      | -129.3443 | 21.9491    | -5.893  | 3.93e-08          |
| GDP            | 6.3242    | 0.2306     | 27.430  | $<$2e-16          |
| Konjunkturbaro.| 2.5795    | 1.1684     | 2.208   | 0.0293            |

Table 6.3: Mallows’s $C_p$ model’s coefficients

Residual standard error: 91.45 on 114 degrees of freedom

$R^2$: 0.981, $R^2_{Adj}$: 0.9802
F-statistic: 1178 on 5 and 114 DF, p-value: < 2.2e-16
Durbin Watson test: DW = 0.70865, p-value = 1.317e-14
Appendix A.3 for sample size approximately 100 at a 5% significance level gives the values $d_l=1.57$ $d_u=1.78$
Since DW is smaller than 1.57, reject the autocorrelation null hypothesis and conclude that the errors are positively correlated.

|            | Estimate | Std. Error | t value | Pr(>|t|) |
|------------|----------|------------|---------|----------|
| (Intercept)| -4602.4942 | 326.0569 | -14.116 | <2e-16   |
| Company taxation | -27.5133 | 6.6910 | -4.112 | 7.36e-05 |
| Repo rate   | -163.6629  | 15.2957 | -10.700 | <2e-16   |
| GDP         | 6.5667    | 0.1969 | 33.356 | <2e-16   |

Table 6.4: BIC model’s coefficients coefficients:

- Residual standard error: 93.17 on 116 degrees of freedom
- $R^2$: 0.9799, $R^2_{Adj}$: 0.9794
- F-statistic: 1889 on 3 and 116 DF, p-value: < 2.2e-16
- Durbin Watson test: DW = 0.67589, p-value = 1.461e-14
- Appendix A.3 for sample size approximately 100 at a 5% significance level gives the values $d_l=1.61$ $d_u=1.74$
- Since DW is smaller than 1.61, reject the autocorrelation null hypothesis and conclude that the errors are positively correlated.

Two multicollinearity tests are performed for both models, the variation inflation factor and the variance decomposition proportions table:

<table>
<thead>
<tr>
<th>Company taxation</th>
<th>GDP</th>
<th>Exchange rate</th>
<th>Konjunkturbaro.</th>
<th>Repo rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.273386</td>
<td>4.253227</td>
<td>2.435832</td>
<td>2.108098</td>
<td>4.664235</td>
</tr>
</tbody>
</table>

Table 6.5: VIF table for Mallows’s $C_p$ model:
<table>
<thead>
<tr>
<th>Number</th>
<th>Condition index</th>
<th>(Intercept)</th>
<th>Company taxation</th>
<th>GDP</th>
<th>Exchange rate</th>
<th>Konjunktur</th>
<th>Repo rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
</tr>
<tr>
<td>2</td>
<td>2.556</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.206</td>
</tr>
<tr>
<td>3</td>
<td>22.184</td>
<td>0.000</td>
<td>0.032</td>
<td>0.009</td>
<td>0.044</td>
<td>0.225</td>
<td>0.058</td>
</tr>
<tr>
<td>4</td>
<td>37.665</td>
<td>0.010</td>
<td>0.154</td>
<td>0.124</td>
<td>0.009</td>
<td>0.303</td>
<td>0.137</td>
</tr>
<tr>
<td>5</td>
<td>71.541</td>
<td>0.058</td>
<td>0.419</td>
<td>0.068</td>
<td>0.936</td>
<td>0.411</td>
<td>0.568</td>
</tr>
<tr>
<td>6</td>
<td>113.865</td>
<td>0.932</td>
<td>0.395</td>
<td>0.800</td>
<td>0.011</td>
<td>0.060</td>
<td>0.029</td>
</tr>
</tbody>
</table>

Table 6.6: Variance decomposition proportions for Mallows’s $C_p$ model

<table>
<thead>
<tr>
<th>Company taxation</th>
<th>GDP</th>
<th>Repo rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.746151</td>
<td>2.987363</td>
<td>2.182044</td>
</tr>
</tbody>
</table>

Table 6.7: VIF table for BIC model

The multicollinearity is not significant for either of the models and no further actions are needed.

<table>
<thead>
<tr>
<th>Number</th>
<th>Condition index</th>
<th>(Intercept)</th>
<th>Company taxation</th>
<th>GDP</th>
<th>Repo rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.012</td>
</tr>
<tr>
<td>2</td>
<td>2.080</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.441</td>
</tr>
<tr>
<td>3</td>
<td>24.778</td>
<td>0.001</td>
<td>0.309</td>
<td>0.131</td>
<td>0.526</td>
</tr>
<tr>
<td>4</td>
<td>85.022</td>
<td>0.999</td>
<td>0.691</td>
<td>0.868</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Table 6.8: Variance decomposition proportions for BIC model
Chapter 7

Results

A regression analysis has been performed using the statistical programming language R. A multiple regression model has been produced with all of the 8 different regressor variables. The full model did suffer from regressors with low significance and strong multicollinearity. The best subsets regression methodology was then used to determine that there is two different models that seem to be most suitable and a selection of the two is therefore needed.

7.1 Final model

The first model has 3 regressor variables: Company taxation, GDP and Repo rate. The second model has 5 regressor variables: Company taxation, GDP, Exchange rate, Konjunkturbarometern and Repo rate. These two models are similar to a high degree, with the first one having a slightly lower BIC-value and the second having a slightly higher $R^2_{\text{Adj}}$-value. Both these differences are small enough to be insignificant. However, the Mallows’s $C_p$-value actually differs slightly, with the second model having a value of 4.78 and the first with a value of 7.13. It is preferable with a low $C_p$-value and more notably a $C_p$-value lower than the model’s number of regressor variables. Having a $C_p$-value higher than this means that the model has substantial bias [15]. Two extra regressor variables does not complicate the model enough for it to be concerning. This reasoning leads to the conclusion that the second model is preferable and it is therefore also the final model, given as

$$y = -4745 - 39.6x_1 + 6.3x_2 + 43.3x_3 + 2.6x_4 - 129.3x_5$$

(7.1)

$y$ = MSCI Sweden Small Cap price index level, $x_1$ = Company taxation, $x_2$ = GDP, $x_3$ = Exchange rate, $x_4$ = Konjunkturbarometern, $x_5$ = Repo rate
This model has a $R^2$-value of 0.981, which means that the model explains 98.1% of the variation in the data. All regressor variables have a p-value lower than 0.05 with an overall p-value of less than 2.2e-16, which means that the null hypothesis can be rejected and a significant relationship with the response variable can be assumed. The final model has a high F-statistic value which means that there is a strong linear relationship between the response and the regressors. However, the Durbin Watson test showed that positive autocorrelation could not be rejected.

Figure 7.1: Partial regression plots

The added variable plots, figure 7.1, illustrate that there are linear relationships between every regressor and the response variable.
Figure 7.2: Residuals Plots

The residuals plots, figure 7.2, show the same behaviour as for the full model, section 6.4.2.
Chapter 8

Discussion

The use of multiple linear regression analysis to evaluate the significance of regressor variables relationship with a response variable has been adequate. There are nevertheless multiple restrictions and complications that can arise when performing regression, as is evident in this thesis. A common restriction that is apparent in research, is the cause effect relationship that many studies attempt to prove but cannot. Regression analysis is a great tool to aid in confirmation of a cause and effect relationship, however it cannot be the sole basis for that conclusion. We encounter this in our attempt to interpret the results of our model. The coefficients in the final model, section 7.1, acts in accordance with current macroeconomic theory, especially for the company taxation and repo rate variables. As expected, these two affect the price index level negatively and it would be convenient to conclude that the Swedish government and Riksbanken definitely should lower taxes and repo rate for the short term. However, as is the restriction, it is not possible to conclusively determine this cause and effect relationship.

We also encounter a significant complication with the presence of autocorrelation in our final model. This is not unexpected, as written in section 4.2.3, since our regressor variables are time-oriented, as is the case for many economic studies, and it is reasonable that the previous day’s value affects the value today. Positive autocorrelation can have severe consequences for the model fit since the ordinary least squares estimators are no longer the best linear unbiased estimators, which often leads to the coefficients standard errors and p-values to be underestimated [15]. However, since our p-values are very small, we find it unlikely that our regressor variables would be rejected at the 95% significance level, with the exception of exchange rate and Konjunkturbarometern because of their comparatively high p-values. This means that the
final model’s variables are most likely correct, but our result could still be misleading or incorrect, i.e. it is possible that our estimated $\beta$s are inaccurate and our $R^2$-value is inflated. It is therefore of great importance that autocorrelation is considered and evaluated in future work.

Lastly, our result may also have been affected by the scope of this thesis and the variable selection. As mentioned in sections 1.4 and 6.1, the response variable MSCI Sweden Small Cap price index is not the best representation of the performance of SMEs. Further, we suspect that there are many more regressor variables that affect SMEs but they are non-quantifiable or have a near constant value for the last 10 years.

8.1 Regression variables

This section will discuss the results for each of the regressor variables.

8.1.1 Repo rate

Since monetary policy has direct effect upon the interest rate and the theory suggest that its a tool for stimulating demand and therefore consumption, our result is expected.

During the last years of expansion, monetary policy should have normalised the repo rate. The repo rate has been at an unprecedented negative rate and an all time low which means that our discussion about lowering the repo rate may therefore not be possible. There are concerns that Riksbanken will not be able to stimulate the economy further with repo rate and thus will be required to use quantitative easing to perform their monetary policy instead. Quantitative easing has its fair share of criticism and some experts believe that it is inefficient [14].

8.1.2 Company taxation

The company taxation rate affects SMEs directly, unlike the interest rates from the monetary policy. The result for company taxation is expected and it suggests that a decrease in taxation would have a positive impact on SMEs. However, this decrease is not very likely to happen due to the current limitations of fiscal policy, as mentioned in section 3.1. Similarly, one problem in this thesis has been that the company taxation only is changed a few times, it being nearly constant limits the ability to conclude the results. It would be beneficial with more variation in the observations. A better variable would have been
the *effective company taxation*, which accounts for company deductions. This would be a better representation of company taxation and would give us more variation in the observations.

### 8.1.3 GDP

GDP is a measurement of all consumption for a specific time period and the fact that it increases with the performance of SMEs is not a surprise. It is expected since an increase in GDP will disperse to all sectors of the economy. However, the cause and effect relationship for GDP and the response variable cannot be determined, which is because of an absence of supporting theory.

### 8.1.4 Exchange rate

The results for exchange rate indicates that when the value of EUR increases compared to SEK, the performance of SMEs increase. This can be explained with the fact that SMEs can export at a lower price than their competitors in countries with EUR as their currency. An interpretation would be that SMEs are more affected by export than import, i.e. that SMEs export more than they import.

### 8.1.5 Economic Tendency Survey (Konjunkturbarometern)

The results for Konjunkturbarometern are expected as well. Konjunkturbarometern is positively correlated with our response variable which means that our hypothesis in section 6.2.6 should be correct, i.e. that if there is a strong belief that the market will enter an expansion period, consumers spend more of their disposable income which would amplify the positive market trend.

### 8.1.6 Inflation, oil spot price and unemployment

Inflation, oil spot price and unemployment rate had no statistical significance with our response variable. This is not surprising for inflation due to companies raising their prices in accordance. Since the repo rate is a tool used by Riksbanken for adjusting both inflation and unemployment, as stated in section 3.1.2, their low significance could be a consequence of their relationship with the repo rate. Likewise, since Sweden does not produce oil, the oil spot price is very dependent on the exchange rate which would explain its low significance.
8.2 Future work / research

Autocorrelation needs to be dealt with when using regression and time-oriented regressor variables. For future work, there is mainly three different ways to deal with autocorrelation. If the autocorrelation occurs because of missing regressor variables, it can be solved by identifying the missing variable and including it. If there is in-depth knowledge about that particular autocorrelation structure, it can be helpful to use general linear model (GLM) or weighted least-squares (WLS). If these methods are unable to solve the autocorrelation issue, the last step would be to include the autocorrelation structure in the model [15].

We also believe that further research of differently sized companies and from different countries would be of interest. This research could compare all different sections and may achieve more conclusive results. A longer time period could also yield interesting results.
Chapter 9

Conclusion

A multiple regression analysis has been performed to examine the significance of the relationship between macroeconomic variables and the performance of a small capitalisation stock index, MSCI Sweden Small Cap. The goal was to research how the success of SMEs in Sweden may be influenced by these variables, and how policy-makers like the Swedish government and Riksbanken can act to ensure the optimal performance of SMEs.

This resulted in a five-variable regression model with a coefficient of determination of 98% and the variables: Company taxation, GDP, Exchange rate, Konjunkturbarometern and Repo rate. As is discussed in previous section, the very high coefficient of determination appear more than satisfactory but the presence of autocorrelation casts doubt on the usefulness of the measure as well as the overall fit of the final model. However, the scope of this thesis does not include the action of correcting autocorrelation and it is left for further research.

We can with our model conclude that company taxation and repo rate have the strongest influence on the performance of SMEs, with both showing a negative correlation, i.e. that a lowering of taxes and repo rate would increase the performance of SMEs. This is however only theoretically according to our model in accordance with macroeconomic theory and, as previously discussed, this cause and effect relationship might not exist in the real world. We recommend nevertheless that the Swedish government and Riksbanken studies these relationships further.
Bibliography


List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Heteroscedastity plot</td>
<td>11</td>
</tr>
<tr>
<td>6.1</td>
<td>Residuals plots</td>
<td>22</td>
</tr>
<tr>
<td>6.2</td>
<td>$R^2_{Adj}$ vs number of variables</td>
<td>24</td>
</tr>
<tr>
<td>6.3</td>
<td>BIC vs number of variables plot</td>
<td>25</td>
</tr>
<tr>
<td>6.4</td>
<td>Mallows’s Cp vs number of variables plot</td>
<td>26</td>
</tr>
<tr>
<td>7.1</td>
<td>Partial regression plots</td>
<td>30</td>
</tr>
<tr>
<td>7.2</td>
<td>Residuals Plots</td>
<td>31</td>
</tr>
</tbody>
</table>
Appendix A

MSCI Sweden Small Cap

The following methodology and information are directly summarized from (MSCI Index Calculation Methodology, 2019), (MSCI Global Investable Market Indexes methodology, 2019) and (MSCI Sweden Small Cap Index (USD) fact sheet, 2019):

MSCI Sweden Small Cap is a stock index designed to measure the performance of the small cap segment of the Swedish equity market. With 102 constituents, the index represents approximately 14% of the free float-adjusted market capitalisation in Sweden."

The price index level is calculated against time with local currency SEK as:

\[
\text{Price Index Level Local}_t = \frac{\text{Price Index Level Local}_{t-1} \times \text{Index Adjusted Market Cap For Local}_t}{\text{Index Initial Market Cap USD}_t} \tag{A.1}
\]

Where:

\[
\text{Index Adjusted Market Cap For Local}_t = \sum_{s \in t} \frac{\text{End of Day Number Of Shares}_{s-1} \times \text{Price Per Share}_s \times \text{Inclusion Factor}_s \times \text{PAF}_t}{\text{FX rates}_{s-1}} \times \frac{\text{ICI}_t}{\text{ICI}_{t-1}}
\]

\[
\text{Index Initial Market Cap USD}_t = \text{41}
\]
\[ \sum_{s \text{ in } l,t} \frac{\text{EndofDayNumberOfShares}_{t-1} \times \text{PricePerShare}_{t-1} \times \text{InclusionFactor}_t}{\text{FXrate}_{t-1}} \]

\( \text{EndofDayNumberOfShares}_{t-1} \) is the number of shares of security s at the end of day t-1.

\( \text{PricePerShare}_t \) is the price per share of security s at time t.

\( \text{InclusionFactor}_t \) is the inclusion factor of security s at time t.

\( \text{PAF}_t \) is the Price Adjustment Factor of the security s at time t.

\( \text{FXrate}_t \) is the FX rate of the price currency of security s vs USD at time t. It is the value of 1 USD in foreign currency.

\( \text{ICI}_t \) is the Internal Currency Index of price currency at time t. It is equal to 1 for the Swedish indices.

MSCI Sweden Small Cap Index: The small cap segment is defined as following:

A market Investable Equity Universe is firstly derived by identifying eligible listings for each security and then applying investability screens. In order to be included in a Market Investable Equity Universe, a company must have the required minimum full market capitalisation, referred to as the Equity Universe Minimum Size Requirement. This is derived using following steps:

- Sort all companies in the Equity Universe in descending order of full market capitalisation and the cumulative coverage of the free float-adjusted market capitalisation.

- When the cumulative free float-adjusted market capitalisation coverage of 99% of the sorted Equity Universe is achieved, the full market capitalisation of the company at that point defines the Equity Universe Minimum Size Requirement.

NOTE: To be eligible for inclusion in a Market Investable Equity Universe, a security must have a free float-adjusted market capitalisation equal to or higher than 50% of the Equity Universe Minimum Size Requirement.
The Market Investable Equity Universe is segmented into the investable market index, standard index (Large + Mid), Large Cap index, Mid Cap index and Small Cap index. The Investable market index, standard index and large cap index are created first, and the Small Cap index is derived as the difference between the Investable Market Index and the Standard Index.

- Large Cap Index: 70% ± 5
- Standard Index: 85% ± 5
- Investable Market Index: 99 + 1% or −0.5%

A.2 Appendix

Appendix B

Tables for $R^2_{adj}$, BIC and Mallow's $C_p$ from best subsets regression. Section 6.3.2

<table>
<thead>
<tr>
<th>Model number</th>
<th>Number of variables</th>
<th>Adjusted $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.9501047</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.8685782</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.9768245</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.9634717</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.9794223</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0.9771390</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>0.9797214</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>0.9793942</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>0.9801767</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0.9798096</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>0.9802515</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>0.9801048</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>0.9801310</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>0.9800551</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>0.9799626</td>
</tr>
<tr>
<td>Model number</td>
<td>Number of variables</td>
<td>BIC</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-551.1777</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-123.9606</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>-438.3995</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>-254.8316</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>-449.9400</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>-437.3127</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>-447.9483</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>-447.1978</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>-446.9335</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>-444.7320</td>
</tr>
<tr>
<td>11</td>
<td>6</td>
<td>-443.5358</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>-442.7692</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>-439.2666</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>-438.7487</td>
</tr>
<tr>
<td>15</td>
<td>8</td>
<td>-434.4824</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model number</th>
<th>Number of variables</th>
<th>Mallows’s C²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>177.832996</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>183.5751695</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>22.491766</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>99.291804</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>1.127689</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>20.346497</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>6.384285</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>7.114438</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>4.782224</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>6.870343</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>5.483519</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>6.198002</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>7.058384</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>7.483183</td>
</tr>
<tr>
<td>15</td>
<td>8</td>
<td>9.000000</td>
</tr>
</tbody>
</table>
Critical values of the Durbin-Watson statistics. Table taken from Introduction to linear regression analysis page 552.

### TABLE A.6 Critical Values of the Durbin—Watson Statistic

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Probability in Lower Tail (Significance Level = α)</th>
<th>k = Number of Regressors (Excluding the Intercept)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>d₁</td>
<td>d₂</td>
</tr>
<tr>
<td>15 .01</td>
<td>.81</td>
<td>1.07</td>
</tr>
<tr>
<td>.025</td>
<td>.95</td>
<td>1.23</td>
</tr>
<tr>
<td>.05</td>
<td>1.08</td>
<td>1.36</td>
</tr>
<tr>
<td>.01</td>
<td>.95</td>
<td>1.15</td>
</tr>
<tr>
<td>20 .025</td>
<td>1.08</td>
<td>1.28</td>
</tr>
<tr>
<td>.05</td>
<td>1.20</td>
<td>1.41</td>
</tr>
<tr>
<td>.01</td>
<td>1.05</td>
<td>1.21</td>
</tr>
<tr>
<td>25 .025</td>
<td>1.13</td>
<td>1.34</td>
</tr>
<tr>
<td>.05</td>
<td>1.20</td>
<td>1.45</td>
</tr>
<tr>
<td>.01</td>
<td>1.13</td>
<td>1.26</td>
</tr>
<tr>
<td>30 .025</td>
<td>1.25</td>
<td>1.38</td>
</tr>
<tr>
<td>.05</td>
<td>1.35</td>
<td>1.49</td>
</tr>
<tr>
<td>.01</td>
<td>1.25</td>
<td>1.34</td>
</tr>
<tr>
<td>40 .025</td>
<td>1.35</td>
<td>1.45</td>
</tr>
<tr>
<td>.05</td>
<td>1.44</td>
<td>1.54</td>
</tr>
<tr>
<td>.01</td>
<td>1.32</td>
<td>1.40</td>
</tr>
<tr>
<td>50 .025</td>
<td>1.42</td>
<td>1.50</td>
</tr>
<tr>
<td>.05</td>
<td>1.50</td>
<td>1.59</td>
</tr>
<tr>
<td>.01</td>
<td>1.38</td>
<td>1.45</td>
</tr>
<tr>
<td>60 .025</td>
<td>1.47</td>
<td>1.54</td>
</tr>
<tr>
<td>.05</td>
<td>1.55</td>
<td>1.62</td>
</tr>
<tr>
<td>.01</td>
<td>1.47</td>
<td>1.52</td>
</tr>
<tr>
<td>80 .025</td>
<td>1.54</td>
<td>1.59</td>
</tr>
<tr>
<td>.05</td>
<td>1.61</td>
<td>1.66</td>
</tr>
<tr>
<td>.01</td>
<td>1.52</td>
<td>1.56</td>
</tr>
<tr>
<td>100 .025</td>
<td>1.59</td>
<td>1.63</td>
</tr>
<tr>
<td>.05</td>
<td>1.65</td>
<td>1.69</td>
</tr>
</tbody>
</table>
