Modelling credit risk with macroeconomic factors.

Jesper Frick

September 7, 2009

Abstract

A model of loan losses for Swedbank has been developed. Loan losses of nine different segments and the two major regions in which Swedbank is active, Sweden and the Baltic region, are mapped to changes in macroeconomic factors of the two regions. The time lags of with which macroeconomic factors explain loan losses are studied. Linear regression is performed of loss ratios on changes of macroeconomic factors at the optimal time lags. A model for the changes of macroeconomic factors is calibrated and a large number of simulations are performed. Using the simulated samples, Swedbank's future loss ratios are predicted one quarter and one year ahead.

Changes of macroeconomic factors in the Baltic region are shown to explain the loan losses of the Baltic region well. The changes of macroeconomic factors of the Swedish region are shown to have a low level of explanation of loss ratios pertaining to the Swedish region. The fact that loss ratios in the Baltic region are better explained by macroeconomic factors could be explained by customers in this region having less buffer to adverse movements of the economy.

Acknowledgements

I would like to thank Georges Mansourati at Swedbank and Filip Lindskog at KTH for their guidance, sharing of knowledge and invaluable support. Moreover, I would like to thank my mother, Annika Eliasson Frick, for proof-reading my master's thesis. Last but not least, I would like to express my gratitude to my girlfriend, family and friends for bearing with me during the last few months.

Stockholm, September 2009

Jesper Frick

Contents

Chapter 1

Introduction

1.1 Problem Formulation

The purpose of this master's thesis is firstly to complete a survey of the literature within the field of dependence and interaction of market and credit risk. Secondly, the aim is to develop a model for Swedbank's credit losses by studying the dependence between credit losses and macroeconomic factors. Using this model the interdependence between credit losses in different segments will be studied.

1.2 Structure

In the remainder of chapter 1 literature related to interaction of credit and market risk is summarized. In chapter 2 a theoretical background of used mathematics is provided. Data from Swedbank and macroeconomic data is explained and presented in chapter 3. In chapter 4 regression models of loss ratios are derived. A model of the macroeconomic factors is calibrated in chapter 5 . These models are used for simulation and prediction of loss ratios in chapter 6 and 7. A conclusion and summary is given in chapter 8.

1.3 Related Literature

There are many risktypes that banks face. The most commonly addressed are credit risk, market risk, operational risk and business risk. Only credit and market risks are within the scope of this literary overview. Market risk can be divided into:

- Equity risk, the risk of adverse stock price changes.
- Interest rate risk, the risk of adverse interest rate changes.
- Currency risk, the risk of adverse foreign exchange rate changes.

- Commodity risk, the risk of adverse commodity price changes.

Credit risk is the risk of loss due to a non-payment of a loan or interest on a loan. Credit risk is usually divided into credit risk from lending to corporate customers and private customers.

The different models can be divided into "top-down" and "bottom-up" models (Overbeck, 2006). In the "top-down" approach risks are first calculated by risk type and then aggregated on group level of the bank. The interdependence of the risks are then calculated using the aggregated risk factors. In the "Bottom-up" approach all risks are first considered on a transaction level and then aggregated to calculate the total risk of the bank.

Different risks are measured over different time horizons. Market risk is usually measured using a VaR measure over a ten day period. VaR is a single period model where the portfolio is assumed to remain the same over the ten day period. Credit risk is usually measured on a one year time horizon using a single period model. Measuring credit and market risk in the same model requires a time scaling of either or both risk types so that they are measured on the same time horizon. Most commonly market risk is scaled to the same time horizon as credit risk. One could take the sum of daily/ten day VaR over one year and consider this your measure of market risk. (Rosenberg & Schuermann, 2006; Larsson, 2009). Aas et al. (2007) first measure market risk using the worst ten day VaR outcome over one year. They then take the sum of the macroeconomic factors used in the calculation to measure credit risk on a one year time horizon. Drehmann et al. (2008) model credit risk and interest rate risk in the same multi period model thus eliminating the need of scaling either of the risks.

One way of measuring interdependence between different risk types is to first map the risks to macroeconomic factors and then model the dependence between these. Here one must decide upon macroeconomic factors. As to corporate credit risk (lending to corporations) Carling et al. (2007) show that yield curve, output gap and household expectations are good explanatory macroeconomic variables. Rosenberg & Schuermann (2006) use monthly spreads for bond indices and treasury yields to model credit risk. They use equity returns (S&P 500), currency returns (10 year constant maturity treasury rate) and interest rate changes (difference of the trade weighted currency index). Aas et al. (2007) use a Norwegian credit loss index reported yearly for credit risk and eleven different indices to model market risk. Drehmann et al. (2008) stress the importance of income gearing as an explanatory variable of both household and corporate probability of defaults. Further they find that property prices affect the probability of defaults. As to modelling interest rate they use output gap, inflation and bank rate as explanatory variables.

One must consider that risks can be affected by changes in macroeconomic factors with different time lags. Larsson (2009) addresses this concern. Larsson studies the optimal time lag in terms of correlation between macroeconomic variables and the risk figures reported by the bank.

Credit risk may also have an accounting delay, i.e. a default may be reported or viewed as a default by the bank with a delay. This could create a calibrating problem since both above concerns create a time lag between actual macroeconomic changes and realizations of credit risk.

1.3.1 Integration of Credit and Market risk (Overbeck, 2006)

The author provides an overview of risk integration approaches. Top-down and bottom-up approaches are explained.

Top-Down Approach

In the top-down approach credit and market risk is first measured separately and then the sum of the two stochastic variables defines the total risk.

$$L_{total} = L_M + L_C$$

Now it remains to obtain the joint bivariate distribution of the vector (L_M, L_C) to calculate the distribution function of L_{total}. Overbeck identifies two approaches, the factor approach and the copula approach. Note that both risks have to have the same time horizon. Typically market risk is measured at a 10-day time horizon and credit risk on a one-year horizon. A common solution is to scale market risk to a one-year time horizon, something the authors argues is an acceptable compromise on a very high aggregation level.

Factor Approach

In the factor approach market and credit risks are driven by a set of factors. For market risk these factors could be interest rates, FX rates, equity prices etc.. For credit risk these factors could be default rates, macroeconomic variables, asset values and equity indexes. All factors are denoted by the vector $\hat{F} = (F_1, \ldots, F_K)$. The loss variables have the representation:

$$\begin{aligned} L_C &= \tilde{L}_C(\hat{F}) \\ L_M &= \tilde{L}_M(\hat{F}) \end{aligned}$$

for some functions $\tilde{L}_C(\cdot)$ and $\tilde{L}_M(\cdot)$. Note that some factors may affect both risks and that K could be a very large number. Now it remains to decide upon a multivariate distribution for the factors, by simulating a large number of scenarios and thereby obtaining the distribution of L_{total}.

Copula Approach

The marginal distributions of market and credit risk are well analysed in literature. What remains to be done in the copula approach is to specify the dependence between the two. Usually a Gaussian or student-t copula is used to describe this dependence. The author argues that a normal distribution could be used as marginal distribution for market risk on an aggregated level. As to credit risk, the Vasicek distribution underlying Basel II is proposed as marginal distribution. By calculating Kendall's tau between market and credit risk a normal copula can be calibrated.

Bottom-Up Approach

In a bottom-up approach one tries to model risk on a transaction level, Drehmann et al. (2008) is an example of a bottom-up approach. This type of approach forces one to reconsider the concept on how to measure risk, thus making it hard to compare with existing models.

1.3.2 Corporate Credit Risk Modelling and the Macroeconomy (Carling et al., 2007)

Carling et al. introduce a duration model for the corporate credit risk of a Swedish bank. The probability of default of firm i in the following quarter given that the firm did not default in the previous quarter is calculated conditionally on firm-specific data as well as common macroeconomic data. The authors assume that each firm's probability of default can be described by a common factor shared by each firm scaled by a firm and macroeconomic state factor.

$$PD = P(T = k \mid T > k-1, x_i, z) \approx P(T = k \mid T > k-1)e^{f(x_i, z, \beta_i, \gamma)}$$

where

T is the discrete random number of quarters until firm defaults

x_i is firm specific data, both time-varying and constant

z is common macroeconomic data, both time-varying and constant

$f(\cdot)$ is a linear function

β_i, γ are coefficients of x and z respectively

The bank provided complete history for each customer. The size of the approved credit lines, actual exposure, the rating class and the industry code are cited as the most important. Complementary information of standard balance sheet and income statement data was gathered from the Swedish credit information agency UC. The macroeconomic factors chosen are yield curve, output gap and household expectations. The yield curve is defined as the difference between a 10 year government bond and 3 month treasury bill. The output gap is defined as the difference between real GDP and potential GDP.

The data set contains quarterly observations of Swedish corporations with loans outstanding at the bank between quarter two of 1994 to quarter two of 2002. Three different models were calibrated to the data. Firstly only using accounting data from the credit information agency and data from the bank, secondly also including credit information agency data on payment remarks and finally also including macroeconomic variables. Maximum likelihood estimation is used to calibrate the model to the data. The measure of fit of the model increases when explanatory variables are added, suggesting that macroeconomic variables are important indicators of default risk.

1.3.3 Risk Capital Aggregation (Aas et al., 2007)

Aas, Dimakos & Øksendal introduce a method to aggregate market, credit, business and operational risk on a one-year time horizon.

A set of risk factors that influence market risk is identified and modelled on a daily resolution using a constant conditional correlation GARCH(1,1)-model. The risk factors are a set of stock indexes, bond indexes, exchange rates, hedge fund indexes and real estate indexes. Let $x_t^{\text{market},k}$ be the log-increment of market risk factor k on day t. Define r_t^k to be the change in asset k associated with a specific day t.

$$r_t^k = \begin{cases} 1 - \prod_{s=t+1}^{t+\delta_k} \exp(x_s^{\text{market},k}) & \text{if class k consists of only long positions} \\ |1 - \prod_{s=t+1}^{t+\delta_k} \exp(x_s^{\text{market},k})| & \text{if class k consists of long and short positions} \end{cases}$$

The market risk is measured not in terms of actual exposures but in terms of risk limits set by the banks management.

$$L_t^{market} = \sum_{k=1}^{K} L_t^k = \sum_{k=1}^{K} E^k r_t^k$$

where

L_t^{market} is the total change of value of the market portfolio on day t

L_t^k is the change of value of the portfolio consisting of assets in class k

K is the total number of asset classes

E_t^k is the anticipated maximum exposure to assets in class k defined as the expected utilisation of exposure limits set by managers at the bank.

The market risk on a one year horizon is defined as the maximum of the above total daily changes.

$$L^{market} = \max(\max_t(L_t^{market}), 0)$$

The market risk factors are scaled to a yearly resolution by taking the sum over one year.

The credit risk factor is chosen to be the Norwegian credit loss ratio reported yearly 1989 through 2004, i.e. only 15 data points. Market and credit risk factors on a yearly resolution are assumed to stem from a multivariate normal distribution. A large number of paths of market risk factors over one year are simulated and summed to a yearly resolution. The correlation between each market risk factor and the credit risk factor is estimated, the distribution of the credit risk factor conditional on the market risk factors is then calculated, resulting in an indirect correlation between the risks through the risk factors. The credit risk factor conditional on the yearly market risk factors is then used as the risk factor in a Basel II IRB model.

Operational and business risks are assumed to be log-normal. The log-normal distributions are chosen so that its quantiles correspond to regulatory risk measures. Operational and business risks are then aggregated with market and credit risk using a normal copula with correlation parameters based on expert knowledge. The authors point out that this is not an optimal method, however since not enough data is available and no obvious risk factors could be identified this approach is the best available method.

1.3.4 A general approach to integrated risk management with skewed, fat-tailed risks. (Rosenberg & Schuermann, 2006)

Rosenberg & Schuermann compare a copula-based approach to risk aggregation with three conventional approaches.

Market risk is defined as annualized trading revenue divided by trading assets. Three risk factors are identified for market risk; equity returns, currency returns and interest rate changes. The equity return is the log of the S&P500 return, the interest rate measure is the log-difference of the 10-year Constant Maturity Treasury rate. The currency return is the log difference of the trade-weighted currency index. Data for these risk factors are available daily from 1 January 1974 to 31 December 2002.

Credit risk is defined as the annualized net interest income less provisions. Two risk factors are identified for credit risk; changes in AA and BBB credit spreads. Data for these risk factors are available monthly from December 1988 to December 2002.

The risk factors are modelled with a multivariate GARCH(1,1) model, a trivariate GARCH for market risk factors and a bivariate GARCH for credit risk factors.

The authors use a panel of quarterly data from regulatory reports for 17 bank holding companies from quarter 1 1994 to quarter 4 2002. Credit risk and market risk is regressed on their respective risk factors using OLS. The volatilities of the market risk factors are also used as explanatory variables for market risk due to the fact that commission based trading revenue is affected by market volatility. Since market risk factors are on a daily resolution and credit risk factors are on a monthly resolution they have to be scaled to a quarterly resolution before the regression.

The GARCH model is used to simulate a large number of paths over one year for the two risk types. Taking the sum of each path and using the $\hat{\beta}$ estimated by OLS one obtains the marginal distributions for the risks. This large data sample can then be used to calibrate a normal copula and thus the joint distribution of the risks. The above approach is compared to more conventional approaches, namely;

1. The different risks have the same distribution and are perfectly correlated, i.e. the portfolio VaR is the sum of the weighted VaR for each risk.

2. The different risks have the same elliptical distribution

3. The different risks have the same normal distribution.

The authors show that assumption 1 results in the highest portfolio VaR and assumption 3 the lowest. Assumption 2 and the copula approach results in portfolio VaR of approximately the same size. Assumption 2 leads to easier computations and the authors argue that it is a good approximation of the copula method.

1.3.5 The integrated impact of credit and interest rate risk on banks: An economic value and capital adequacy perspective (Drehmann et al., 2008)

The aggregation of interest risk and credit risk is made using an "bottom up" approach, i.e. the two risks are modelled using a common set of macroeconomic factors. The economic value, EVA, of an asset with maturity at time T is the risk adjusted discounted value of future coupon payments and principal. The EVA of all assets at time t must be larger than the face value, FVL of the assets at time t. The expected risk premium and expected yield curve depend on the same set of macroeconomic risk factors. As these macroeconomic factors change so will the discounted value of future payments but since these remain fixed up until repricing, a asset liability mismatch will arise. The risk weighted assets, RWA, are calculated either by the Basel I or Basel II IRB framework. The share holder funds, SF of the bank must be of a certain percentage of the RWA for the bank to fulfil capital adequacy conditions. The regulatory minimum of 4% is used to calculate capital adequacy conditions. The aove can be summarized in the following two conditions:

- Condition 1 - Economic value

$$\mathrm{EVA}_t > \mathrm{FVL}_t$$

- Condition 2 - Capital Adequacy

$$\frac{\mathrm{SF}_t}{\mathrm{RWA}_t} > k$$

Credit risk is modelled using a latent variable model with level, slope and curvature as the latent variables and output gap, inflation and bank rate as the observable variables. The PDs are linked to the macroeconomic factors using a model developed by Bank of England (Bunn et al., 2005), where the PDs of households and corporations are modelled as a linear combination of macroeconomic factors.

Further the model estimates the risk of the bank itself changing credit rating resulting in a shift of interest paid on liabilities.

1.3.6 Inter-Risk Correlation within Economic Capital (Larsson, 2009)

As the title suggests inter-risk correlation is studied in this masters thesis. The risks studied are credit, market and business risk.

Credit risk is quantified by the actual gross loss divided by the outstanding amount. Credit risk is divided into eight subgroups, by region (the four Nordic countries) and customer type (household or corporate). Macroeconomic factors used to model credit risk are; GDP, inflation, unemployment rate, interest rate, household's savings and debts, output gap and household consumption.

Market risk is quantified by the actual quarterly market loss divided by a quarterly VaR measure at confidence level of 99%. Three different market risks are identified; interest rate risk, equity risk and exchange rate risk. Common macroeconomic factors used to model all market risks are; GDP and output gap. Interest rate risk, equity risk and exchange rate risk are further modelled by short and long term bond indices, stock indices and four different exchange rates respectively.

The correlation between macroeconomic factors and credit risk and market risk is then evaluated using different time lags in order to find the time lag with the highest correlation. Once the optimal time lag has been decided upon, the risks are mapped to the macroeconomic factors using OLS and each explanatory factor is tested for significance.

The time period for which data for all risks exists is from 2002 through 2008. The relationships between risks and macroeconomic factors in this time period is used to historically simulate the time period with the highest correlation between risk types. All risks are assumed to be of the same elliptical distribution and the total inter-risk correlation is calculated using a linear Pearson correlation matrix.

1.3.7 Risk Aggregation and dependence modelling with copulas (Svensson, 2007)

Svensson (2007) investigates risk aggregation using copulas. Firstly the individual market risks, interest rate risk and equity risks are aggregated using a copula. Then all risks are aggregated using four different methods (summation, variance covariance, Monte-Carlo and copula). The Monte-Carlo and copula approach are shown to give very similar results, this since normal marginal distributions are assumed and a gaussian copula is used. The variance/covariance method is shown to over-estimate the economic capital as does the summation method.

Chapter 2

Theoretical Background

In this chapter ordinary least squares estimation is briefly explained, along with measures of goodness of fit and confidence intervals of estimated parameters. Further, principal component analysis is explained.

2.1 Regression model

2.1.1 Ordinary least squares estimation

The dependent variable, y is assumed to be explained by the explanatory variables, $x_1, \ldots, x_n$ through the equation:

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \epsilon \tag{2.1}$$

The residual ϵ represents factors other than the explanatory variables that affect y. β_0 is the intercept. The residual is assumed to be uncorrelated with the explanatory variables and have zero mean:

$$E(\epsilon) = E(\epsilon \mid x_1, \ldots, x_n) = 0$$

From m observations one creates the set of equations:

$$y_j = \beta_0 + \sum_{i=1}^{n} \beta_i x_{i,j} + \epsilon_j,\ j = 1, \ldots, m \tag{2.2}$$

In ordinary least square (OLS) estimation, the parameters $\beta_1, \ldots, \beta_m$ are estimated by $\hat{\beta}_1, \ldots, \hat{\beta}_m$ chosen so that the sum of squared residuals is minimized. OLS estimation is an unbiased estimation. For a more detailed discussion please see Wooldridge (2003).

2.1.2 Goodness of fit

There are numerous ways to measure how good a model fits the data. In this section the ones used in this master's thesis are presented.

R^2 and adjusted R^2

R^2 measures the fraction of the sample variation in y that is explained by x (Wooldridge, 2003). It is defined as:

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}$$

where n is the number of observations and $\hat{y}_i$ is the OLS estimate of y_i. OLS estimation minimizes the R^2 per definition. R^2 never decrease when an extra explanatory variable is added. An alternative measure of goodness of fit is the adjusted R^2, $\bar{R}^2$, which impose a penalty for adding more explanatory variables to a model.

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

where n is the number of observations and k is the number of explanatory variables.

T-test

If the variance of one $\hat{\beta}_i$ in the OLS estimation is estimated by the sample variance s^2, interval estimates according to the student-t distribution can be calculated. Using these one can test if that particular $\hat{\beta}_i$ has the value zero in its interval estimate at some significance level. If this is found to be true, one must consider removing the corresponding explanatory variable from the linear model. The t-statistic is given by:

$$t_n(\beta) = \frac{\hat{\beta} - \beta}{s(\beta)^2} \tag{2.3}$$

The confidence interval for β is given by:

$$I_\beta = [\hat{\beta} - t(\beta)s(\hat{\beta}), \hat{\beta} + t(\beta)s(\hat{\beta})]$$

Non-parametric Bootstrap of points

Using the original sample pairs $((y_1, \mathbf{x_1}), \ldots, (y_n, \mathbf{x_n})$ one creates a new sample by drawing randomly with replacement from the original sample. Each element is drawn with probability $1/n$. The resulting sample is called the bootstrap-sample, denoted by $((y_1^*, \mathbf{x_1}^*), \ldots, (y_n^*, \mathbf{x_n}^*)$.

Using the bootstrap sample a new estimation of $(\beta_1, \ldots, \beta_n)$ is made, $\hat{\beta}^*$. This is repeated B times, where B is a large number, resulting in bootstrap samples; $\hat{\beta}_1^*, \ldots, \hat{\beta}_B^*$. From this bootstrap distribution the statistic of interest can be calculated.

In particular, confidence intervals at confidence level α for $(\hat{\beta}_1, \ldots, \hat{\beta}_n)$ can be estimated. The bootstrap samples are sorted, denoted $\hat{\beta}_{[1]}^*, \ldots, \hat{\beta}_{[B]}^*$, where $[i]$ denotes the i:th largest sample. The confidence interval for $\hat{\beta}_i$ is given by:

$$C_i = [\beta_{[B(\alpha-1)/2],i} \; \beta_{B(\alpha+1)/2],i}]$$

2.2 Stepwise Regression

To choose the "best" explanatory variables in an OLS regression, Draper & Smith (Draper & Smith) believe the Stepwise regression to be a good procedure. It uses partial F-test repeatedly to determine which explanatory variables to use in the model. A partial F-test is a measure of how adding an explanatory variable, to a regression in which it is not already included, affects the regression. The order in which these tests are performed is of significance since if a test is performed on an explanatory variable highly correlated with one that is all ready included in the test the effect of adding the extra variable is very small. The F-test is performed by using the t-statistic in equation 2.3. The stepwise regression is performed according to the following algorithm:

1. Include the explanatory variable with the highest correlation with the dependent variable.

2. Check if this explanatory variable is significant or not by an partial F-test at inclusion level α. If it is not, reject the model and use the model $\hat{Y} = mean(y)$ i.e. the mean of the observations of the dependent variable. If it is found to be significant, continue to the next step.

3. Calculate the new partial F-statistics on all variables not in the equation and include the one with the highest value.

4. Examine the partial F-statistics of the variables already included. If any of these fall below the decided removal significance level, remove it.

5. Perform step 3-4 until equilibrium is reached. Note that one variable may be moved out and then moved in again since the partial F-statistic changes according to which explanatory variables that are included in the regression.

2.3 Principal component analysis

Assume that a p-dimensional vector $\mathbf{x}$, where p is large, is to be studied. Principal component analysis can be used to reduce the dimension of $\mathbf{x}$ by removing redundant information. This is achieved by transformation of $\mathbf{x}$ to a new set of ordered variables, the principal components, where the first few contain information of most of the variation of all variables in $\mathbf{x}$.

A linear combination, $a_1^T\mathbf{x}$, of the elements of x having maximum variance is found, this is the first principal component. Next, $a_2^T\mathbf{x}$ is found by the same criteria as above but such that $a_1^T\mathbf{x}$ and $a_2^T\mathbf{x}$ are uncorrelated. This is the second principal component. There a p principal components to be found, where each new principal component is uncorrelated to all other previous principal components.

Denote the sample covariance matrix of $\mathbf{x}$ S. It can be shown, see for example Jolliffe (2002), that the kth principal component is given by $Z_k = \alpha_k^T x$ where α_k is the eigenvector of S corresponding to its kth largest eigenvalue λ_k.

By studying the weights in $\mathbf{a}_k$ a greater understanding of which components that are driving the process can be achieved.

Chapter 3

Data

In this chapter data is explained and presented. Swedbank data is presented along with key historical changes in Swedbank's company structure. Loan losses and lending by segments and regions are discussed and corresponding loss ratios are calculated. Swedish and Baltic macroeconomic data are also presented and studied. All data used is public. This limits the available data but allows for interested readers to recreate the model.

3.1 Swedbank Data

The following notations are used for lending:

$$L^r_{i,t} = \text{Lending to region } i \text{ in end of period } t,\ i = (Sweden, Baltic)$$
$$L^s_{j,t} = \text{Lending to segment } j \text{ in end of period } t$$

The following notations are used for credit losses:

$$CL^r_{i,t} = \text{Credit losses in region } i \text{ in period t}$$
$$CL^s_{j,t} = \text{Credit losses in segment } j \text{ in period t}$$

The segments are shown in table 3.2

3.1.1 History of Swedbank

To understand the Swedbank data, acquisitions and mergers in Swedbank's history are studied. In table 3.1 the important historical events are listed.

In 1997, Föreningsbanken and Sparbanken merged to form Föreningssparbanken. Föreningssparbanken aqcuired more than 50% of Hansabank late 1999. In 2006, Föreningssparbanken changed name to Swedbank.

Föreningsbanken was formed in 1991 by a merger of twelve regional agricultural cooperative banks. Sparbanken was formed in 1992 by a merger of 11 regional savings banks.

1991	12 regional agricultural cooperative banks merged to form Föreningsbanken
1991	Hansabank was founded in Estonia
1992	A merger between 11 regional savings banks forming Sparbanken Sverige
1994	Föreningsbanken was listed on the Stockholm Stock Exchange
1995	Sparbanken Sverige shares are listed on the Stockholm Stock Exchange
1995	Hansabank started operations in Latvia
1996	Hansabank started operation in Lithuania
1997	FöreningsSparbanken formed through merger between Sparbanken Sverige and Föreningsbanken
1997	Hansabank acquired a stake in Hoiupank
1999	Swedbank acquired more than 50% of Hansabank
2002	Hansabank started operations in Russia
2004	Hansabank acquired Kvest bank in Moscow
2005	Swedbank acquired 100% of Hansabank
2006	The Annual General Meeting resolves to change the bank's name to Swedbank
2007	Swedbank enters Ukraine through the acquisition of TAS-Kommerzbank

Table 3.1: Milestones in Swedbank's history

3.1.2 Loan losses and lending by segment

The segments for which loan losses and lending are reported are listed in table 3.2. Segment 8 is the sum of segment 2 through 7. Segment 9 is the sum of segment 1 and 8.

Loan loss and lending data by segment is available on a yearly basis from 1992 through 1998.

Following the merger of Sparbanken and Föreningsbanken in 1997, Föreningssparbanken kept Sparbankens reporting methods, thus yearly data from Sparbanken is used from 1992 through 1995. Föreninssparbankens proforma yearly data is used 1996. A data break can be seen in figure 3.1 this year. From 1997 through 1998 yearly data from Föreningssparbanken is used. For this period, lending is assumed to be constant over a one year period. Quarterly losses are assumed to be one fourth of the yearly loan losses. All yearly data are from year-end reports.

Data for loan losses and lending is available on a quarterly basis from quarter 1 of 1999 until present quarter 2 of 2009, from public quarterly reports.

no.	Segment
1.	Private Customers
2.	Real estate management
3.	Retail, hotels, restaurants
4.	Construction
5.	Manufacturing
6.	Transportation
7.	Other incl. agriculture
8.	Total corporate
9.	Total

Table 3.2: Swedbank's segments

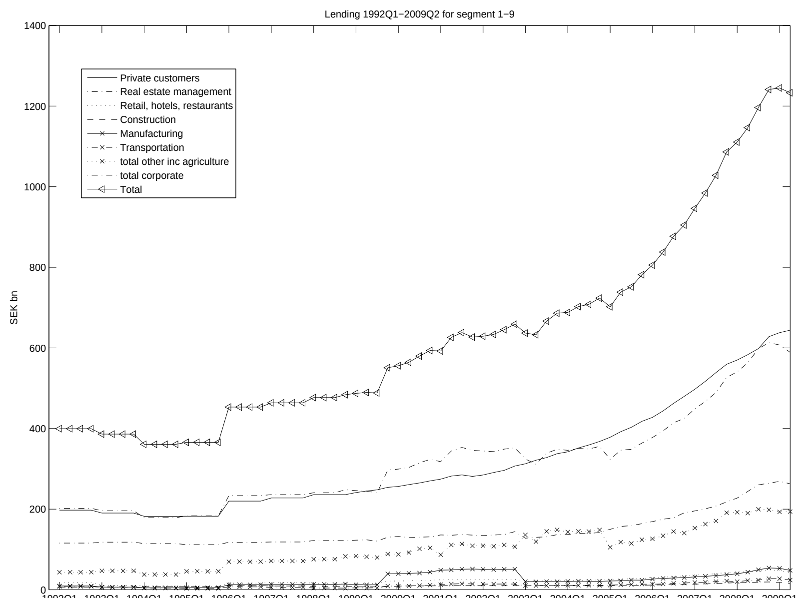
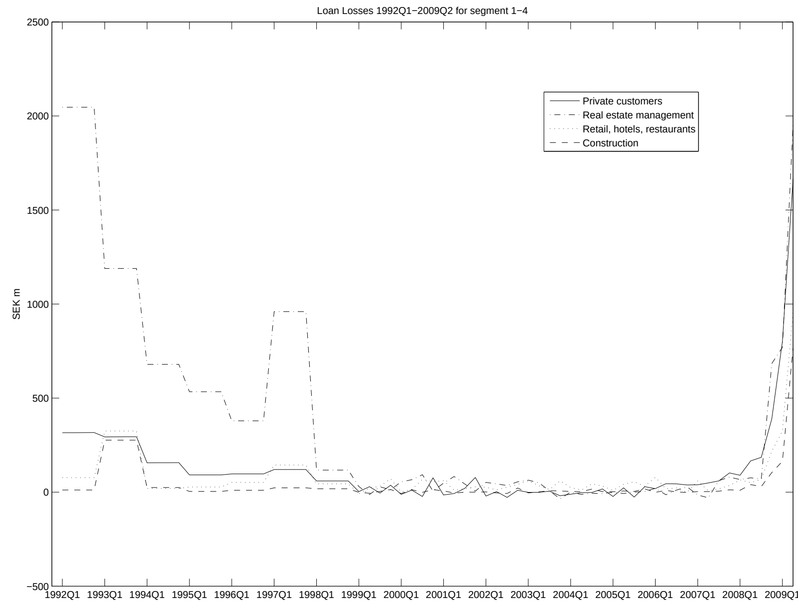


Figure 3.1: Lending by segment. Lending of the whole Swedbank group to the different segments. The merger of Föreningsbanken and Sparbaken can be seen 1996Q1.



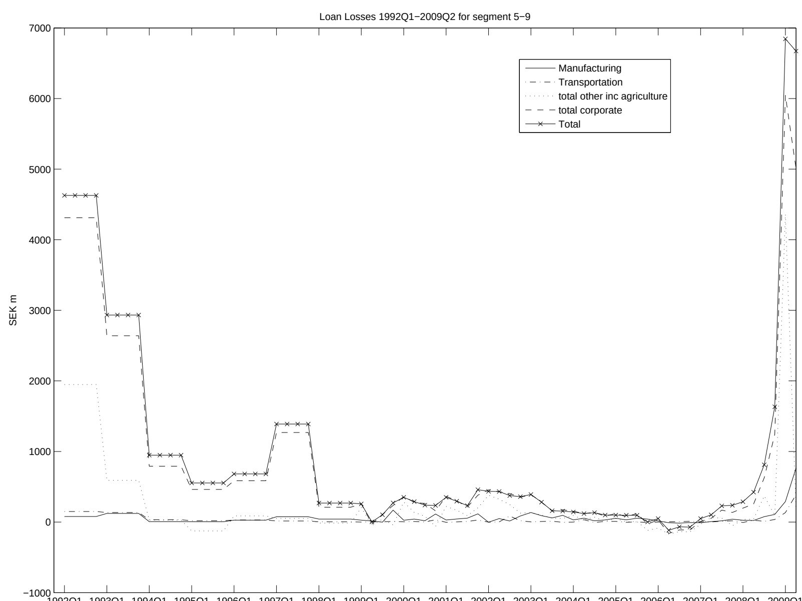


Figure 3.2: Swedbank's credit losses by segment.Credit losses are of the whole Swedbank group to the different segments.

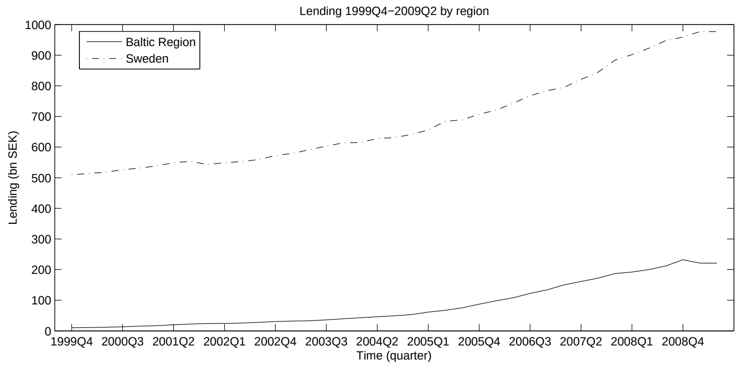


Figure 3.3: Swedbank's lending by region, data available from quarter 4 of 1999 to quarter 2 of 2009

3.1.3 Loan Losses and Lending by region

As of quarter 4 of 1999, when Swedbank acquired more than 50% of Hansapank, lending and loan losses for Sweden and the Baltic region are reported. Since Swedbank is active in Russia lending in Russia is included in the Baltic region. Lending and loan losses are reported excluding repurchase agreements. Recently Swedbank has expanded its operations to Ukraine, however due to the short period of Swedbank's activity in this region, loan losses and lending to Ukraine is not included.

3.1.4 Loss ratios

Loss ratios, $LR^r_{i,t}$ and $LR^s_{i,t}$ are calculated by eqn 3.1 and 3.2 respectively

$$LR^r_{i,t} = \frac{CL^r_{i,t}}{L^r_{i,t-1}} \tag{3.1}$$

$$LR^s_{i,t} = \frac{CL^s_{i,t}}{L^s_{i,t-1}} \tag{3.2}$$

The segments are shown in table 3.2. The regions are, as mentioned before: Sweden and the Baltic. The loss ratios by segment are shown in figure 3.5 and loss ratios by region are shown in figure 3.6. A larger volatility of loss ratios can be seen in the Baltic region. A smaller loan portfolio, and thus smaller diversification effects within the portfolio, is one possible explanation.

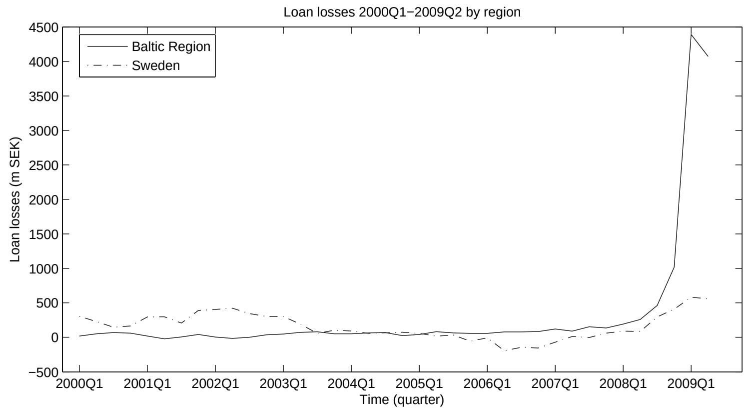


Figure 3.4: Swedbank's credit losses by region, data available from quarter 4 of 1999 to quarter 2 of 2009

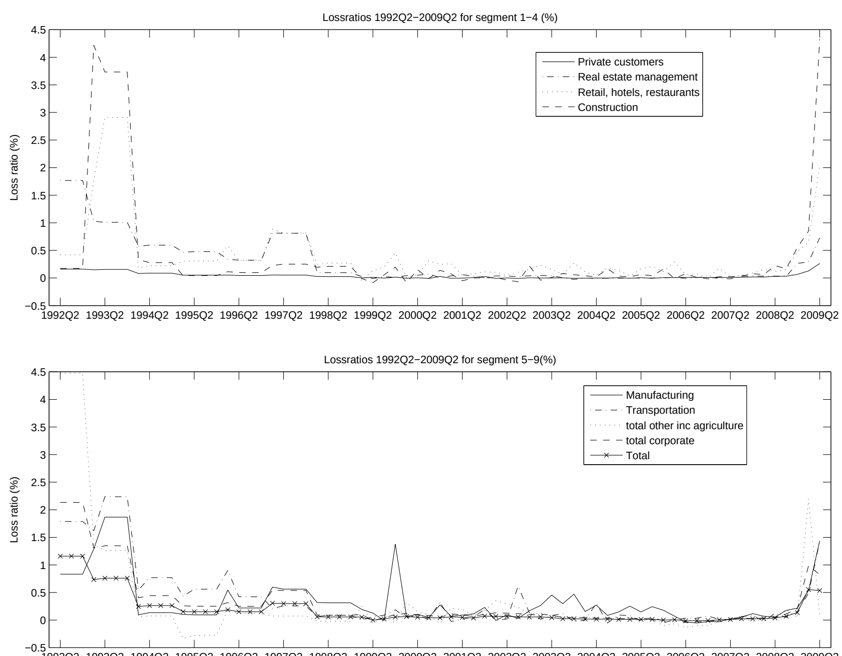


Figure 3.5: Swedbank's loss ratios by segment.

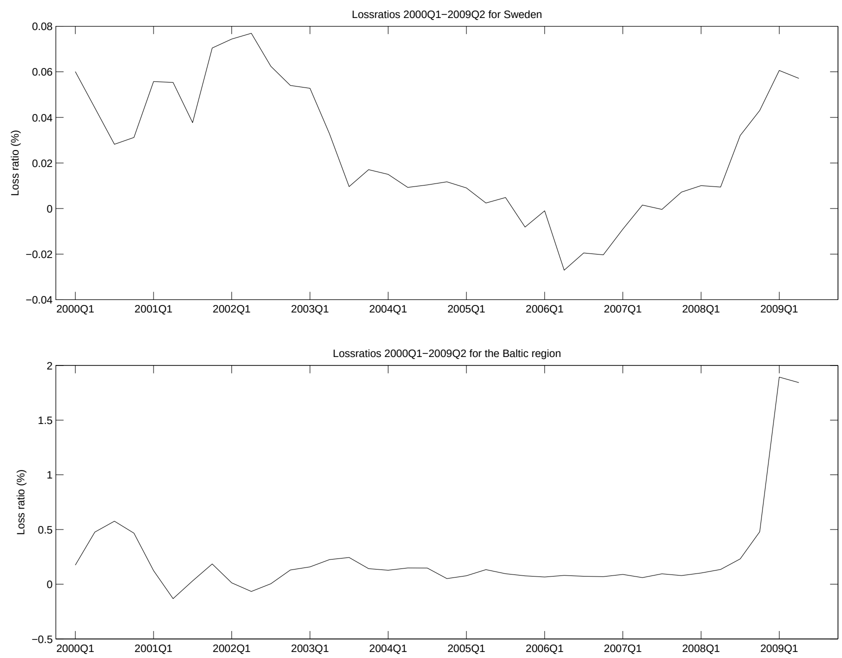


Figure 3.6: Swedbank's loss ratios by region

3.2 Macroeconomic factors

In this section all macroeconomic factors investigated for their level of explanation of loss ratios are presented. Figures of time series of macroeconomic factors and changes thereof are presented in Appendix A.

3.2.1 Swedish macroeconomic factors

In this section the macroeconomic factors for the Swedish region are presented. Seasonal factors are explained and removed. Finally the changes of the Swedish macroeconomic factors are derived and presented.

Swedish Gross domestic product

As can be seen in figure A.1 Swedish gross domestic product, GDP, shows a seasonal behaviour, with a period of one year. This is removed by smoothing moving average, see Brockwell & Davis (2002)

A trend component can clearly be seen, however after taking the difference the time series is assumed to be stationary. The quarterly change of the seasonally adjusted GDP, $G\tilde{D}P$ is calculated as:

$$\Delta G\tilde{D}P_t = \frac{G\tilde{D}P_t - G\tilde{D}P_{t-1}}{G\tilde{D}P_{t-1}} \tag{3.3}$$

where t is time in quarters. The data is gathered from Statistiska Centralbyrån.

Consumer price index

Consumer price index, CPI, serves as a measure of the inflation. It is an index of the cost of a basket of products, calculated so that changes in the basket do not change the index. It is assumed to be non-periodical, but a clear trend can be seen in figure A.3. However the change in CPI is assumed to be a stationary time-series, where the quarterly change is calculated in line with equation 3.3. The data is gathered from Statistiska Centralbyrån.

Swedish unemployment rate

Statistiska Centralbyrån has recently changed its definition of unemployment to that recommended by the European Union. However the older definition is available further back in time and will thus be used in this thesis. The difference between the two is that full-time students are regarded as unemployed in the new definition. Unemployment has a periodical part, which is removed in line with that of GDP, i.e. using smoothing moving average. The data is gathered from Statistiska Centralbyrån.

Swedish 3 month treasury bill

A Swedish 3 month treasury bill is a short-term debt instrument issued by the Swedish National Debt Office. It is issued by a competitive bidding process. A treasury bill is considered to be liquid and riskless. Data is gathered from Sveriges Riksbank and is calculated as the mean of daily quotations of the yield of each quarter.

Swedish 10 year government bond

Swedish 10 year government bonds are issued by the Swedish National Debt Office to finance long term financial needs. The data is gathered from Sveriges Riksbank and is calculated as the mean of the yield over each quarter.

Swedish TCW index

The TCW (Total Competitiveness Weights) index is a method of measuring the value of the Swedish krona against a basket of other currencies. It is possible to see how much the value of the krona has changed by studying this index. A high value on the index means that the krona has weakened and a low value means it has strengthened (Riksbanken, 2009)

Business confidence indicator

The business confidence indicator is based on a survey conveyed by the Swedish National Institute of Economic Research (NIER) and intends to provide a summary of the situation of a particular industry. Data is available sufficiently far back in time for the two industries of manufacturing and construction. The business confidence indicator takes into account answers of a number of different question relating to new orders, output and employment. The business confidence indicator is calculated in net figure, i.e. the balance between the percentage of firms reporting an increase and those reporting a decrease of a certain variable. The data is already seasonally adjusted by and gathered from the Swedish National Institute of Economic Research.

OMXS30

OMXS30 is a market value weighted stock market index of the 30 most traded stocks on the Stockholm Stock Exchange. The closing price of the last day of each quarter is used. The change of the price is calculated in line with equation 3.3.

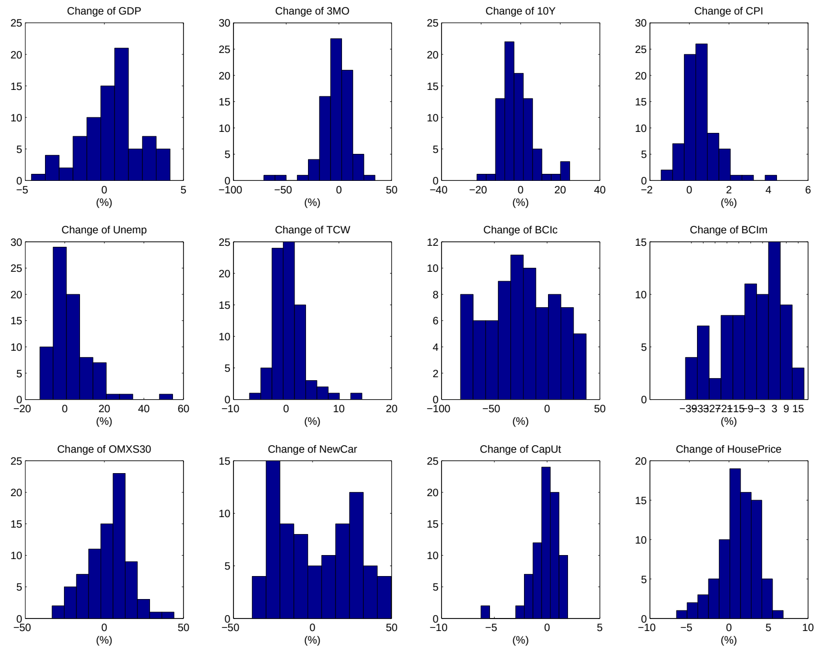


Figure 3.7: Histograms of changes in Swedish macroeconomic factors, 1990Q2-2009Q2.

Capacity utilisation in industry

Capacity utilisation in industry is the actual capacity utilization as percentage of available production capacity in industry. Industry is defined as mining and manufacturing. The data is gathered from Statistics Sweden and is already seasonably adjusted.

New car registrations

New car registrations in Sweden is reported by BIL Sweden monthly, the quarterly figures are simply the sum of these.

Real Estate Price Index

Real Estate Price Index gives the change of price for one- and two-dwelling buildings for permanent living in Sweden. 1981 is the reference year. The data is gathered from Statistics Sweden.

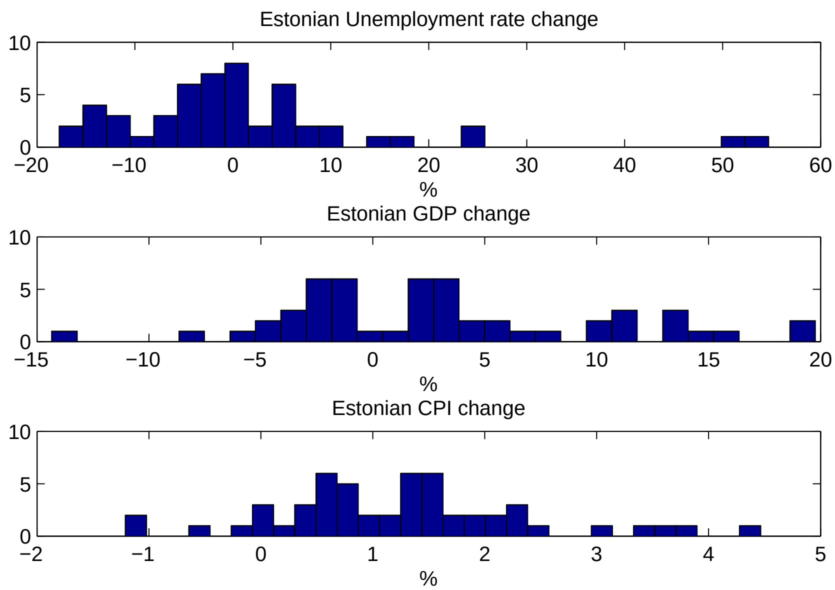


Figure 3.8: Histograms of changes in Estonian macroeconomic factors, 1998Q2-2009Q2

3.2.2 Baltic macroeconomic factors

Due to the lack of macroeconomic data sufficiently far back in time from Latvia and Lithuania, Estonian macroeconomic data will serve as a proxy for the entire Baltic region. The macroeconomic variables believed to capture the state of the economy in the Baltic region are GDP, CPI and unemployment rate.

Quarterly GDP and unemployment rate are reported with a lag of a few months and are currently not available for quarter 2 of 2009. Data is gathered from Statistics Estonia (www.stat.ee).

3.3 Time lags

Losses may occur in the financial reports of Swedbank with a delay. Further, changes in the economic climate may take some time to effect companies and private individuals, i.e. losses may be explained by macro economic factors with a time lag. This will be taken in account by studying the correlation of loss ratios and change of macroeconomic factors at different time lags.

One possible approach is to assume that Swedbank's loan losses are best explained by macroeconomic factors at the time lag with the highest correlation. The correlation of Swedbank's loss ratios of segments and Swedish and Baltic macroeconomic factors will be studied. Correlations between loss ratios of the two regions and their respective macroeconomic factors will also be studied.

One can note that there is a larger lag between Swedish unemployment and loss ratios than that of the Baltic region. This could indicate that there is a higher tolerance to adverse movements in the economy in the Swedish region.

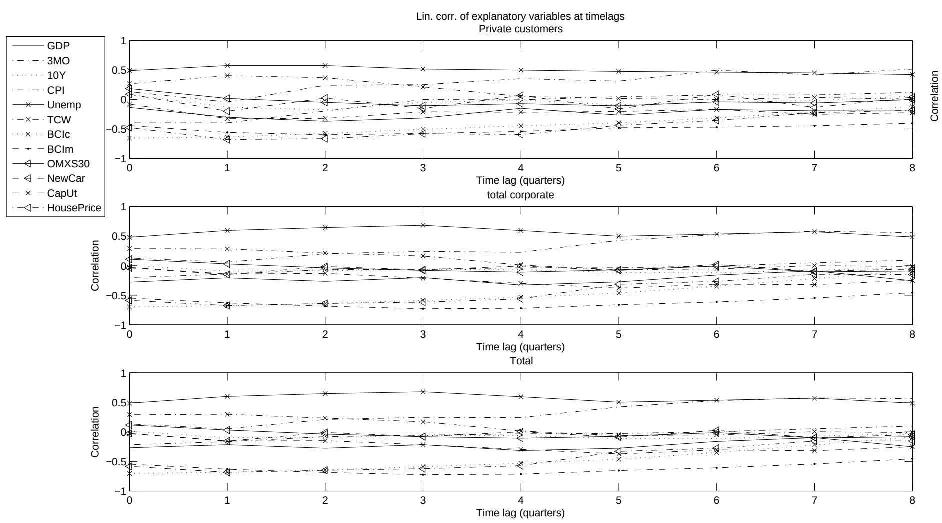


Figure 3.9: Correlation of change in Swedish Macroeconomic factors and Swedbank's loss ratios of segments at different time lags. Correlation is used to study at what time lag loan losses are affected by changes in the macroeconomic factors.

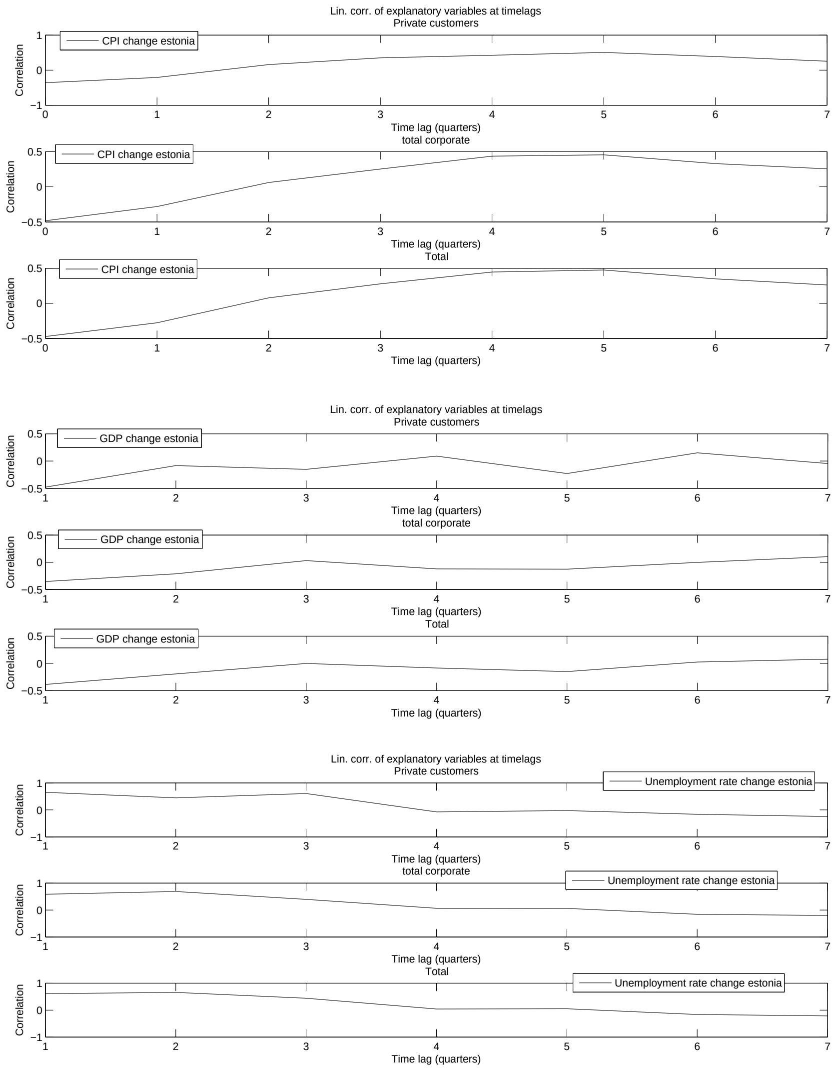


Figure 3.10: Correlation of change of Estonian CPI, GDP and unemployment rate with Swedbank's loss ratios of segments at different time lags. Correlation is used to study at what time lag loan losses are affected by changes in the macroeconomic factors.

	Priv.Cust.	Tot. Corp.	Tot.
Estonia unemp	1	2	2
Estonia GDP	1	1	1
Estonia CPI	5	5	5
Swedish GDP	2	4	4
Swedish 3MO	1	0	0
Swedish 10Y	2	6	5
Swedish CPI	8	7	7
Swedish Unemp	2	3	3
Swedish TCW	1	0	1
Swedish BCIc	0	0	0
Swedish BCIm	2	3	3
Swedish OMXS30	0	4	0
Swedish NewCar	1	1	1
Swedish CapUt	1	5	5
Swedish HousePrice	1	1	1

Table 3.3: Time lags with which macro economic factors have the highest correlation with loss ratios of the different segments. The number of the lags correspond to the peaks in graphs 3.9 and 3.10

.

	Lag
Estonia unemp	2
Estonia GDP	1
Estonia CPI	5

Table 3.4: Time lags with which Estonian macroeconomic factors have the highest correlation with loss ratios of the Baltic region. The number correspond to the peaks in figures 3.12.

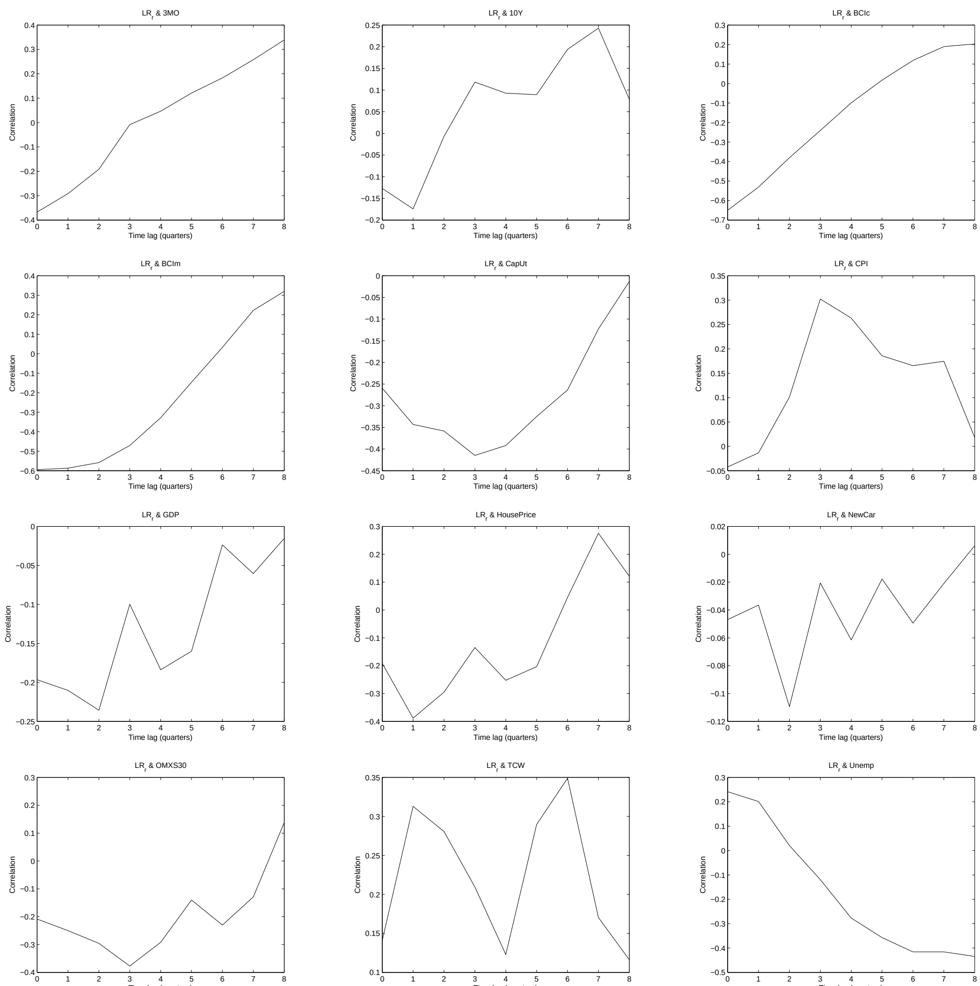


Figure 3.11: Correlation of change of Swedish Macroeconomic factors and Swedbank's total loss ratios in Sweden at different time lags. Correlation is used to study at what time lag loan losses are affected by changes in the macroeconomic factors.

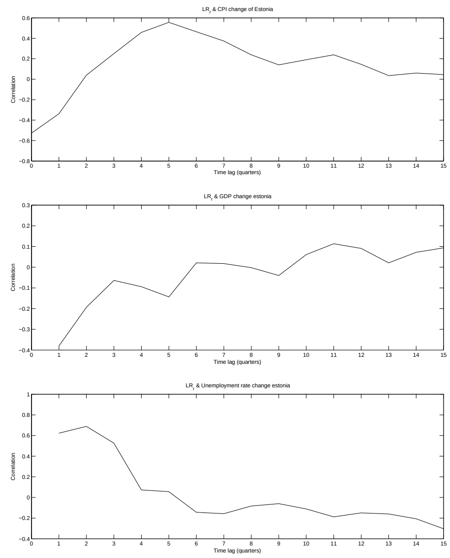


Figure 3.12: Correlation of change of Estonian CPI, GDP and unemployment rate with Swedbank's total loss ratio in the Baltic region at different time lags. Correlation is used to study at what time lag loan losses are affected by changes in the macroeconomic factors. Due to delays in the reporting of Estonian GDP and unemployment rate the smallest time lag of these is one quarter.

GDP	2
3MO	0
10Y	7
CPI	3
Unemp	8
TCW	6
BCIc	0
BCIm	0
OMXS30	3
NewCar	2
CapUt	3
HousePrice	1

Table 3.5: Time lags with which Swedish macro economic factors have the highest correlation with loss ratios of the Swedish region. The number correspond to the peaks in figure 3.11.

3.4 Calculation of principal components

An approach which does not ignore the non-zero correlation of other time-lags but maintaining the small dimensions of maximum correlation approach, is to use principal component analysis. All available time lags are reduced to linear combinations of the macroeconomic factors at different time lags according to section 2.3. PCA will be performed separately on Swedish and Baltic macroeconomic factors.

PCA on Swedish macroeconomic factors Data on changes of Swedish macroeconomic factors are available from 1990Q2 to 2009Q2, 77 observations. Loss ratios by segment are available from 1992Q2 to 2009Q2, 69 observations. The possible lags are thus $(0, 1, \ldots, 8)$. Let the matrix of original observations of swedish macroeconomic factors be denoted X which consists of:

$$X = [X^1\, X^2\, \ldots\, X^{10}\, X^{11}\, X^{12}] \tag{3.4}$$

where X^i is the vector of the observations of the change of the individual macroeconomic factors. A new matrix is formed with all possible lags of the Swedish macroeconomic factors.

$$X_{alllags} = \begin{pmatrix} X_t^1 & X_{t-1}^1 & \cdots & X_{t-8}^1 & X_t^2 & X_{t-1}^2 & \cdots & X_{t-8}^{12} \end{pmatrix}$$

$X_{alllags}$ contains 69 observations of 96 variables. PCA is performed on $X_{alllags}$ and the eight first principal components are found to explain almost 90% of the variation of $X_{alllags}$.

PCA on Baltic macroeconomic factors Data on changes of Estonian GDP and unemployment rate are available from 1996Q2 to 2009Q1, a total of 52 observations. Data on changes of Estonian CPI id available from 1996Q2 to 2009Q21, a total of 53 observations.

Loss ratios for the Baltic region are available from 2000Q2 to 2009Q2, a total of 37 observations. The possible lags are thus $(0, 1, \ldots, 16)$ for estonian CPI and $(1, 2, \ldots, 16)$ for Estonian GDP and unemployment rate. However, it does not seem reasonable that loss ratios should be affected by change in the economic climate more than two years back. The possible lags are thus restricted to $(1, 2, \ldots, 8)$ for GDP and unemployment and $(0, 1, \ldots, 8)$ for CPI, resulting in a matrix of 25 variables (columns) with 38 observations (rows).

PCA is performed on a matrix of all possible combinations of lags above and the 8 first principal components are found to explain almost 94% of the variation of the 25 variables.

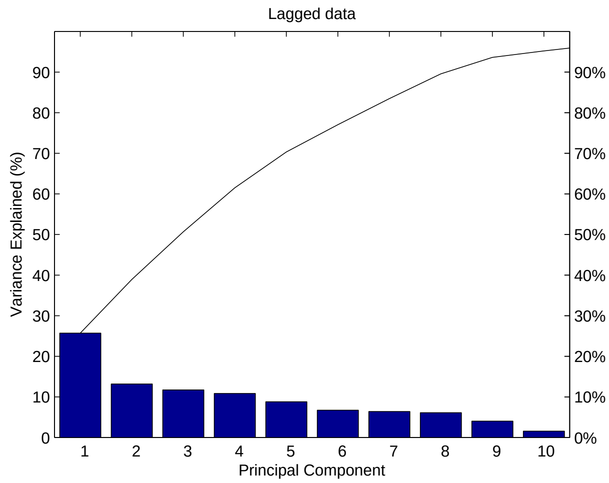


Figure 3.13: Percentage of variance explained by the first 8 principal components when principal component analysis has been performed on Swedish $X_{alllags}$, all possible lagged values of Swedish macroeconomic observations. The first 8 out of 54 principal components explain almost 90% of the total variation.

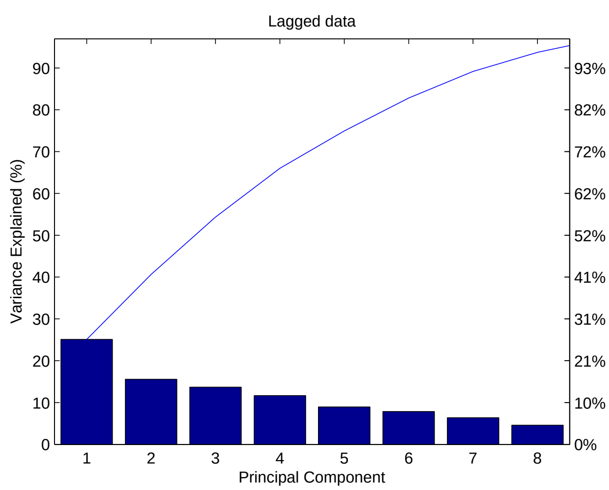


Figure 3.14: Percentage of variance explained by the first 8 principal components when principal component analysis has been performed on all possible lagged values of Estonian macroeconomic observations. These first 8 out of 25 principal components explain almost 94% of the total variation.

Chapter 4

Regression of loss ratios on macroeconomic factors

To study the explanatory power of macroeconomic factors on loss ratios a number of different models are used. Loan losses of the segments "Private Customer", "Total Corporate" and "Total" are regressed on both all Swedish and all Baltic macroeconomic factors. This model is then refined by using stepwise regression.

Loan losses of the two different regions are regressed on their respective macroeconomic factors, this model is also refined by using stepwise regression.

Lastly, the above is repeated but using principal components of macroeconomic factors as explanatory variables.

4.1 Model 1, loss ratios by segment regressed on macroeconomic factors.

First loss ratios of the the segments "Private Customer", "Total Corporate" and "Total" are regressed on all Swedish and all Baltic macroeconomic factors using OLS.

$$LR^s_{j,t} = \beta_0 + \sum_{i=1}^{n} \beta_i X^i_{t-1-\delta_i} + I\left(\sum_{k=1}^{m} \gamma_k Y^k_{t-1-\delta_k}\right) \tag{4.1}$$

$X^i_{t-1-\delta_i}$ are the macroeconomic variables of Sweden at time lag δ_i. $Y^k_{t-1-\delta_k}$ are the macroeconomic variables of the Baltic region at time lag δ_k. The time lags with the largest absolute value of correlation presented in table 3.3 are used. The $\hat{\beta}$:s are reported for three segments; private customers, total corporate and total. Swedbank had no lending, and thus no loan losses, in the Baltic region before quarter 1 of 2000, to account for this an indicator

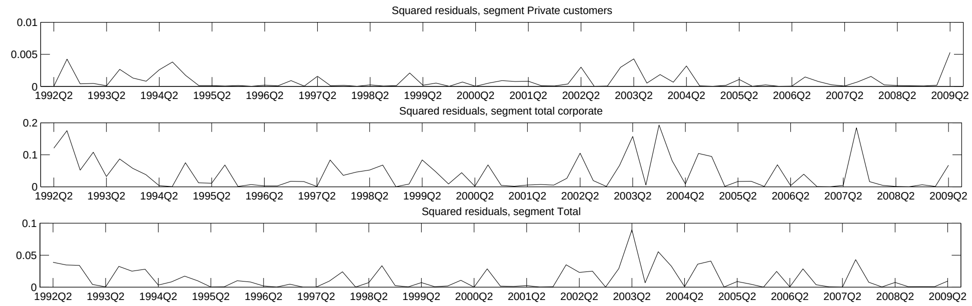


Figure 4.1: Squared OLS residuals of model 1

variable I is used in equation 4.1 which takes the value 0 if $t < 2000Q1$ and 1 otherwise.

The results of this regression can be seen in table 4.1 for the segments "Private Customer", "Total Corporate" and "Total". The residuals of these three models can be seen in figure 4.1. Studying the signs of the estimated beta coefficients one can get a feeling if the regression makes sense. For the segments "private customers" Estonian unemployment at time lag one has a positive coefficient. If the unemployment increased the previous quarter so will the loss ratio, the same is true for Swedish unemployment but with a time lag of two quarters. The estimated coefficient of both countries GDP is negative, which also makes sense. In times of economic contraction loss ratios increase.

Using stepwise regression the number of explanatory variables in the regression explained in equation 4.1 are reduced significantly while keeping the explanatory power at the same level. The explanatory variables chosen by the stepwise regression can be seen in table 4.2. The available macroeconomic factors for the model to choose from are 15, stepwise regression reduces this number significantly to 7 for the segment "private customers", 9 for the segment "Total Corporate" and 6 for the segment "total". The adjusted R^2 decrease marginally for all segments in the stepwise regression.

Variable/Segment:	Private	Total Corporate	Total
Intercept	0,0437⋆	0,1953‡	0,1377⋆
Estonia unemp	0,0008†	0,0092‡	0,0033
Estonia GDP	-0,0016	-0,0063	-0,0063
Estonia CPI	-0,0098	-0,0065	-0,0278
Swedish GDP	-0,0073‡	-0,0063	0,0018
Swedish 3MO	0,0006	0,0055†	0,0012
Swedish 10Y	-0,0002	0,0042	0,0025
Swedish CPI	0,0066	0,2205⋆	0,1220⋆
Swedish Unemp	0,0019‡	0,0131‡	0,0083 ⋆
Swedish TCW	0,0025	0,0293⋆	0,0103†
Swedish BCIc	-0,0004†	-0,0017	-0,0007
Swedish BCIm	0,0001	-0,0087‡	-0,0042†
OMXS30	-0,0001	0,0055‡	-0,0010
Swedish NewCar	0,0001	-0,0030‡	-0,0013
Swedish CapUt	-0,0085‡	-0,0323	-0,0244
Swedish HousePrice	-0,0050†	-0,0601⋆	-0,0273†
R^2	0,7409	0,8576	0,8287
Adj. R^2	0,6675	0,8174	0,7803

Table 4.1: Estimated β values for regression of model 1 on all macroeconomic factors, regressing loss ratios by segment on the macroeconomic factors. † $\Leftrightarrow$ p-value < 0.1, ‡ $\Leftrightarrow$ p-value < 0.05 and $\star \Leftrightarrow$ p-value < 0.01. The p-value is the probability of the hypothesis of that particular coefficient being zero is true.

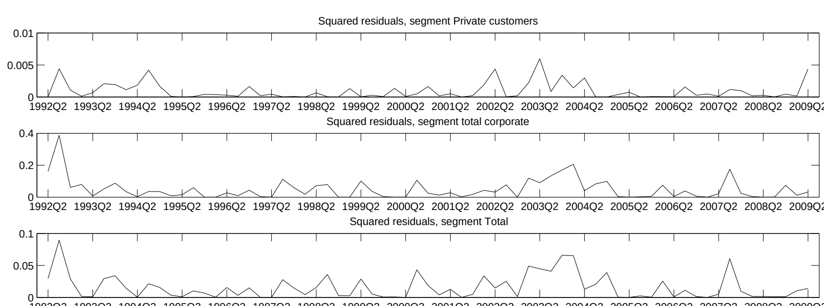


Figure 4.2: Squared residuals of stepwise regression of model 1

Variable/Segment:	Private	Total Corporate	Total
Intercept	0,0395	0,1869	0,1338
Estonia GDP	-0,0021	-	-0,0103
Estonia unemp	-	0,0062	-
GDP	-0,0067	-	-
CPI	-	0,2185	0,1217
Swedish Unemp	0,0017	0,0155	0,0107
Swedish TCW	-	0,0208	-
Swedish BCIc	-0,0006	-	-
Swedish BCIm	-	-0,0105	-0,0046
Swedish OMXS30	-	0,0070	-
Swedish NewCar	-	-0,0032	-
Swedish CapUt	-0,0068	-	-
Swedish HousePrice	-0,0055	-0,0684	-0,0320
R^2	0,696	0,834	0,784
Adj. R^2	0,666	0,812	0,767

Table 4.2: Estimated β values by stepwise regression of model 1, regressing loss ratios by segment on the macroeconomic factors with the highest explanatory power according to stepwise regression. No number given indicates that a variable is not included in model for that particular segment. The hypothesis that one particular coefficient is zero can be rejected at confidence level larger than 95% for all included coefficients.

4.2 Model 2, Loss ratios by region regressed on respective macroeconomic factors

Two different models are calibrated for each region. Firstly a model using macroeconomic factors lagged with the time lag of highest correlation, decided upon in Section 3.3. Secondly, stepwise regression is used to decide which macroeconomic factors to include and at what time lag.

Loss ratios of the two different regions are regressed on the corresponding macroeconomic factors using OLS.

$$
\begin{aligned}
LR^R_{swe,t} &= \beta_0 + \sum_{i=1}^{n} \beta_i X^i_{t-1-\delta_i} \\
LR^R_{balt,t} &= \alpha_0 + \sum_{i=1}^{n} \alpha_i Y^i_{t-1-\zeta_i}
\end{aligned}
\tag{4.2}
$$

$X^i_{t-1-\delta_i}$ are the macroeconomic variables of Sweden at time lag δ_i, $Y^i_{t-1-\zeta_i}$ are the macroeconomic variables of Estonia at time lag ζ_i. The time lags with the largest absolute value of correlation as presented in tables 3.5 and 3.4 is used. The model gives an adjusted R^2 of 0,57 for the Swedish region and 0,55 for the Baltic region. The low values of adjusted R^2 and the high p-values indicates that these models do not fit the data well. However one must remember the high level of uncertainty due to the small sample size.

Instead of choosing the explanatory variables as the macroeconomic factors lagged with that of the maximum correlation one could include macroeconomic factors of all time lags between 0 and 8 in equation 4.2 and use stepwise regression to decide which explanatory variables to include. The estimated β coefficients of included variables by stepwise regression can be seen in table 4.5 for the Baltic region, the high values indicate a high level of explanation of Baltic loss ratios by Baltic macroeconomic variables.

For all Swedish macroeconomic factors at all lags the hypothesis that the corresponding coefficient is zero can't be rejected at confidence level 95%. Due to this fact stepwise regression can not be performed for the Swedish region. This could be explained by Swedish customers tolerating larger adverse movements of the economic climate before not being able to fulfil their obligations towards Swedbank.

The confidence levels should however be used with care. The small data sample of 38 points leads to uncertainties in the measures of goodness of fit as well as general uncertainties in the model.

Variable	$\hat{\beta}$	p-value
Intercept	0,1377	0,1180
Estonia unemp	0,0158	0,0004
Estonia GDP	-0,0140	0,0544
Estonia CPI	0,1248	0,0510
R^2	0,5831	
Adjusted R^2	0,5463	

Table 4.3: Estimated β values by model 2, regressing loss ratios of Baltic region on macroeconomic factors with time lags of maximum correlation.The p-value is the probability that one particular coefficient is zero.

	beta	p-value
Intercept	0,0174	0,0343
Swedish GDP	-0,0016	0,6022
Swedish 3MO	0,0003	0,3193
Swedish 10Y	0,0004	0,4755
Swedish CPI	0,0035	0,7086
Swedish Unemp	-0,0009	0,1152
Swedish TCW	0,0055	0,0114
Swedish BCIc	-0,0005	0,0161
Swedish BCIm	-0,0010	0,0574
Swedish OMXS30	0,0005	0,1599
Swedish NewCar	0,0001	0,5850
Swedish CapUt	-0,0089	0,1134
Swedish HousePrice	-0,0001	0,9835
R^2	0,71	
Adjusted R^2	0,57	

Table 4.4: Estimated β values of model 2, regressing loss ratios of Sweden on Swedish macroeconomic factors with time lags of largest correlation. The p-value is the probability that one particular coefficient is zero.

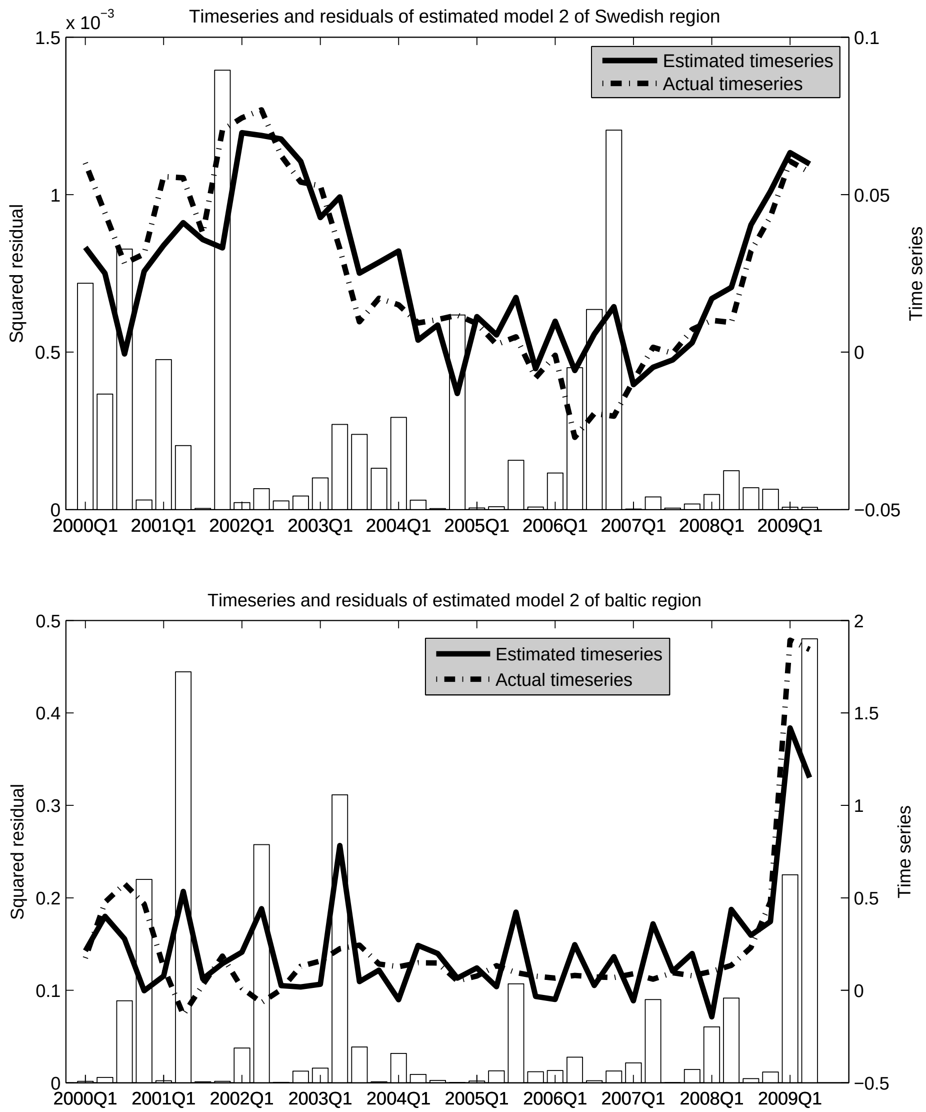


Figure 4.3: Squared OLS residuals of model 2 when regressing on the macroeconomic factors lagged with time lag of largest correlation

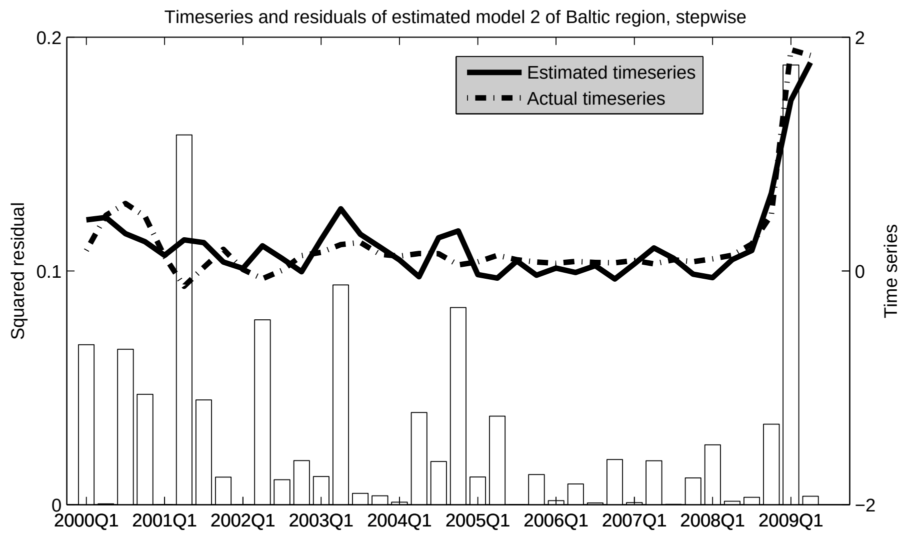


Figure 4.4: Time series of Loss ratios of Baltic region and estimated time series when regressing on the Baltic macroeconomic factors with lags decided by stepwise regression.

Variable	$\hat{\beta}$
Intercept	0,1173
Estonia CPI lagged 7	0,1277
Estonia GDP lagged 5	0,0075
Estonia GDP lagged 6	0,0110
Estonia unempl. lagged 5	0,0171
R^2	0,8213
Adjusted R^2	0,7997

Table 4.5: Estimated β values by stepwise regression of model 2, regressing loss ratios of Baltic region on macroeconomic factors with time lags decided upon by stepwise regression. The hypothesis that one particular coefficient is zero can be rejected at confidence level larger than 95% for all included coefficients.

4.3 Model 3, Loss ratios by segment regressed on principal components

Loss ratios of of Swedbank's segments are regressed on the principal components using OLS.

$$LR^s_{i,t} = \beta_0 + \sum_{i=1}^{n} \beta_i PC_t^{swe,i} + I\left(\sum_{i=1}^{n} \beta_i PC_t^{balt,i}\right) \tag{4.3}$$

where $PC_t^{swe,i}$ and $PC_t^{balt,i}$ are the principal components calculated in section 3.4. I is an indicator function which takes the value 1 if $t > 1999Q4$ and zero otherwise. The explanatory power of this regression is summarized in table 4.6.

Variable/Segment	Private	Corporate	Total Corporate
R^2	0,79929	0,68506	0,69899
Adjusted R^2	0,73753	0,58815	0,60637

Table 4.6: R^2 and adjusted R^2 for model 3, regression of loss ratios by segment on principal components of macroeconomic data lagged all possible time steps.

4.4 Model 4, Loss ratios by region regressed on the corresponding PC:s

Loss ratios of the two different regions are regressed on the corresponding principal components of macroeconomic factors lagged all possible time steps.

$$\begin{aligned} LR^R_{swe,t} &= \beta_0 + \sum_{i=1}^{n} \beta_i PC_t^{swe,i} \\ LR^R_{balt,t} &= \alpha_0 + \sum_{i=1}^{n} \beta_i PC_t^{balt,i} \end{aligned} \tag{4.4}$$

The regression is very poor for the Swedish region and he hypothesis that all coefficients are zero can not be rejected. For the Baltic region the regression has a very high level of explanation, the adjusted R^2 is 0,78.

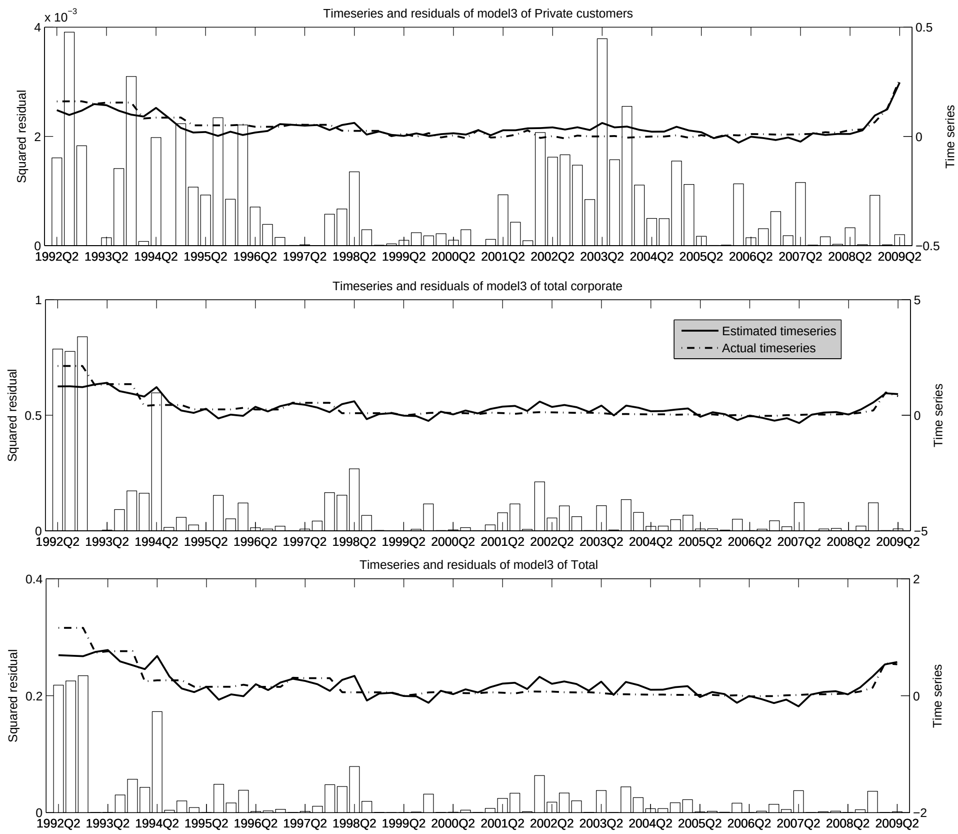


Figure 4.5: Timeseries of actual and estimated LR by OLS and Squared residuals of the three segments "Private", "Corporate" and "Total" using model 3. The residuals are of the same magnitude as model 1 but captures the spike of loss ratios the first quarters of 2009 better.

R^2	0,243241667
p-value	0,352732348
estimate of error variance	0,000814834
Adjusted R square	0,034480748

Table 4.7: Statistics of descriptive power of model 4 for the Swedish region. The p-value is the probability that all coefficients are zero. The large p-value means that the hypothesis that all coefficients are zero can not be rejected. Larger estimate of error variance, much smaller R^2 and adjusted R^2 is almost zero. Using the original macroeconomic variables with the time lag with the highest correlation leads to a better model.

R^2	0,833565875
p-value	2,29297E-09
estimate of error variance	0,036775933
Adjusted R square	0,787653012

Table 4.8: Statistics of descriptive power of model 4 for the Baltic region. The p-value is the probability that all coefficients are zero. R^2 and adjusted R^2 are larger suggesting a model with a better fit than model 2 regressing on macroeconomic factors with the largest correlation. The explanatory power of this model and the explanatory power of the model 2 with time steps chosen by stepwise regression is of the same magnitude. The loss ratios of the Baltic region is better described by macroeconomic factors than the Swedish region suggesting a lower tolerance to adverse movements in the economic climate.

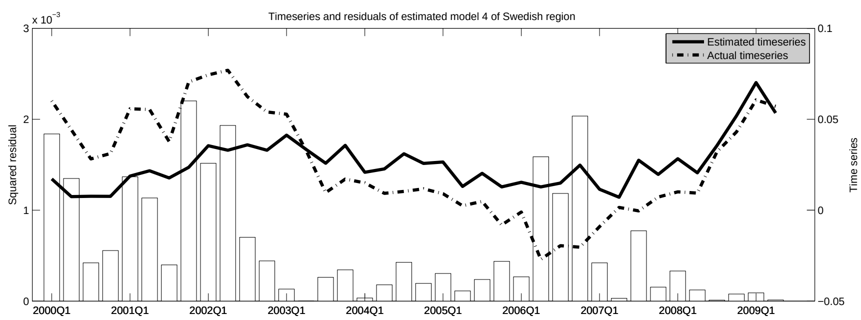


Figure 4.6: Timeseries of actual loss ratios and estimated loss ratios and the resulting Squared OLS residuals of model 4 for the Swedish region.

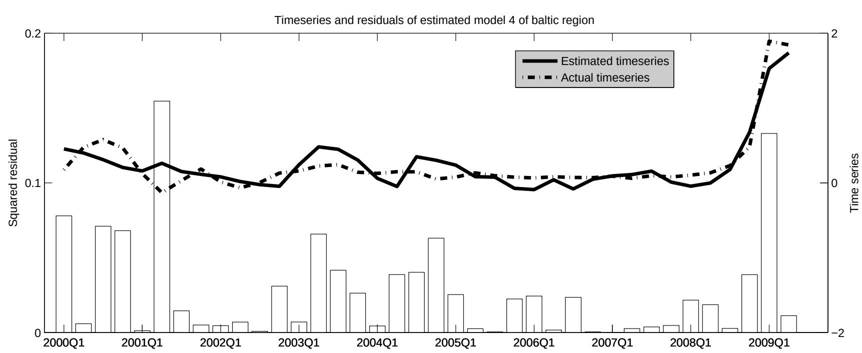


Figure 4.7: Timeseries of actual loss ratios and estimated loss ratios and the resulting Squared OLS residuals of model 4 for the Baltic region

4.5 Summary of regression models

Loan losses for the Baltic region are equally well modelled by model 2 using stepwise regression and model 4 using principal component analysis. Choosing explanatory variables by taking the lags with largest correlation leads to a model with smaller explanatory power.

Loan losses for the Swedish region are very poorly explained by principal components of all time lags of Swedish macroeconomic factors. Further choosing which macroeconomic factors and at what time lag time lags should be included by stepwise regression is not possible due to the fact that confidence intervals of the corresponding β coefficients includes zero at a 95 % confidence level.

As to the loan losses of the segment "Private Customers" regressing on principal components of all factors of all possible time lags has good explanatory power. Choosing explanatory variables by taking the macroeconomic factors lagged according to highest correlation leads to a marginally lower explanatory power. Choosing which macroeconomic factors to use by stepwise regression also leads to marginally lower explanatory power.

The reverse is true for loan losses of the segments "Total Corporate" and "Total". Choosing explanatory variables by taking the macroeconomic factors lagged according to highest correlation and by stepwise regression leads to good explanatory power. Regressing on principal components of all macroeconomic factors of all possible time lags leads to a model which have significantly lower explanatory power.

Chapter 5

Modelling of macroeconomic factors

A model of the macroeconomic factors will be used to predict the loss ratios of Swedbank. Macroeconomic factors from the two regions, Sweden and the Baltic region, will be modelled separately. Using the estimated model the distribution of macroeconomic changes will be simulated and using the regression models the future credit losses of Swedbank can be estimated. There are almost 80 observations of the Swedish macroeconomic factors, and about half of that for the Baltic region which will lead to prediction uncertainties in both models.

5.1 Model of Swedish Macroeconomic factors

Two separate models for the macroeconomic factors of the Swedish model are to be calibrated. Firstly a model of the macroeconomic factors. Secondly a model of the principal components of the macroeconomic factors of all possible time lags is to be calibrated.

5.1.1 Modelling of Swedish Macroeconomic factors

A multivariate autoregressive model is to be calibrated from the observations of Swedish macroeconomic factors. In the same way that loan losses depend on macroeconomic factors with a time lag, the inter-dependence between Swedish macroeconomic factors may differ at different time lags. As can be seen in the left figure of 5.1 the correlation between unemployment and GDP seems close to zero. In the right figure of 5.1 a scatter plot of Swedish unemployment at time t and Swedish GDP at time $t-1$ is shown. Here one can see a non-zero correlation, i.e. one must consider dependencies over time. A multivariate autoregressive model does this. The vector of Swedish macroeconomic factors at time t, X_t^{swe} comes from an AR-process of order

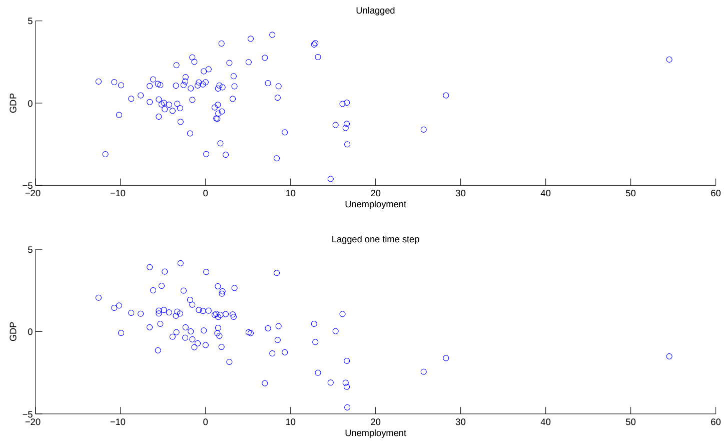


Figure 5.1: Top figure is a scatter plot of Swedish GDP and unemployment. Bottom is of Swedish unemployment at time t and Swedish GDP at (t-1)

p if it can be written as:

$$X_t^{swe} = \Theta_1 X_{t-1}^{swe} + \ldots + \Theta_p X_{t-p}^{swe} + Z_t \tag{5.1}$$

$$Z_t \sim WN(\mathbf{0}, \Sigma) \tag{5.2}$$

where

X_t^{swe} is a vector of the Swedish macroeconomic factors at time t

Θ_i is a matrix of coefficients of previous observations of Swedish macroeconomic factors

Z_t is a vector of white noise random variables

Σ is the covariance matrix of Z_t

What needs to be estimated in the multivariate AR model is the distribution and covariance matrix of Z_t, the matrices $\Theta_1, \ldots, \Theta_p$ and p itself. Since we have removed seasonal components of the Swedish macroeconomic time series and model the change of the macroeconomic factors, stationarity is assumed. Under this assumption the auto-covariance matrix Γ of the macroeconomic process can be written as:

$$\Gamma(t+h, t) = \Gamma(h) = \begin{pmatrix} \gamma_{1,1}(h) & \cdots & \gamma_{1,n}(h) \\ \vdots & \ddots & \vdots \\ \gamma_{n,1}(h) & \cdots & \gamma_{n,n}(h) \end{pmatrix} \tag{5.3}$$

where $\gamma_{i,j}(h)$ is the covariance between macroeconomic variable j and i at time lag h and n is the number of macroeconomic variables, in this case 12. An estimator of $\Gamma(h)$ is

$$\hat{\Gamma}(h) = \begin{cases} \frac{1}{n}\sum_{t=1}^{n-h}(\mathbf{X}_{t+h} - \hat{\mu})(\mathbf{X}_t - \hat{\mu})^T & \text{for } 0 \leq h \leq n-1 \\ \hat{\Gamma}(-h) & \text{for } -n+1 \leq h \leq 0 \end{cases} \tag{5.4}$$

The Yule-Walker estimates for the parameters in equation 5.1 are obtained by multiplying equation 5.1 with $\mathbf{X}_{t-j}^T$, taking expectations and using the sample covariance.

$$\hat{\Sigma} = \hat{\Gamma}(0) - \sum_{j=1}^{p} \hat{\Theta}_j \hat{\Gamma}(-j) \tag{5.5}$$

$$\hat{\Gamma}(i) = \sum_{j=1}^{p} \hat{\Theta}_j \hat{\Gamma}(i-j) \quad i = 1, \ldots, p \tag{5.6}$$

The order p is chosen by Schwarz's Bayesian Criterion (Neumaier & Schneider, 2001). From a set of plausible orders, $p_{min}, \ldots, p_{max}$, the optimal order, p_{opt} is found. In this calibration p_{min} is set to 1 and p_{max}, is set to 4, macroeconomic factors are thus assumed to be affected by earlier macroeconomic factors for a maximum of one year. A larger value of p_{max} is not possible due to the limited number of observations. Calculations in line with Neumaier & Schneider (2001) show that the p_{opt} maximizing the Schwarz's Bayesian Criterion is $p_{opt} = 1$. A multivariate autoregressive model of order one is thus the best fitting model for the data. To solve the system of equations in 5.5 it thus suffices to estimate $\hat{\Gamma}(0), \hat{\Gamma}(1)$ and $\hat{\Sigma}$. Once the matrix $\hat{\Theta}_1$ is estimated the residuals Z_t can be calculated. As stated before the residuals are assumed to be white noise but it remains to decide which type of distribution they stem from. In figure 5.2 histograms of the twelve components of Z_t are shown.

In figure 5.3 qq-plots against a normal distribution are shown for all components of the residual vector. In figure 5.4 qq-plots of the residuals with a t-distribution with 1 degree of freedom as reference distribution can be seen. In figure 5.5 qq-plots of the residuals with a t-distribution with 5 degrees of freedom as reference distribution can be seen. The qq-plots with this reference distribution are approximately linear which suggest that a multivariate t(5) distribution can be used in the simulation. The decision has been made by eye-ball investigation. The dependence structure between the variables is based on the estimated covariance matrix scaled according to the degrees of freedom of the multivariate t(5) distribution.

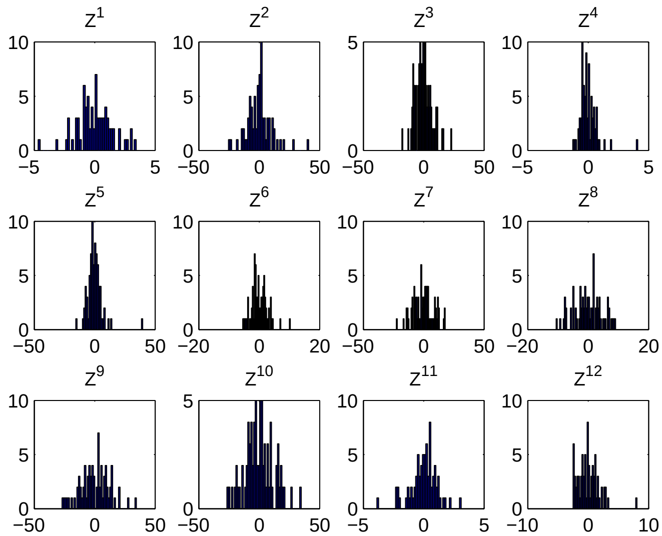


Figure 5.2: Histograms of residuals of the estimated multivariate model of Swedish macroeconomic factors.

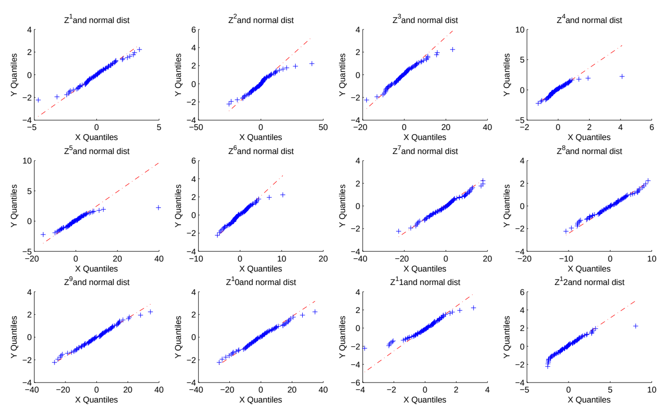


Figure 5.3: QQ-plots of residuals of the AR-model of Swedish macroeconomic factors with a normal reference distribution. Since the qq-plots are curved the residuals are more heavy tailed than a normal distribution.

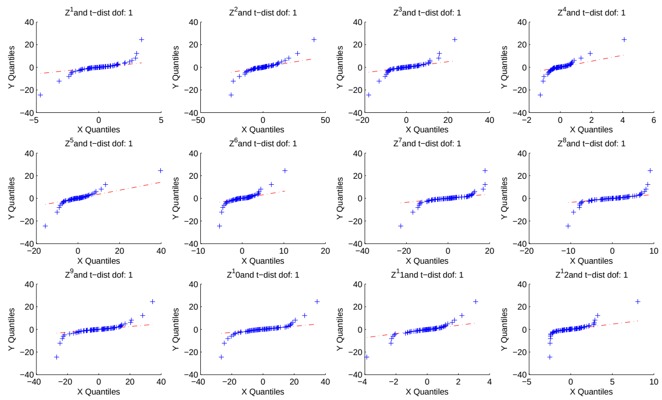


Figure 5.4: QQ-plots of residuals of the AR-model of Swedish macroeconomic factors with a t-distribution with 1 degree of freedom as reference distribution. Since the qq-plots are curved the residuals are less heavy tailed than a t(1) distribution.

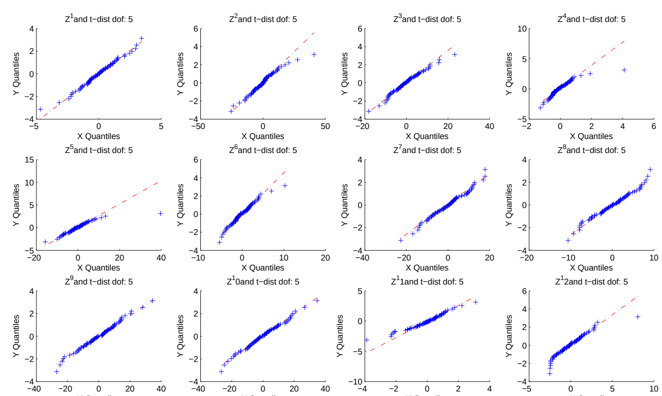


Figure 5.5: QQ-plots of residuals of the AR-model of Swedish macroeconomic factors with a t-distribution with 5 degrees of freedom as reference distribution. The qq-plots are approximately linear which suggest that a multivariate t(5) distribution can be used in the simulation.

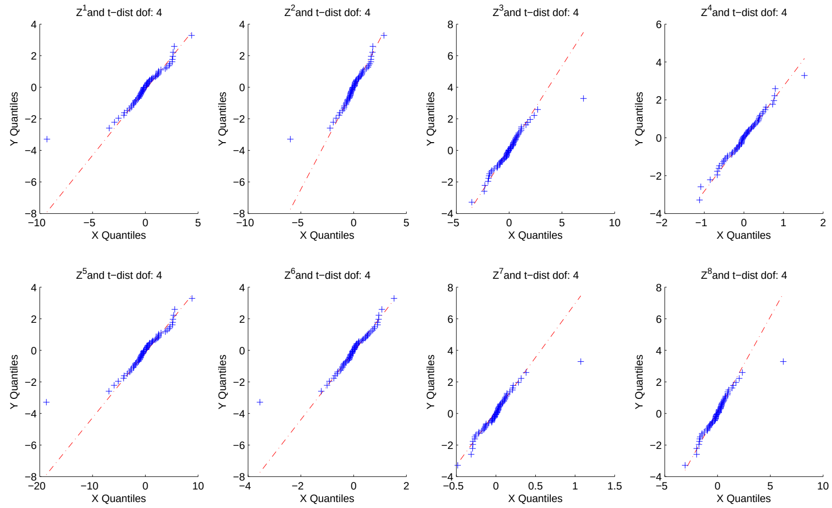


Figure 5.6: QQ-plots of residuals of the AR-model of Swedish principal components of macroeconomic factors with a t(4)-distribution as reference distribution. The qq-plots are approximately linear which suggest that a multivariate t(4) distribution can be used in the simulation.

5.1.2 Modelling of principal components of all lags of Swedish macroeconomic data

As stated in section 2.3 the principal components are uncorrelated which means that $\hat{\Gamma}(0)$ is a diagonal matrix. The optimal order of the multivariate model is chosen from $p = 1, \ldots, 4$. The p_{opt} maximizing the Schwarz's Bayesian Criterion is $p_{opt} = 4$. Thus $\hat{\Theta}_1, \ldots, \hat{\Theta}_4$ have to be estimated. The estimation of these are carried out in the same way as in the estimation of the Swedish macroeconomic factor model. Once $\hat{\Theta}_1, \ldots, \hat{\Theta}_4$ have been estimated the residuals can be calculated.

By eye-ball investigation of qq-plots in figure 5.6, 5.7 and 5.8 the residuals are considered to stem from a t(4) multivariate distribution.

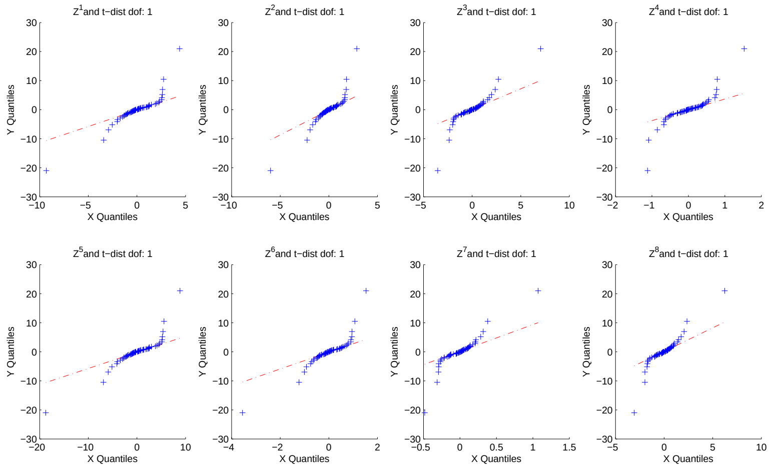


Figure 5.7: QQ-plots of residuals of the AR-model of Swedish principal components of macroeconomic factors with a t-distribution with 1 degrees of freedom as reference distribution. The qq-plots are non-linear which suggests that the residuals stem from a less large tail distribution.

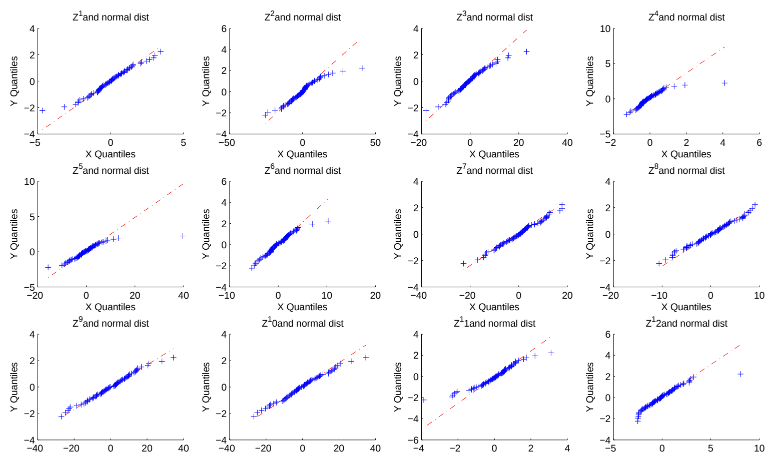


Figure 5.8: QQ-plots of residuals of the AR-model of Swedish principal components of macroeconomic factors with a normal reference distribution. The qq-plots are non-linear which suggests that the residuals stem from a more heavy tailed distribution.

5.2 Model of Baltic Macroeconomic factors

Two separate models for the macroeconomic factors of the Swedish model are to be calibrated. Firstly a model of the macroeconomic factors, secondly a model of the principal components of the macroeconomic factors of all possible time lags is to be calibrated.

5.2.1 Modelling of Baltic Macroeconomic factors

A multivariate autoregressive model is to be calibrated from the observations of Baltic macroeconomic factors. Since only Estonian CPI is available for quarter 2 of 2009, the last observation of this macroeconomic factor is removed from the sample. The simulation thus has to simulate the data for all three variables including the present quarter. The optimal order of the multivariate model is chosen from $p = 1, \dots, 4$. The p_{opt} maximizing the Schwarz's Bayesian Criterion is $p_{opt} = 3$. Thus $\hat{\Theta}_1, \dots, \hat{\Theta}_3$ have to be estimated. The estimation of these are carried out in the same way as in the estimation of the Swedish macroeconomic factor model. Once $\hat{\Theta}_1, \dots, \hat{\Theta}_3$ have been estimated the residuals can be calculated, histograms of these are shown in figure 5.9. The distribution of the residuals are decided upon by eye-ball investigation of the qq-plots in figure 5.10 which appears to be approximately linear.

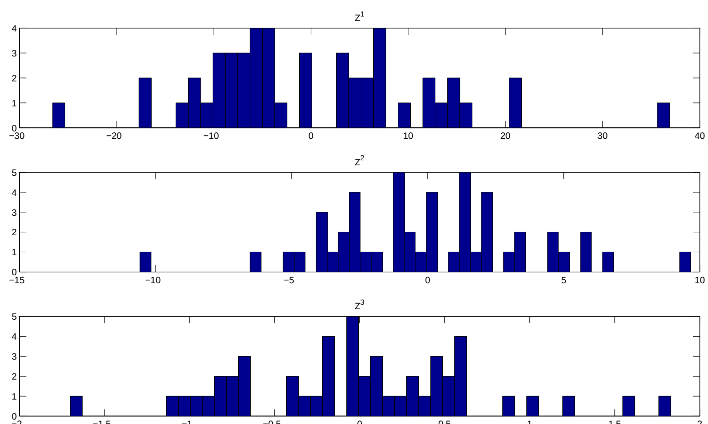


Figure 5.9: Histograms of residuals of the estimated multivariate model of Baltic macroeconomic factors.

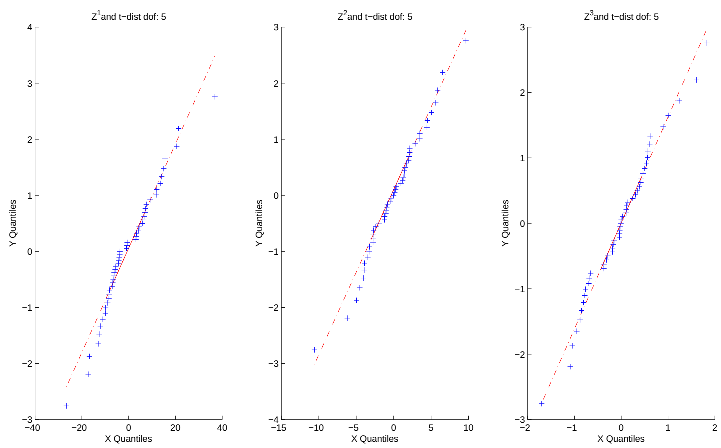


Figure 5.10: QQ-plots of residuals of AR-model for the Baltic region with a t(5) reference distribution. The qq-plots are approximately linear which suggest that a multivariate t(5) distribution can be used in the simulation of Baltic macroeconomic factors.

5.2.2 Modelling of principal components of all lags of Baltic macroeconomic data

As stated in section 2.3 the principal components are uncorrelated which means that $\hat{\Gamma}(0)$ is a diagonal matrix. The available observations for the Baltic macroeconomic factors are half of those available for Sweden. The optimal order of the multivariate model is chosen from $p = 1, \ldots, 4$. The p_{opt} maximizing the Schwarz's Bayesian Criterion is $p_{opt} = 4$. Thus $\hat{\Theta}_1, \ldots, \hat{\Theta}_4$ have to be estimated. The estimation are carried out in the same way as in the estimation of the Swedish macroeconomic factor model. Once $\hat{\Theta}_1, \ldots, \hat{\Theta}_4$ have been estimated the residuals can be calculated, histograms of these are shown in figure 5.11.

For marginal distribution a normal distribution seems plausible according to the qq-plots in figure 5.12.

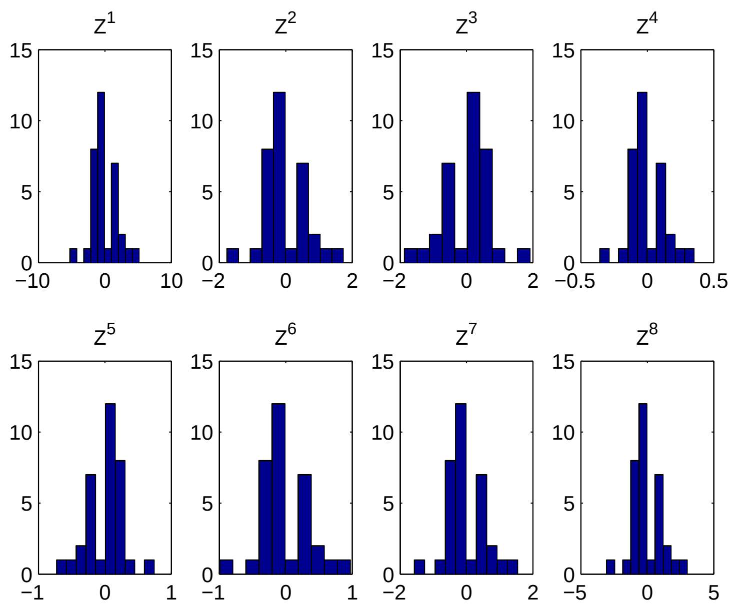


Figure 5.11: Histograms of residuals of the estimated multivariate model of principal components of Baltic macroeconomic factors.

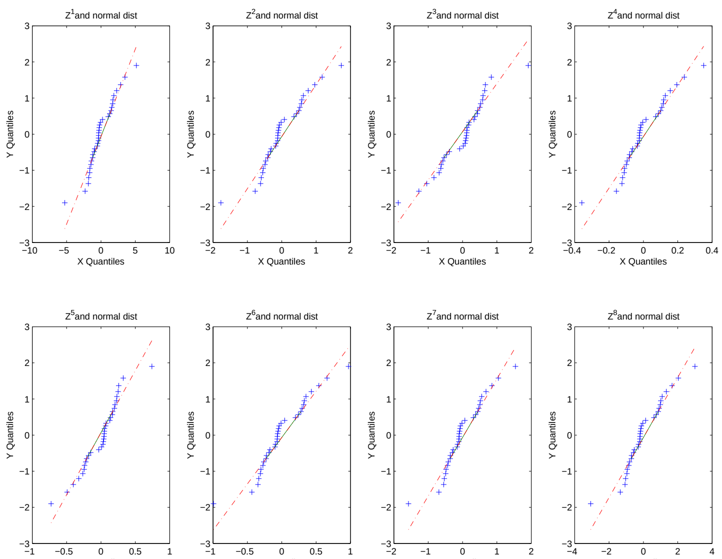


Figure 5.12: QQ-plots of residuals of AR-model for the Baltic region. with a normal reference distribution. The qq-plots are approximately linear which suggest that a multivariate normal distribution can be used in the simulation of principal components of Baltic macroeconomic factors.

Chapter 6

Simulation

In this chapter macroeconomic factors and principal components of macroeconomic factors lagged all possible time steps of the Baltic and Swedish region are simulated a large number of times. These simulations are used to predict future loss ratios, both by segments and by region.

6.1 Simulation of Swedish macroeconomic factors

6.1.1 Simulation of Swedish macroeconomic factors

Simulation of the Swedish macroeconomic factors will be done using the model calibrated in section 5.1.1. Macroeconomic factors will be used to predict loss ratios one year ahead, i.e. the distribution of four time steps have to be simulated. Since we know the values of the changes of macroeconomic factors today these will be used in the simulation. The macroeconomic factors of tomorrow will be simulated according to equation 6.1 where X_t^{swe} are the macroeconomic factors of today and Z_{t+1} are drawn from a multivariate t distribution with three degrees of freedom.

$$X_{t+1}^{swe} = \hat{\Theta}_1 X_t^{swe} + Z_{t+1}, \qquad Z_{t+1} \sim t(\mathbf{0}, \hat{\Sigma}, 3) \tag{6.1}$$

The simulation of X_{t+2}^{swe} is performed in the same way but using the simulation of the previous time step as the known previous component, and so on for the simulations of X_{t+3}^{swe} and X_{t+4}^{swe}. Simulating a large number of paths over four time steps gives the distribution of the Swedish macroeconomic factors of the next year.

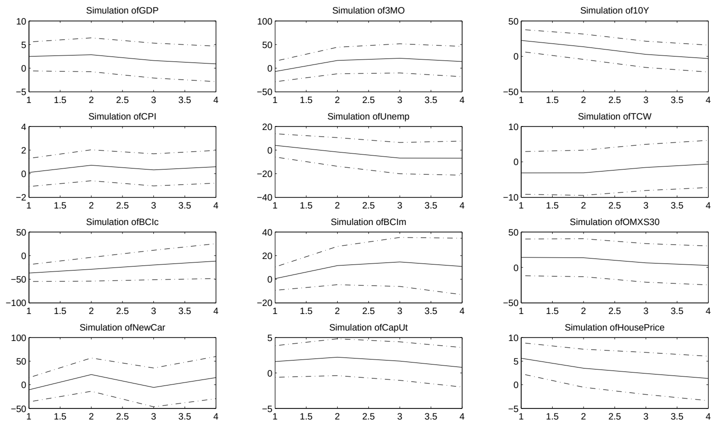


Figure 6.1: Simulations of changes of Swedish macroeconomic factors. The graphs show the mean and 95% confidence intervals. 500,000 simulations have been performed.

6.1.2 Simulation of principal components of Swedish macroeconomic factors of all time lags

The simulation of Swedish principal components will be done according to the following model:

$$PC_{t+1}^{swe} = \hat{\Theta}_1 PC_t^{swe} + \hat{\Theta}_2 PC_{t-1}^{swe} + \hat{\Theta}_3 PC_{t-2}^{swe} + \hat{\Theta}_4 PC_{t-3}^{swe} + Z_{t+1} \quad (6.2)$$

$$Z_{t+1} \sim t(\mathbf{0}, \hat{\Sigma}, 4) \quad (6.3)$$

The simulations are carried out four time steps ahead. The resulting outcomes can be seen in figure 6.2.

6.2 Simulation of Baltic macroeconomic factors

6.2.1 Simulation of Baltic macroeconomic factors

Baltic Macroeconomic factors will be used to predict loss ratios one year ahead. Since the value of today of all but one variable is missing in the sample the distribution of five time steps have to be simulated. The model to be simulated is:

$$X_{t+1}^{balt} = \hat{\Theta}_1 X_t^{balt} + \hat{\Theta}_2 X_{t-1}^{balt} + +\hat{\Theta}_3 X_{t-2}^{balt} + Z_{t+1}, \qquad Z_{t+1} \sim t(\mathbf{0}, \hat{\Sigma}, 5) \quad (6.4)$$

The simulated paths can be seen in figure 6.3

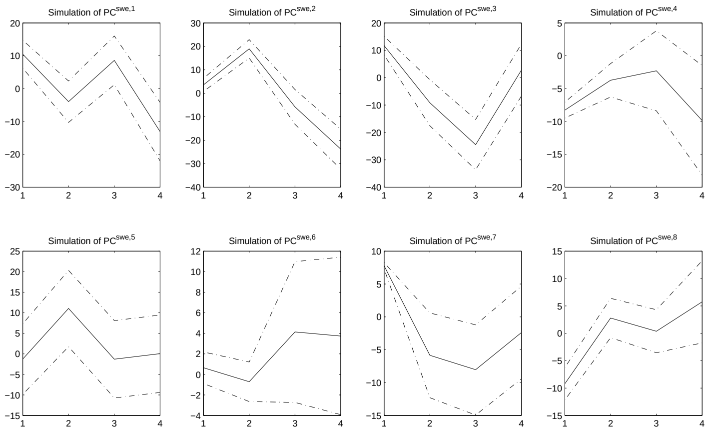


Figure 6.2: Simulations of changes of principal components of all lags of Swedish macroeconomic factors. The graphs show the mean and 95% confidence intervals. 500,000 simulations have been performed. Remember that it is not the actual changes of macroeconomic factors that are being modelled.

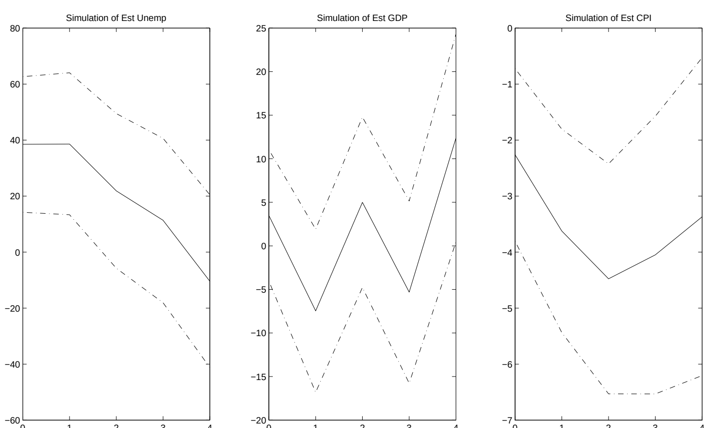


Figure 6.3: Simulations of changes of Baltic macroeconomic factors. The graphs show the mean and 95% confidence intervals. 500,000 simulations have been performed.

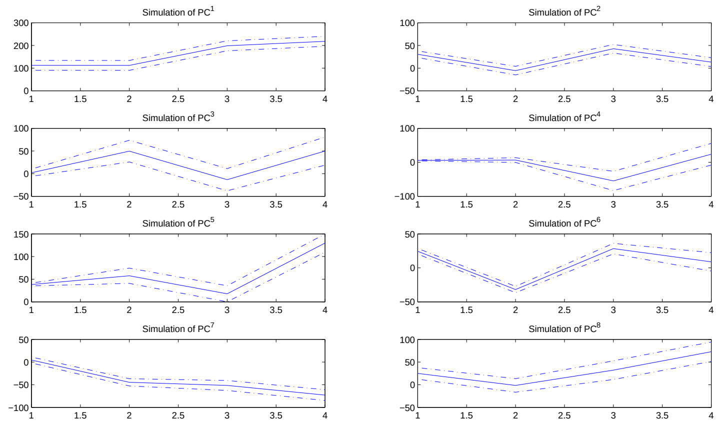


Figure 6.4: Simulations of changes of principal components of all lags of Baltic macroeconomic factors. The graphs show the mean and 95% confidence intervals. 500,000 simulations have been performed. Remember that it is not the actual changes of macroeconomic factors that are being modelled.

6.2.2 Simulation of principal components of Swedish macroeconomic factors of all time lags

The simulation of Baltic principal components will be done according to the following model:

$$PC_{t+1}^{balt} = \hat{\Theta}_1 PC_t^{balt} + \hat{\Theta}_2 PC_{t-1}^{balt} + \hat{\Theta}_3 PC_{t-2}^{balt} + \hat{\Theta}_4 PC_{t-3}^{balt} + Z_{t+1} \qquad (6.5)$$

$$Z_{t+1} \sim N(\mathbf{0}, \hat{\Sigma}) \qquad (6.6)$$

The simulations are carried out four time steps ahead. The resulting outcomes can be seen in figure 6.4.

Chapter 7

Prediction

In this chapter the models mapping loss ratios to macroeconomic factors and the simulations of macroeconomic factors are used to predict future changes of loss ratios. The prediction of loss ratios over the next year is calculated as:

$$LR^{year} = \left(\prod_{k=1}^{4} (1 + LR_{t+k}) \right) - 1 \tag{7.1}$$

where LR^{year} is on a one year time scale and LR_{t+k} in on a quarterly scale. The above formula is a reasonable approximation if the lending is constant and only the loan losses vary over the time period.

7.1 Prediction of loss ratios in the Swedish region

Using the model calibrated in section 4.2, where the macroeconomic factors have been lagged by the time lags of maximum correlation, future loan losses for the Swedish region will be predicted. The distribution of predicted loan losses over the next quarter is shown in figure 7.1. Comparing this distribution with the last loss ratio in figure 3.6 the predicted loss ratio seems to remain the same. In figure 7.2 the distribution of predicted total loss ratio of the next year is shown with most weight between 0.1% and 0.4%. Since the regression of Swedish loss ratios on principal components had such a poor explanatory power, no prediction using this model will be used.

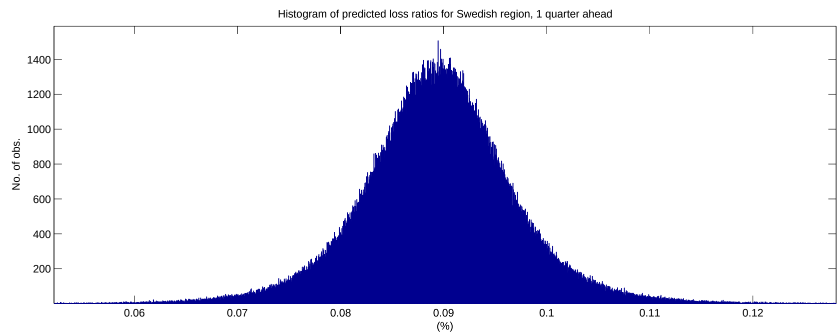


Figure 7.1: Distribution of predicted loss ratios of the Swedish region for the next quarter. 500,000 simulations

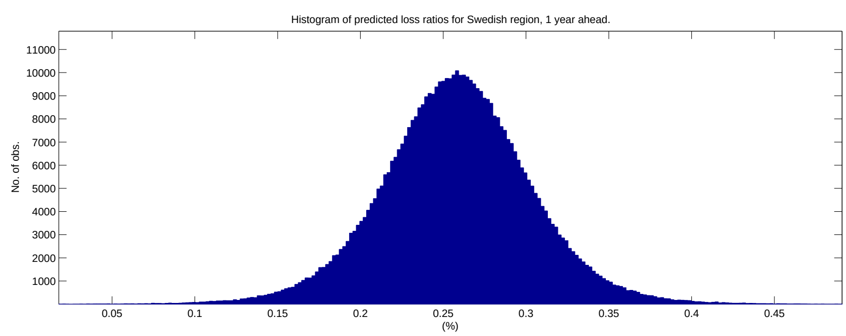


Figure 7.2: Distribution of predicted loss ratios of the Swedish region for the next year (four quarters ahead). 500,000 simulations of the path from next quarter to four quarters ahead

7.2 Prediction of loss ratios in the Baltic region

A total of three different model will be used to predict Baltic loss ratios.

1. The model decided by regression on Baltic macroeconomic factors lagged with the time step with the highest correlation.

2. The model decided by stepwise regression of macroeconomic factors of all plausible time lags.

3. The model decided upon by regression on principal components of macroeconomic factors of all time lags.

Since the model decided upon by stepwise regression has a minimum lag of 5 time steps, the prediction is given by data already known. The predicted loss ratio of the next quarter is 0.3835% and the loss ratio of the next year is 3.0732 %. The model using Baltic macroeconomic factors lagged with the time step with the highest correlation predicts significantly lower loss ratios for the Baltic region than does the principal component prediction.

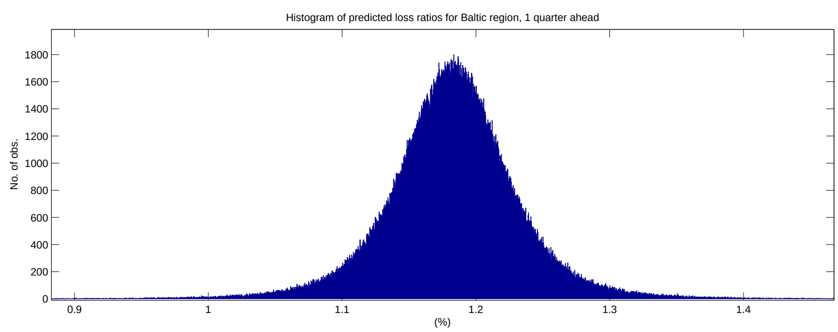


Figure 7.3: Distribution of predicted loss ratios of the Baltic region by the model of macroeconomic factors at time lag of largest correlation over the next quarter. 500,000 simulations

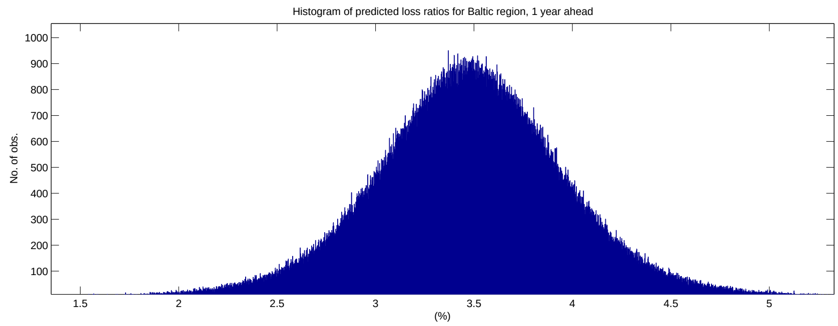


Figure 7.4: Distribution of predicted loss ratios of the Baltic region by the model of macroeconomic factors at time lag of largest correlation over the next year. 500,000 simulations

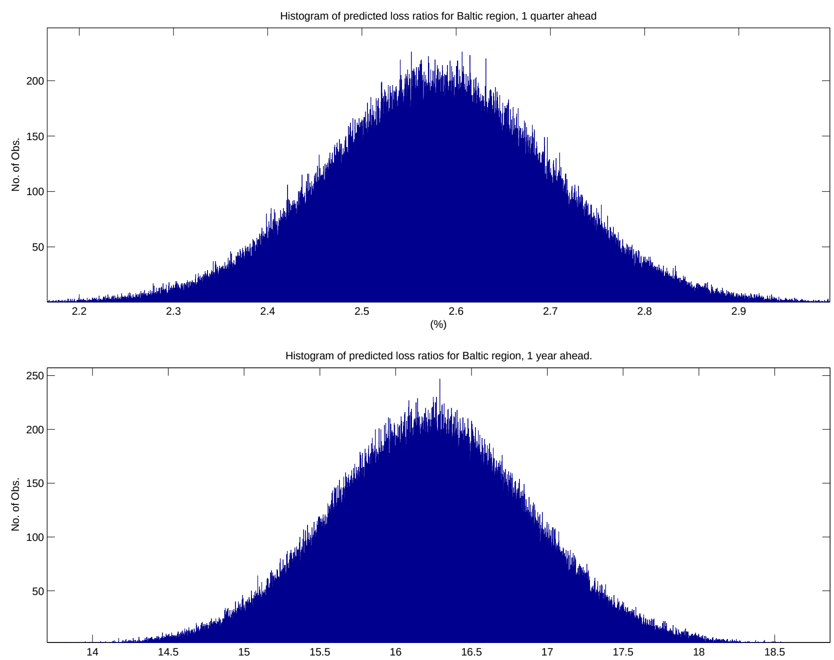


Figure 7.5: Distribution of predicted loss ratios of the Baltic region by the model of principal components for the next quarter and year. 500,000 simulations

7.3 Prediction of loss ratios by segment

Modelling loss ratios by segments by all macroeconomic factors and modelling loss ratios by macroeconomic factors chosen by stepwise regression have the same explanatory power. Only one of the predictions will be presented here, the one of prediction by the model that incorporates all macroeconomic factors of both regions. The distribution of predicted loss ratios by segment can be seen in figure 7.6 for the next quarter and in figure 7.7 for the next year.

Modelling loss ratios of the segments using principal components of macroeconomic factors of all time lags for the two regions also had acceptable explanatory power. The distribution of predicted loss ratios by segment using this method can be seen in figure 7.8 for the next quarter and in figure 7.9 for the next year.

To see how these simulated values relate to the historical values a scatter plots of both predictions between the segment "Private Customers" and "Total Corporate" is studied. As can be seen in figure 7.10 the predicted values varies a little from the historical values but are of the same magnitude when using macroeconomic factors with the highest correlation as predictors. In figure 7.11 prediction by principal components of all lags of macroeconomic factors can be seen. The values here are also of the same magnitude but show a clearer dependence between the loss ratios of the two segments.

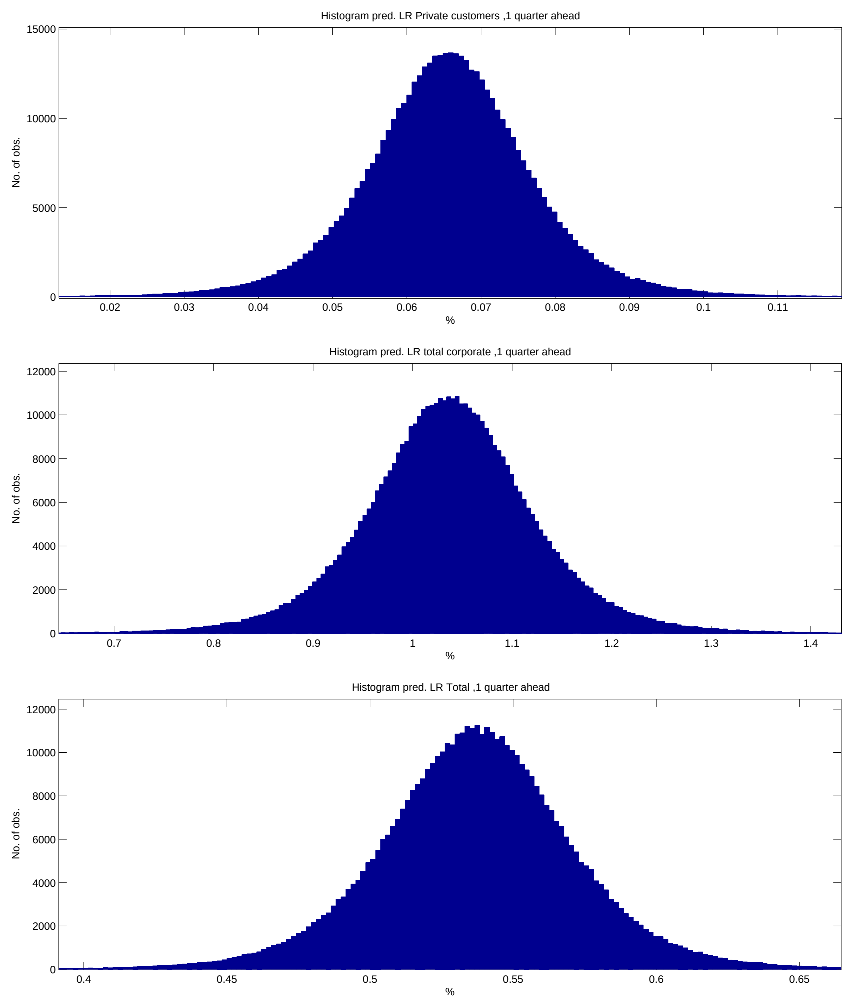


Figure 7.6: Distribution of predicted loss ratios of Swedbank's loss ratios of "Private", "Total Corporate" and "Total" segments over the next quarter. Modelling loss ratios by segments by all macroeconomic factors lagged with the time lag of highest correlation.

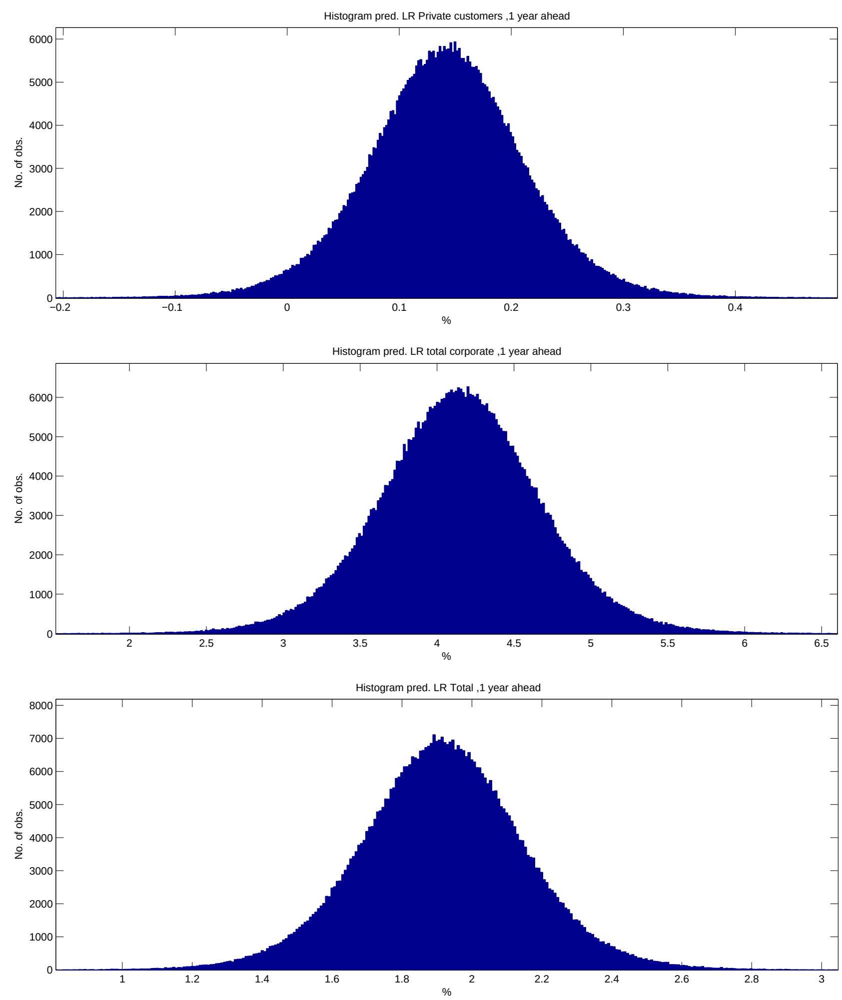


Figure 7.7: Distribution of predicted loss ratios of Swedbank's loss ratios of "Private", "Total Corporate" and "Total" segments for the next year. Modelling loss ratios by segments by all macroeconomic factors lagged with the time lag of highest correlation

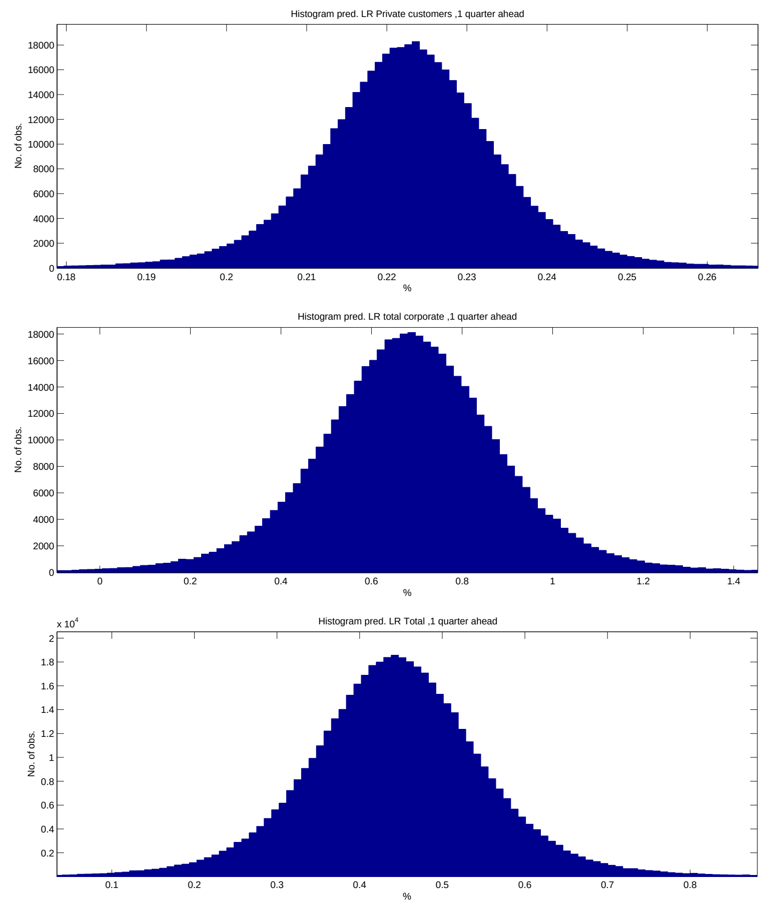


Figure 7.8: Distribution of predicted loss ratios of Swedbank's loss ratios of "Private", "Total Corporate" and "Total" segments over the next quarter. Prediction by principal components of macroeconomic factors of all time lags of both regions.

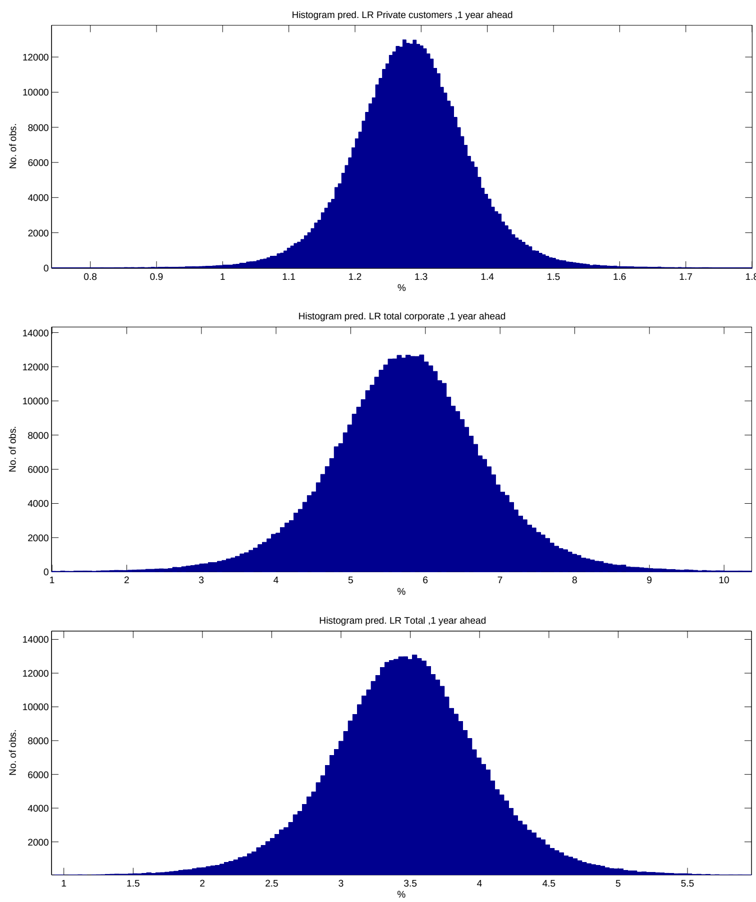


Figure 7.9: Distribution of predicted loss ratios of Swedbank's loss ratios of "Private", "Total Corporate" and "Total" segments over the next year. Prediction by principal components of macroeconomic factors of all time lags of both regions.

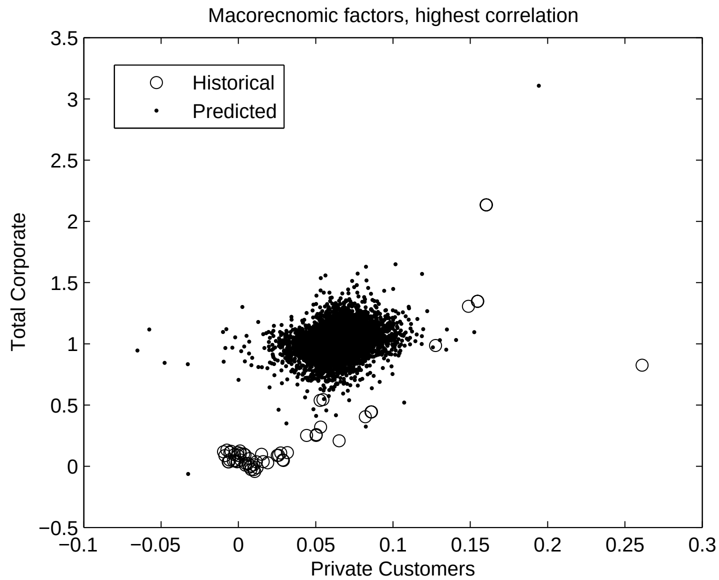


Figure 7.10: Scatter plot of prediction of segment "Private Customers" and "Total Corporate" by using macroeconomic factors with the highest correlation as predictors.

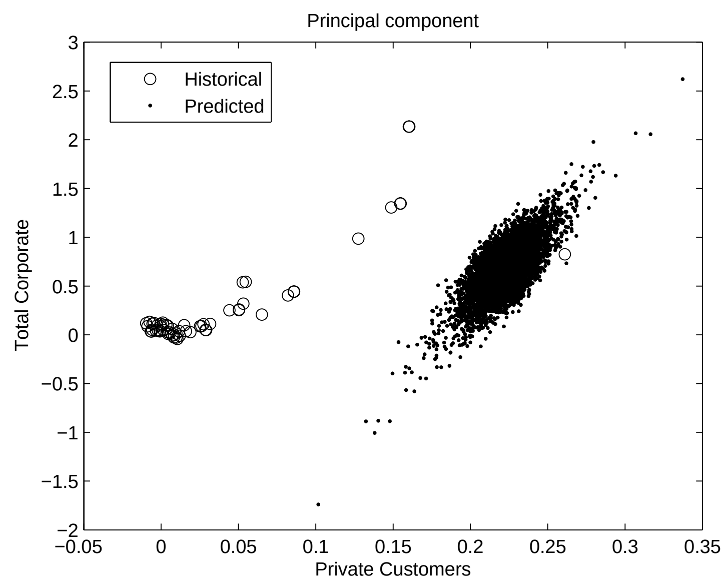


Figure 7.11: Scatter plot of prediction of segment "Private Customers" and "Total Corporate" by using principal components.

Chapter 8

Conclusion

In this master's thesis prediction models for Swedbank's loan losses of different segments and regions have been developed by regressing loss ratios on changes of macroeconomic factors. Through modelling and simulation of the macroeconomic factors, loss ratios over the next year have been predicted.

To perform the regression assumptions about the relationship between loan losses and macroeconomic factors have to be made. In this master's thesis loss ratios are assumed to depend linearly on changes of macroeconomic factors. Further, a number of different macroeconomic factors at different time lags are tested for goodness-of-fit to choose which macroeconomic factors to include in the model.

Loss ratios of the different segments are equally well explained by changes of macroeconomic factors lagged with the time lag of maximum correlation, a subset thereof chosen by stepwise regression and principal components of changes of macroeconomic factors of all plausible time lags. The adjusted R^2 of these three models varies from 0.66 to 0.85 depending on which model is used and what segment is modelled.

Swedbank's loss ratios in Sweden modelled by Swedish macroeconomic factors lagged with the time lag with highest correlation gives an adjusted R^2 of 0,57, to be considered low. The level of explanation is decreased when using principal components of Swedish macroeconomic factors of all plausible time lags to 0,034 which is unacceptably low.

Loss ratios of the Baltic region are modelled well by all three ways of choosing explanatory variables of the regression. The highest level of explanation is obtained by regressing on Baltic macroeconomic factors lagged with time lags chosen by stepwise regression and by regressing on principal components of changes of macroeconomic factors of all plausible time lags. The adjusted R^2 of these two models are 0.82 and 0.83 respectively.

The fact that loss ratios in the Baltic region are better explained by macroeconomic factors could be explained by customers in this region having less buffer to adverse movements of the economy. Customers in the Baltic region might have trouble fulfilling their obligations to Swedbank after a shorter time of unemployment or after a general downturn in the economy. This can also be seen in the generally higher maximum correlations of Baltic loss ratios and changes of Baltic macroeconomic factors.

The predicted values of loss ratios over the next quarter and year seem plausible since they are predicted given the extreme values of macroeconomic data over the last few quarters. One can note that the dependence structure between loss ratios of the segments "Private" and "Total Corporate" varies a lot between the models using principal components of macroeconomic factors of all time lags and macroeconomic factors with the time lag of highest correlation. The predicted loss ratios of the Baltic region over the next year are almost four times larger when using principal components instead of macroeconomic factors with the time lag of highest correlation. This emphasises the need of caution when using this model for prediction of future loss ratios. The available data on loss ratios is very limited, thus the model should be used with caution.

The model presented in this thesis could be extended by mapping commission incomes from the business unit Swedbank Markets to the macroeconomic factors used in this master's thesis, thereby obtaining a model of the dependence between credit loss ratios and net commission incomes .

Appendix A

Figures of macroeconomic factors

A.1 Swedish macroeconomic factors

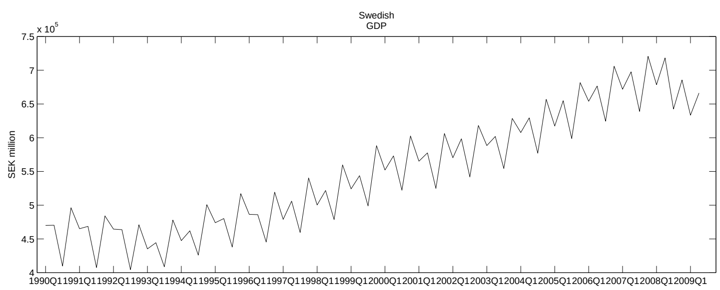


Figure A.1: Swedish GDP (SEK million)

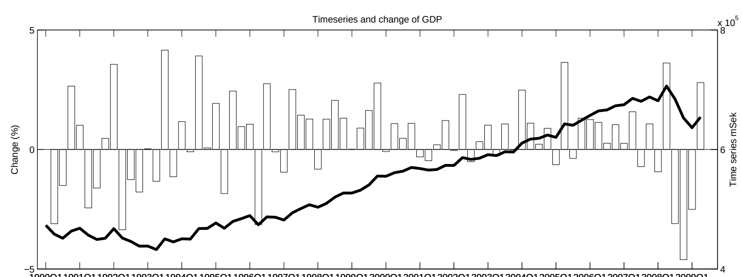


Figure A.2: Time series (SEK million) and quarterly change (%) of seasonally adjusted Swedish GDP.

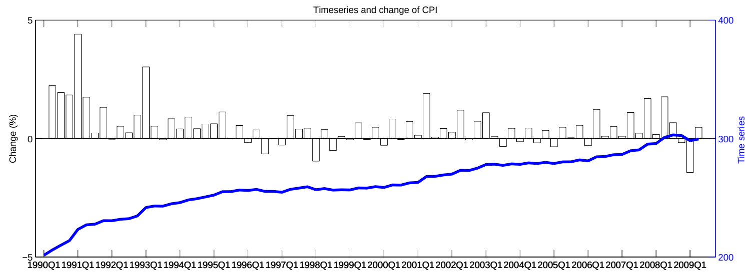


Figure A.3: Time series and quarterly change of Swedish CPI (1980=100)

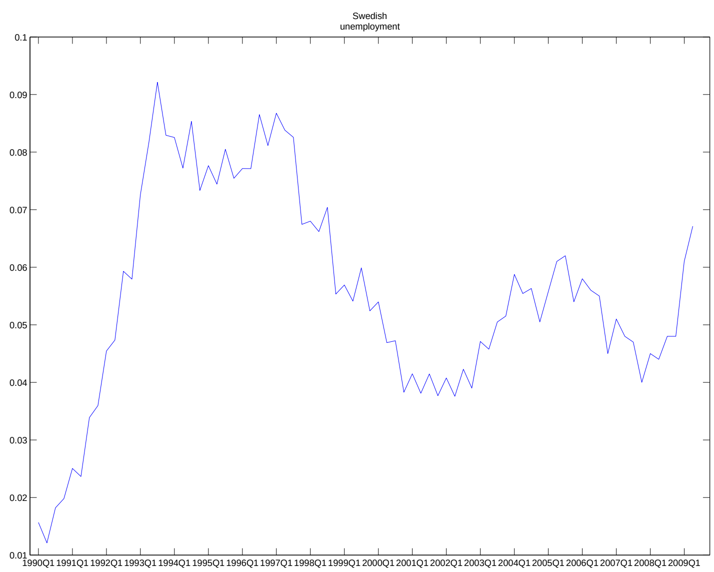


Figure A.4: Unemployment rate, Sweden (% of workforce))

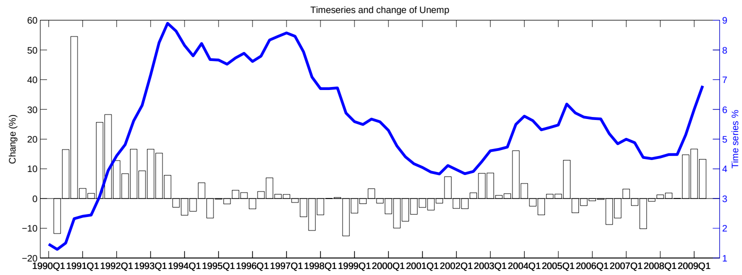


Figure A.5: Time series and quarterly change of Seasonally adjusted Swedish unemployment rate (% of workforce)

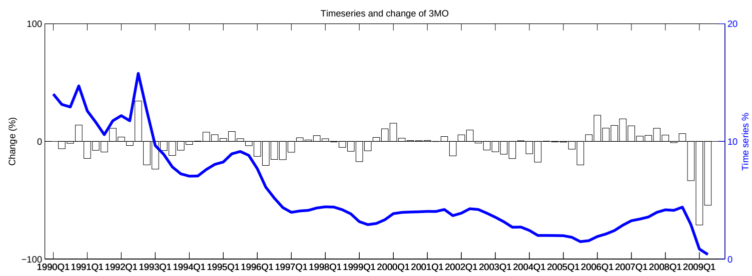


Figure A.6: Time series and quarterly change of Swedish 3 month treasury bill

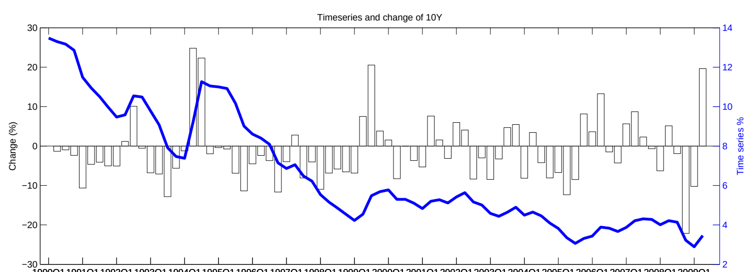


Figure A.7: Time series and quarterly change of Swedish 10 year government bond

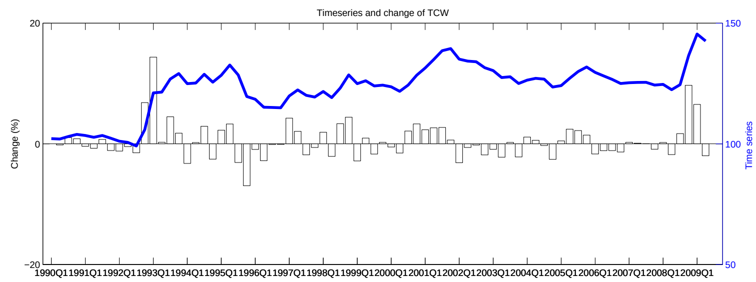


Figure A.8: Time series and quarterly change of Swedish TCW index (1992=100)

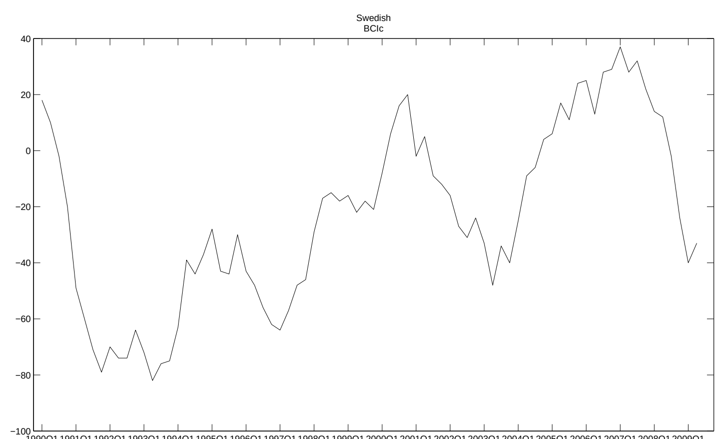


Figure A.9: Swedish Business confidence indicator, construction

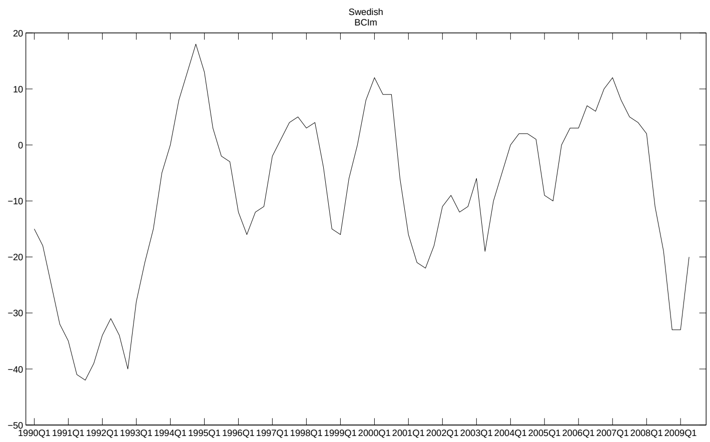


Figure A.10: Swedish Business confidence indicator, manufacturing

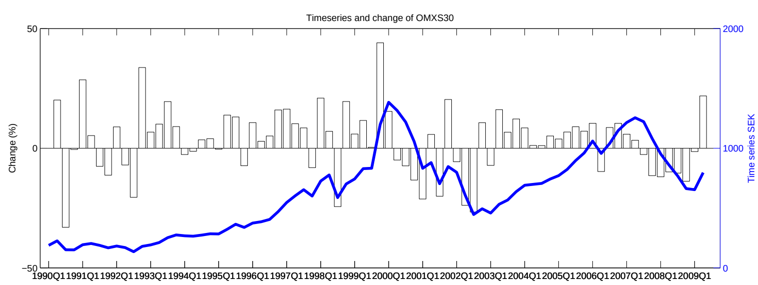


Figure A.11: Time series and quarterly change of OMXS30.

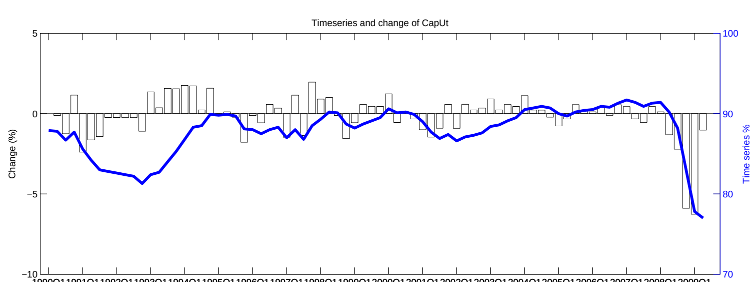


Figure A.12: Time series and quarterly change of Capacity utilisation in industry

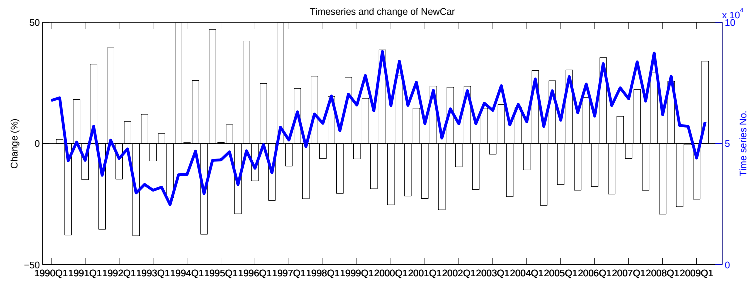


Figure A.13: Time series and quarterly change of New car registrations in Sweden

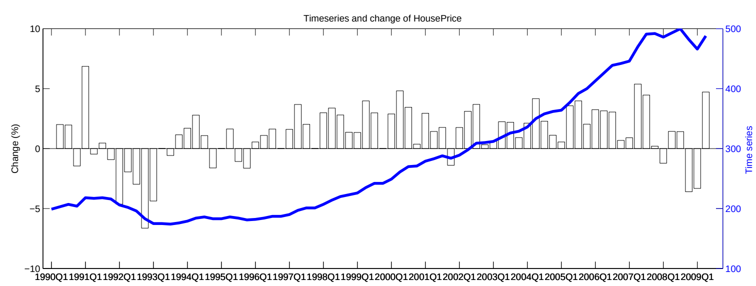


Figure A.14: Time series and quarterly change of Swedish Real Estate Price Index (1981 = 100)

A.2 Baltic macroeconomic factors

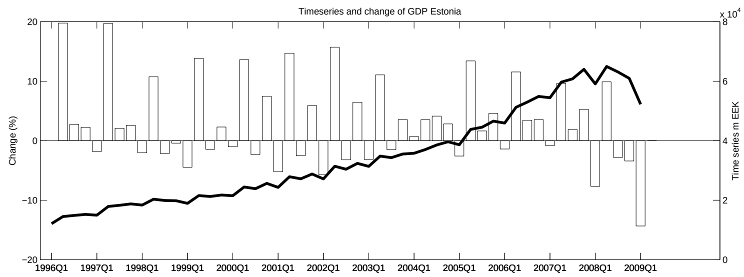


Figure A.15: Time series (m Estonian Kroon) and quarterly change (%) of GDP Estonia

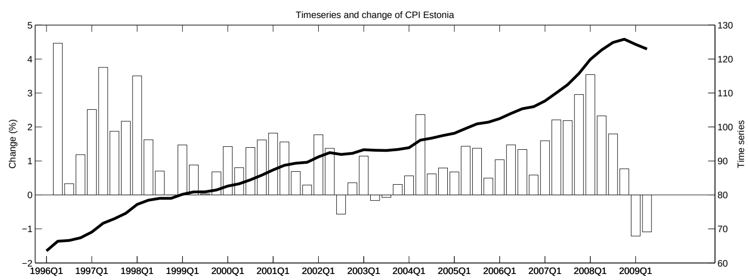


Figure A.16: Time series and quarterly change of Estionian CPI

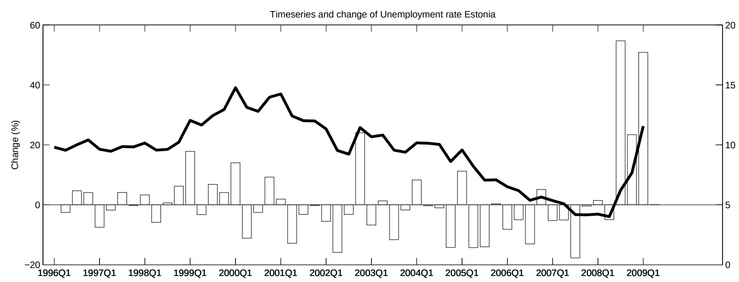


Figure A.17: Time series and quarterly change of Estonian Unemployment rate (%)

Bibliography

Aas, K., Dimakos, X., & Øksendal, A. (2007). Risk Capital Aggregation. *Risk Management*, *9*, 82–107.

Brockwell, P. & Davis, R. (2002). *Introduction to time series and forecasting*. Springer.

Bunn, P., Cunningham, A., & Drehmann, M. (2005). Stress testing as a tool for assessing systemic risk. The Bank of England Financial Stability Review pages 116–126.

Carling, K., Jacobson, T., Lindén, J., & Roszbach, K. (2007). Corporate Credit Risk Modeling and the Macroeconomy. *Journal of Banking and Finance*, *31*, 845 – 868.

Draper, N. & Smith, H. Applied regression analysis. 1998. *New Yoek: John Wiley and Sons Inc.*

Drehmann, M., Sorensen, S., & Stringa, M. (2008). The integrated impact of credit and interest rate risk on banks: An economic value and capital adequacy perspective. The Bank of England working paper no. 339.

Jolliffe, I. (2002). *Principal component analysis*. Springer verlag.

Larsson, H. (2009). Inter-risk correlation within economic capital. Master's thesis, KTH.

Neumaier, A. & Schneider, T. (2001). Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software*, *27*(1), 27–57.

Overbeck, L. (2006). Integration of credit and market risk. In M. K. Ong (Ed.), *Risk Management* (pp. 341 – 365). Academic Press.

Riksbanken (2009). www.riksbank.se.

Rosenberg, J. V. & Schuermann, T. (2006). A general approach to integrated risk management with skewed, fat-tailed risks. *Journal of Financial Economics*, *79*(3), 569–614.

Svensson, E. (2007). Risk aggregation and dependence modelling with copulas. Master's thesis, Lund Institute of Technology.

Wooldridge, J. (2003). *Introductory econometrics, a modern approach.* South-Western Mason, OH.