DIPLOMARBEIT

Titel der Diplomarbeit

Modeling Credit Risk through the Austrian Business Cycle: An Update of the OeNB Model

Verfasserin

Angela Zirknitzer

Angestrebter akademischer Grad

Magistra der Sozial- und Wirtschaftswissenschaften
(Mag. rer. soc. oec.)

Wien, im September 2010

Studienkennzahl lt. Studienblatt:	157
Studienrichtung lt. Studienblatt:	Internationale Betriebswirtschaft
Betreuer/Betreuerin:	Univ.-Prof. Dr. Alexander Stomper

Table of contents

Abstract

This thesis is about how to measure the influence of macroeconomic variables such as GDP or consumer price index on default rates in nine industrial sectors.
The software used to calculate the results is Stata. It includes the software commands needed to program the model.

The statistical procedures used in this thesis are suggested in a paper by Boss et al. and published by the Austrian central bank in order to capture the information of a large data set and to bridge the differences between the economic cycles and the credit cycles.
Only the first model of the two suggested will be implemented here.
24 macroeconomic variables will be transformed in a principal component analysis to number the set down to five linear combinations which explain about 71% of the model's variance. Then these five factors will be included in the OLS regression to determine their statistical and economical significance.

Results show that either PCA factor one, three or five is statistically significant in each sector, at least one of them, sometimes two. However, they are not economically significant in any sectors.
The GDP is found statistically significant in the majority of sectors and one of the few variables which are also economically significant.
The consumer price index is statistically significant in many sectors, however insignificant economically. The labor productivity is, like the GDP, one of the few variables which is statistically significant in the majority of sectors as well as economically significant.
The statistical outliers such as short-term interest rate, industrial production or PCA factor five are each statistically only in one or two sectors and are not economically significant at all.
The unemployment rate, although overall statistically significant, is not economically significant in any sector, except the service sector.

Introduction

The goal of this thesis is the implementation of an OeNB model, which was proposed by Michael Boss et al.[1] to capture the impact of a change in macroeconomic variables on probabilities of default, using the statistical software Stata.
The OeNB published this paper in its financial market report in 2009 in response to the financial crisis. The authors aim to find a link between macroeconomic environment and credit risk. It's an update of the OeNB's previous credit risk model and faces two challenges:

1. how to exploit the information held by a large data set and
2. how to bridge the difference between the different time horizons of the business cycles and the credit cycles.

The authors address the first problem by a regression analysis based on a principal component analysis, the second problem by conducting a threshold approach. Their paper originally suggests a procedure which can be split into four parts:

1. Regression models for each of the Austrian corporate sectors (two different approaches)
2. Using the Austrian Quarterly Model provided by the OeNB to create a macroeconomic scenario. This serves as stress test for the models.
3. Performing a principal component analysis (PCA) to avoid the arbitrary selection of variables
4. Applying a threshold approach to bridge the differences in credit cycles and business cycles

The probabilities of default p at time t are modeled in a logistic function of an industry-specific macroeconomic index $y_{t,s}$ at time t in sector s which depends on the current value of the observed macroeconomic variables:

$$G(y_{t,s}) = p_{t,s,} = \frac{1}{1+e^{-y_{t,s}}}$$ [2]

The authors use two different approaches on this equation: The first one is by Wilson (1997) who proposes to take the inverse of the logistic function to calculate the

[1] Boss et al., 2009: 92-108
[2] Boss et al., 2009: 92-108

values of the macroeconomic index y_t based on the observed default probabilities p at time t:

$$y_t = -\ln\left(\frac{1}{p_t} - 1\right)$$

The macroeconomic index however is not stationary and the authors address this problem by taking differences $\Delta y_t = y_t - y_{t-4}$ and $\Delta x_t = x_t - x_{t-4}$

Then the regression equation is estimated for a macroindex y at time t

$$\Delta y_t = \sum_{i-\cap}^{K} \beta \Delta x_{i,t} + \varepsilon_t = X_t \beta + \varepsilon_t \text{ with } \Delta x_{0,t} := 1$$

where $\Delta x_{1,t}, \Delta x_{2,t}, \ldots \Delta x_{K,t}$ denote the year-on-year changes of macroeconomic variables in the data set and ε_t the normally distributed standard error.

Furthermore this method includes a principal component analysis to reduce the problem of collinearity and to number down the variables included and to generate linear combinations of them. The authors include the first five components which explain most of the variables' variance.

The second approach is by Papke and Wooldridge[3] (1996). In contrast to the first method, the probabilities of default are not transformed, but their estimation explicitly accounts only for data between 0 and 1. The estimation equation looks like this:

$p_t = G(\Delta X_t \beta) + \varepsilon_t$, for $\varepsilon_t \sim N\left(0, \delta G(\Delta X \beta)\{1 - G(\Delta X \beta)\}\right)$

Instead of an OLS optimization method the maximum likelihood method is applied.

The log likelihood is given by $\ln L(\beta) = \sum_{t=1}^{T} \left\{ p_t \ln\left[G(\Delta X_t \beta)\right] + (1 - p_t)\ln\left[G(\Delta X_t \beta)\right]\right\}$

As mentioned before, the aim of my thesis is to show a way to implement the model, including software commands and result window outputs and discuss the results for a shorter time-series data set.

The mathematical and statistical procedures I will use are the ones proposed in this paper by Boss et al., however, I will only calculate the regression model based on Wilson's approach as well as the precedent principal component analysis. I will also briefly show how to program Papke & Woolridge's approach for fractional data between 0 and 1. The threshold model, which was already tested on the data by

[3] Papke & Woolridge (1996): 619-663

Boss et al., did not bring any satisfying results and therefore will be excluded from my thesis.

This thesis is written in cooperation with the Österreichiche Volkbanken AG.

In the first chapter I will give you some background information on the statistical procedures conducted in the model: the principal component analysis and the following standard regression.

In the second chapter I will discuss the data I am using and list the input variables for the model.

The third chapter will give a short overview of the software I am using in the forth chapter to implement the model in Stata, including several software commands I used and result window outputs.

The conclusion will sum up the results, compare the sectors and show and the model's limitations.

1. Statistical Procedures

This chapter will give some background information on the statistical procedures involved. Starting with quite a big data set of 39 variables and 98 observations each, the principal component analysis will try to figure out the structure of the data set and cutting it down to the most significant (linear) variable combinations. These new combinations will then later be used in the logistic regression analysis to estimate the relationship between the macroeconomic variables and the probabilities of default.

1.1. The principal component analysis (PCA)

According to J. Edward Jackson (1991), the principal component analysis is a technique from the multivariate statistics and transforms a number of related variables into a new set of uncorrelated variables that will explain most of the variables' variance. The goal is to get as few significant linear combinations of the input variables as possible explaining as much of the variance as possible in order to reduce the number of variables and therefore see the structure of the data. This method is commonly used on large sets of variables where one expects a certain degree of correlation between the variables (redundancy).[4]

To conduct the PCA on the data set, the mean has first to be subtracted.
Then, as mentioned before, the n x n covariance matrix is calculated for n variables. Because the covariance matrix is square, we are later able to calculate the eigenvectors and eigenvalues for the matrix.
For a data set with n variables, the correlation matrix looks like this:

$$s = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1n} \\ s_{12} & s_2^2 & \cdots & s_{2n} \\ \vdots & \vdots & & \vdots \\ s_{1n} & s_{2n} & \cdots & s_n^2 \end{bmatrix}$$

where the diagonal is the variance of variable x_i and s_{ij} is the covariance between the i^{th} and j^{th} variable. Covariances that are not zero indicate a relationship between two

[4] Jackson (1991): 10-11

variables and pretend that it is linear, the strength of this relationship represented by the correlation coefficient $r_{ij} = \frac{s_{ij}}{s_i s_j}$ [5]

In order to determine the principal components, we have to create a diagonal matrix L which can be obtained by pre- and post-multiplying it with an orthogonal matrix U, which contains the eigenvectors of covariance matrix S:

(Eq 1) $L = U'SU$

The elements of matrix L are the eigenvalues of covariance matrix S. This procedure is, geometrically, a principal axis rotation of the original axes.[6]

Later in the implementation chapter the five most significant components are chosen to remain in the data set.

1.2. The standard regression

Subsequent to the principal component analysis is the standard regression.

In this paper, the probabilities of default are estimated by a logistic function already given in the introduction:

(Eq.2) $G(y_{t,s}) = p_{t,s,} = \frac{1}{1+e^{-y_{t,s}}}$ [7]

where $y_{t,s}$ denotes the industry-specific macroeconomic index in sector s at time t.

The probabilities of default in my data set are stated in relative frequencies, not in dichotomous numbers 1 or 0, so the question arises: How does one fit a model when the dependent variables are proportions?

One way is the logit transformation also proposed by Wilson and applied by Boss et al.

Kleinbaum (1996) defines the logit transformation as following: “The logit transformation, denoted as logit P(X), is given by the natural log (i.e. to the base e) of the quantity P(X) divided by one minus P(X), where P(X) denotes the logistic model as previously defined.”[8]

---

[5] Jackson (1991): 10-11

[6] Jackson (1991): 6-9

[7] Boss et al., 2009: 92-108

[8] Kleinbaum (1996): 17

The values for the macroeconomic index y_t are calculated as shown by the following equation:

(Eq. 3) $y_t = -\ln\left(\frac{1}{p_t} - 1\right)$[9]

The resulting y_t are our dependent variables that later will be part of the inputs for the regression. p_t denote the probabilities of default. Now a standard regression with ordinary least squares optimization can be conducted. The standard linear regression observes the dependence of one variable to another. Its line reflects the relationship between two or more variables and estimates its strength.

A typical regression model can look like this:

(Eq. 4) $Y = \beta_0 + \beta_1 X + \varepsilon$

Y is the dependant variable which has the value $\beta_0 + \beta_1 X$ for a given X. ε refers to an error term caused by an unobserved variable. While ε changes for each variable, β_0 and β_1 are fixed terms that can be estimated by using the observations in the data set. β_0 is also called the intercept and indicates the value of Y when X is zero. β_1 is the slope of the line, indicating how much Y would change per one-unit of change in X. [10]

Estimates b_0 and b_1 for the unknowns β_0 and β_1 can be calculated using the observations for Y and X in the data set to predict a $\hat{Y}$. Therefore, the equation can be written as:

(Eq. 5) $\hat{Y} = b_0 + b_1 X$

$\hat{Y}$ can thus be predicted for a given X when b_0 and b_1 are determined.

In order to get a line which optimally reflects all of the observations, I am using the least squares method. For n available sets of observations the sum of squares of deviations from the regression line can be written as:

(Eq. 6) $S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$ **for i = 1, 2, ... n**

Now we estimate b_0 and b_1 to be numbers for which the sum of squares S is as small as possible. Y_i and X_i are observations taken from the data set. The line we get is the

[9] Boss et al., 2009: 92-108

[10] Draper & Smith (1981): 8-11

one that minimizes the sum of squares of all discrepancies between observations and the line.[11]

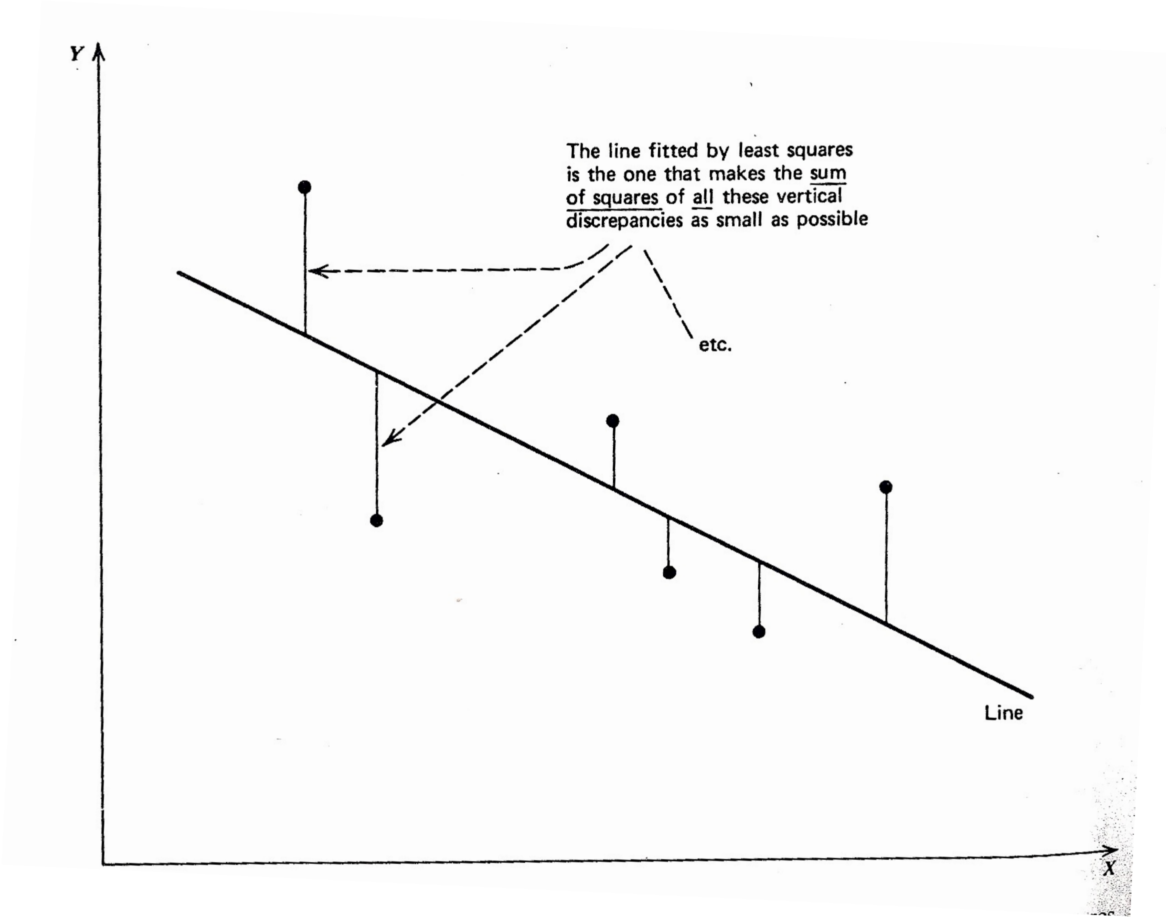

Figure 1: OLS

[11] Draper & Smith (1981): 11-22

2. Data

2.1. Macroeconomic variables

There are two time series indices of data I had at my disposal for this study: An index for the time period 1985-2009 for the macroeconomic variables and another index from 1985-2009 for the probabilities of default, provided by Eurostat. The data is stated quarterly. My data set spans a shorter time horizon than the original data used in the paper, therefore the results might deviate from the results obtained by Boss et al.

These 24 factors are tested and transformed in the principal component analysis:

- Total capital cost (CAC)
- Private credit, amount outstanding (CPN)
- Domestic demand, real (DDR)
- Government budget balance (GB)
- Government debt gross (GDN)
- Government disposal income (GYN)
- Harmonized index of consumer prices (HIC)
- Interest payments on government debt (INN)
- Total investment, real (ITR)
- Real marginal product of capital (MPC)
- Imports, real (MTR)
- Net foreign assets (NFA)
- Net factor income (NFN)
- Private consumption, real (PCR)
- Direct tax paid by households (PDN)
- Average labor productivity (PRO)
- Private sector disposal income, real (PYR)
- Unit labor costs, adjusted (ULA)
- Unemployment rate (URX)
- Value added tax (VAT)
- Real compensation per employee (WURYD)
- Export, real (XTR)

- GDP, real (YER)[12]

These variables are the input for the PCA, spanning a period of 24 years. The first five factors, which explain around 71% of the variance, are used as input for the following standard regression. Furthermore, these macro variables serve as explanatory variables:

- GDP, real (YER)
- Private consumption, real (PCR)
- PCR/GDP
- Unemployment rate (URX)
- Average labor productivity (PRO)
- Private sector disposable income, real (PYR)
- Total investment real (ITR)
- Investment in equipment, real (IER)
- IER/GDP
- Unit labor costs, adjusted (ULA)
- Exports, real (XTR)
- XTR/GDP
- Short-term interest rate, real (STI real)
- Long-term interest rate (LTI real)
- Short-term interest rate, nominal (STI)
- Long-term interest rate, nominal (LTI)
- Industrial production, real (IPEXE)
- Oil price
- Harmonized index of consumer prices (HIC)[13]

2.2. Probabilities of default

The probabilities of default were provided by the Kreditschutzverband von 1870 (Austrian creditor association), quoted in relative frequencies. The numbers are calculated by taking the number of companies that have filed for bankruptcy plus the number of companies that have filed for bankruptcy but have been rejected divided by the number of companies.

The data is split into the Austrian corporate sectors:

[12] Boss et al. (2009): 92-108
[13] Boss et al. (2009): 92-108

- Agriculture
- Production
- Construction
- Trading
- Tourism
- Transport
- Services
- Others
- The overall probabilities[14]

2.3. Skew data

Some variables show a high skew, which means that their distribution is asymmetric. Due to that, the logarithm has been taken.

The following variables have been log transformed:

- Total capital cost
- Domestic demand, real
- Government budget balance
- Government debt, gross
- Government disposal income
- Interest payments on government debt
- Total investment, real
- Imports, real
- Net foreign assets
- Private consumption, real
- Direct tax paid by households
- Private sector disposal income, real
- Total tax revenues
- Value added tax
- Real compensation per employee
- Export, real
- GDP, real

[14] Boss et al. (2009): 92-108

3. Statistical software Stata

Stata is a statistical software licensed by StataCorp LP. It can handle any kinds of data, time-series, panel or cross-sectional data. Through a command line or a do-file you can read and organize data, draw graphs or conduct statistical analyses. A broad user community provides self-written programs for processes not included in the default setting.[15]

The first thing I learned about Stata is that it is case sensitive, no matter whether it comes to variable names or commands. Basically, I was using four windows as interface:

In the **review window** you can again see the commands you have just entered. If black, Stata is able to fulfill the command, if red there is an error in your command, for instance a spelling mistake. If this happens you do not have to retype the whole command, just click on the command line in the review window and it will be copied into your command window. There you can correct it before submitting it again. Furthermore, you can save all your commands listed in the review window in a log-file. When you are still unfamiliar with Stata and haven't figured out certain commands yet, you almost always have the possibility to do it by menu or dialog. There, you don't have to think about the syntax and can simply enter the inputs you need in the assigned space and get your result. This might take a lot longer than to just write a command line, but it is easier to use at the beginning. Once you submit your data in the menu or dialog, the command for your action is written in the review window and you now know what to type the next time instead of using a dialog or menu.

In the **variables window** a list of variables of the data set you are currently using is shown. You can see the variables' names, labels, type and format. By clicking on one variable, it will be pasted into the command window. In this window you can only alter variable names, delete them, attach and edit a label or add notes.

[15] Baum (2006): 1-3

In the **command window** you enter your commands. If you want to see the history of commands you have already submitted, type *history*.

In the **result window** the results of your commands are shown. This can be a regression table with residuals, coefficients, t-values etc., it can be the notification that changes in your data set have been made, it can be a notification that a command cannot be executed due to an error, it can be a list of variables and so on. Graphs are usually shown in a separate window.

I will now describe some of the basic commands which I have used daily. Some of them are not used in the do-file, but I was using them regularly when I was working with the software and figuring out the program. For commands with a more complicated syntax, I will give an example.

3.1. Basic commands

- **use**: this command is used when you wish to open a Stata file (.dta) in Stata. It can be a file already on your computer or a file you can access on the Internet.[16]
- **save** *filename:* before you exit Stata, you can save your data set in a new Stata file by using save and the name of the new set, e.g. *save newfile*. If you wish to save the changes made in your data set without saving the changes in a new file, you write *save, replace*.[17] If you do not save your changes or explicitly tell Stata to discard them, you will not be able to quit the program.
- **clear:** this command clears Stata's memory. When you wish to exit Stata and do not want to save the changes you have made in your data set, you will have to write *clear* in order to exit.[18]
- **_n and _N**: in your data set of observations _n refers to the current, _N to the highest observation. I used _n for the calculation of the year-on-year changes in the index.[19] They only work in combination with a command.
- **if and in**: to restrict Stata to just a certain range of variables, you use the *if* expression or in range command[20], e.g. if you want Stata only to operate on

[16] Baum (2006): 7
[17] Baum (2006): 10
[18] http://www.stata.com/help.cgi?clear, 13.6.2010
[19] Baum (2006): 8
[20] Baum (2006): 11

the first ten observations, you'll write *in 1/10*. For logical conditions you will use the *if* expression, e.g. if you have have a data set with a variable "goodlooks" which takes values from 1-10 and you only wish to use a command on those over 5, you will write *if goodlooks > 5.* Similar to _n and _N, if and in only work in combination with commands.

- **list**: this command lists the contents of the Stata file you are currently using.[10]
- **generate** [*type*] newvar =exp [*if*] [*in*][21]: if you want to create a new variable or change an existing one, you use this command. I only used it when creating a new variable. [type] refers to the variable types (byte, integer, float, etc.) and =exp refers to the calculation and condition the new variable has to fulfill, e.g. *generate bmi = weight/(height^2).* The brackets [ ] symbolize that these additional specifications are optional and do not have to be included in the command.
- **label variable** *varname* [**"***label***"**][22]: the label command attaches a label to your data set or to a certain variable (syntax shown here). Especially when you have many variables and you use short abbreviated names for them, you run the risk of losing sight of which name refers to which variable. In this case, you can just have a look at the label. An example could look like this: *label variable goodstudent "Students with an average grade <= 2,5; from 1950-2010".*
- **drop and keep:** if you wish to discard a certain variable or a whole list of variables, you uses the drop command. If you wish to delete all variables except a few selected ones, you use the command *keep*.[23]
- **egen** [*type*] newvar = fcn(arguments) [*if*] [*in*] [,*options*][24]: egen's structure is similar to the one of *generate* and is used as an extension to it. The fcn refers to the action you want Stata to take, for instance **mean(exp)** creates the mean of exp (variable). Many useful egen commands have been created and are available on the user community. I have used the *egen* commands particularly when standardizing my data.[25]

---

[21] http://www.stata.com/help.cgi?label, 13.6.2010
[22] http://www.stata.com/help.cgi?label, 13.6.2010
[23] Baum (2006): 20
[24] http://www.stata.com/help.cgi?egen, 13.6.2010
[25] Kohler & Kreuter (2009): 84-86

4. Implementation

4.1. Getting started

After this short theoretical introduction I will now show how to implement the model in Stata. Stata commands will be written in *italics*. I have studied Stata software language primarily using these books:

1. *An Introduction to Modern Econometrics Using Stata* by Christopher Baum
2. *Data Analysis Using Stata* by Ulrich Kohler and Frauke Kreuter
3. *Getting Started with Stata* by StataCorp.
4. *Stata Time-Series Reference Manual* by Stata Corp.

I am not able to directly quote which command I have learned from which book because the contents are overlapping at some parts. Furthermore, I will show some excerpts from the result window to underline how Stata's commands are used.
The data I am using is a time series. After the command *use* one can make Stata identify the data as time series by introducing a date variable and the Stata command *tsset*:

- **tsset** timevar [**,** *options*]: this command declares to Stata that the data set now in use is a time series.[26]
- **format** varlist **%***fmt*: this command sets a variable's output format. %fmt refers to the format which can be a date, a string (word) or numerical.[27] In our case, we use the one for a quarterly date format *%tq.*

One can create a date variable using the command *generate* which I have mentioned in the basic command chapter. Given that we only have quarterly data, the *q(1985q1)* tells Stata to create a variable starting from the first quarter (*q*) in 1985, the _*n-1* indicates the ascending trend. The *format* command translates the created variable into a readable format.
Taking one look at the result window, we find that Stata has now created a variable time, in quarterly intervals, spanning a time horizon from the first quarter of 1985 to the second quarter in 2009.

[26] http://www.stata.com/help.cgi?tsset, 22.6.2010
[27] http://www.stata.com/help.cgi?format, 22.6.2010

```
.
. generate time = q(1985q1) + _n -1

. format time %tq

.
. tsset time
        time variable:  time, 1985q1 to 2009q2
                delta:  1 quarter
```

Figure 2: Stata result window

4.2. PCA

This section will focus on the implementation of the principal component analysis which statistical background I have explained in chapter 2.1.

Boss et al. propose to use the year-on-year changes in the macroeconomic variables' index as input for the analysis.

One can easily calculate these deltas for the macroeconomic index by using the mentioned _n option. _n refers to the current observation of the data set. If we wish to calculate the year-on-year changes of the single variables we can do so by creating a delta variable which is calculated by subtracting observation n at time t-4 from the current observation n at time t. In the result window you can see that Stata reports four missing values. These are the first four values where Stata cannot find an observation _n at time t-4. You can see the creating of two variables of the data set, delta CAC (total capital cost) and delta CPN (Private credit, amount outstanding). Because this way of implementing creates a large set of new variables, I decided to label the new variables so I can see in the variables window what variable's delta they stand for. I do so by using the *label* command.

```
. gen dcac = cac[_n] - cac[_n-4]
(4 missing values generated)

. lab var dcac "Delta CAC"

. gen dcpn = cpn[_n] - cpn[_n-4]
(4 missing values generated)

. lab var dcpn "Delta CPN"
```

Figure 3: Creating the deltas

As I cannot use missing values in the principal component analysis, I have to delete them, using the command *drop*. The term ==. refers to a missing value.

```
.
. drop if dcac == .
(4 observations deleted)
```

Figure 4: Drop observations with missing values

Before starting with the principal component analysis, the data has to be standardized according to the paper by Boss et al. by subtracting the mean from a variable and dividing it by its standard deviation.
For the standardization I was using the *egen* command which is an extension to the before mentioned *generate* command. Here again its syntax:

- **egen** *newvar* = fcn(*arguments*) [*if*] [*in*] [,*options*][28]: the function fcn, in our case the mean, determines the *arguments*, whether they are a single variable or a list of variables or others. To calculate the mean of a variable, I am creating a new “mean“ variable for each macro variable I want to use in the PCA. The result window displays the correct command, as example variable I have again used CAC and CPN.

[28] http://www.stata.com/help.cgi?egen, 24.6.2010

```
. egen cacMEAN = mean(dcac)

. egen cpnMEAN = mean(dcpn)
```

Figure 5: Egen command for calculating the mean

Then I am using the same *egen* command with a different function fcn to calculate the standard deviation, sd(*arguments*):

```
. egen cacSD = sd(dcac)

. egen cpnSD = sd(dcpn)
```

Figure 6: Egen command for calculating the standard deviation

Now I can create the standardized variables I will be using in the principal component analysis by using Stata's command *generate.* This command subtracts the mean from each observation, divides it by the standard deviation and saves the result in the new variable cacSTAND or cpnSTAND:

```
. generate cacSTAND = (dcac - cacMEAN) / cacSD

. generate cpnSTAND = (dcpn - cpnMEAN) / cpnSD
```

Figure 7: Generating the standardized variables

The standardized variables are the input for the PCA. The command for this analysis is – very simple – *pca*:

```
pca yerSTAND xtrSTAND wurydSTAND vatSTAND urxSTAND ulaSTAND totrevSTAND
pyrSTAND proSTAND pdnSTAND pcrSTAND nfnSTAND nfaSTAND mtrSTAND itrSTAND
innSTAND hicSTAND gynSTAND gdnSTAND gbSTAND ddrSTAND cpnSTAND cacSTAND
mpcSTAND
```

Figure 8: PCA

The linear combinations will be ordered according to their significance (eigenvalues). The result window in table 1 further indicates that a calculation for 24 components and 94 observations has been conducted. The proportion shows the explanatory power for each linear composition, e.g. Comp1 explains 30,51% of the total variance. The next column indicates the cumulative explanatory power. I will make the cut after the first five components which then will explain 71,1% of the total variance. Those five factors will be taken as input variable for the logistic regression.

Why use eigenvectors to calculate the principal components? Because eigenvectors are orthogonal to the matrix, no matter how many dimensions it has. This means the data can be expressed using eigenvectors instead of the x and y axes. The eigenvectors with the highest eigenvalues are the most significant principal components, indicating how the data is related along its line. Thus the eigenvectors are organized by eigenvalue. One can set a level of significance beneath which the eigenvectors are ignored. Although some information will be lost if combinations are excluded, it will not be too much information if the dropped eigenvectors have small eigenvalues.[29]

[29] http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf, 16.6.2010

Principal components/correlation — Number of obs = 94; Number of comp. = 24; Trace = 24

Rotation: (unrotated = principal) — Rho = 1.0000

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	7.32301	3.93913	0.3051	0.3051
Comp2	3.38387	.564442	0.1410	0.4461
Comp3	2.81943	.852231	0.1175	0.5636
Comp4	1.9672	.399081	0.0820	0.6456
Comp5	1.56812	.122947	0.0653	0.7109
Comp6	1.44517	.224194	0.0602	0.7711
Comp7	1.22098	.390365	0.0509	0.8220
Comp8	.830614	.122622	0.0346	0.8566
Comp9	.707992	.141713	0.0295	0.8861
Comp10	.566279	.059604	0.0236	0.9097
Comp11	.506675	.0990298	0.0211	0.9308
Comp12	.407646	.133446	0.0170	0.9478
Comp13	.274199	.0130861	0.0114	0.9592
Comp14	.261113	.0639378	0.0109	0.9701
Comp15	.197175	.0284379	0.0082	0.9783
Comp16	.168737	.0260544	0.0070	0.9853
Comp17	.142683	.0722509	0.0059	0.9913
Comp18	.070432	.0141165	0.0029	0.9942
Comp19	.0563155	.0236621	0.0023	0.9966
Comp20	.0326534	.0108427	0.0014	0.9979
Comp21	.0218107	.00713528	0.0009	0.9988
Comp22	.0146754	.00564053	0.0006	0.9994
Comp23	.00903488	.00486069	0.0004	0.9998
Comp24	.00417419	.	0.0002	1.0000

Table 1: Result window PCA

The next table shows a list of the variables’ eigenvectors, for component 1 to 12. The list for the next 12 components can be found in the appendix.

Principal components (eigenvectors)

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9	Comp10	Comp11	Comp12
yerSTAND	0.3403	0.0218	-0.0830	0.1232	0.0338	-0.0870	0.0457	0.0582	0.2955	-0.0460	-0.0679	-0.0057
xtrSTAND	0.3193	-0.0295	-0.2256	-0.0855	0.0449	0.0164	0.1351	-0.0059	0.0834	0.0516	0.0700	0.1735
wurydSTAND	-0.1105	0.1369	0.3174	0.1456	0.1416	-0.1323	-0.0599	0.5853	0.3575	-0.1483	-0.1265	0.3049
vatSTAND	0.0537	0.2594	0.1035	-0.0295	-0.4516	0.0384	-0.3315	-0.2296	0.3893	0.2902	0.1733	0.0159
urxSTAND	-0.1728	-0.0165	0.2250	-0.2780	0.0599	0.3504	0.3052	0.1510	0.2521	0.2521	0.1358	0.1651
ulaSTAND	-0.2187	-0.0648	0.0833	0.3875	-0.1647	0.2033	-0.1046	0.2065	-0.3164	0.0223	-0.1854	0.0113
totrevSTAND	-0.2028	0.3433	-0.0125	0.1260	0.2242	0.1475	0.0084	-0.0125	-0.0794	0.2590	0.3205	0.0242
pyrSTAND	0.1634	-0.1158	0.2624	0.2567	0.2694	-0.2655	-0.1790	-0.1499	0.0660	0.0836	0.1511	-0.3908
proSTAND	0.3354	0.0542	0.0897	-0.1249	0.0676	0.0383	0.0351	0.0841	0.1614	-0.0863	-0.2573	0.0471
pdnSTAND	0.0262	0.1353	-0.3996	0.2633	0.0825	0.1555	-0.1665	0.2024	0.2374	0.4197	-0.0155	-0.2241
pcrSTAND	0.1837	0.1312	0.3216	0.2250	-0.2317	-0.0300	0.2284	-0.1221	-0.1836	0.0671	0.2460	0.0873
nfnSTAND	0.1332	-0.2540	-0.1647	-0.0122	0.3416	-0.0458	-0.3067	0.2061	-0.2760	0.3303	0.1291	0.3202
nfaSTAND	-0.0481	-0.4048	-0.0506	0.1130	-0.1983	-0.2737	0.2143	0.2136	0.1040	0.2243	-0.1127	-0.3152
mtrSTAND	0.3263	0.0084	-0.1200	-0.0988	0.0722	0.0077	0.2189	-0.0737	-0.1813	0.0333	0.0504	0.1577
itrSTAND	0.2987	0.1054	0.1381	0.1697	-0.0796	0.0123	0.0333	0.1986	-0.2525	0.1752	-0.0113	0.0728
innSTAND	0.0476	-0.1917	0.2378	0.1069	0.3414	0.3781	0.3903	-0.0928	0.1127	0.1197	0.0367	-0.2792
hicSTAND	0.0143	-0.1221	-0.2402	0.3547	-0.1883	0.4692	0.0591	-0.1793	0.0507	-0.0633	-0.3037	0.1929
gynSTAND	-0.0580	0.4371	-0.1137	0.1883	0.1902	-0.0189	0.1167	0.0209	-0.0441	-0.2535	-0.1786	-0.2541
gdnSTAND	0.1545	-0.0423	0.2392	0.0101	0.2582	0.3535	-0.4724	-0.1285	0.0004	-0.2616	0.0247	-0.0457
gbSTAND	-0.0151	0.4641	-0.1922	-0.0974	0.1537	-0.1109	0.1796	0.0296	-0.1310	0.0542	-0.0595	-0.0672
ddrSTAND	0.2718	0.1396	0.2339	0.2568	-0.1697	-0.0313	0.0954	0.0988	-0.1366	0.0566	0.0251	0.0742
cpnSTAND	-0.0741	-0.1213	-0.2229	0.4448	0.1369	-0.1726	0.1539	-0.1760	0.2822	-0.2767	0.3979	0.3303
cacSTAND	0.3488	0.0541	-0.0590	-0.0457	0.0282	0.0841	-0.0383	-0.0020	0.1289	0.0328	-0.2043	-0.0541
mpcSTAND	0.1672	-0.0421	-0.2015	-0.1050	-0.2446	0.2729	-0.0271	0.4669	-0.0539	-0.3699	0.5261	-0.3162

In order to cut down the sample and use the first five factors for the logistic regression, the scores for the time period 1985-2009 have to be predicted. A way to do so is to use the *predict* command in combination with the *score* command and adding how many factors you wish to be predicted. The result window will indicate how many components are skipped:

```
. predict factor1 factor2 factor3 factor4 factor5, score
(19 components skipped)

Scoring coefficients
    sum of squares(column-loading) = 1
```

Figure 9: Predict PCA factors in Stata

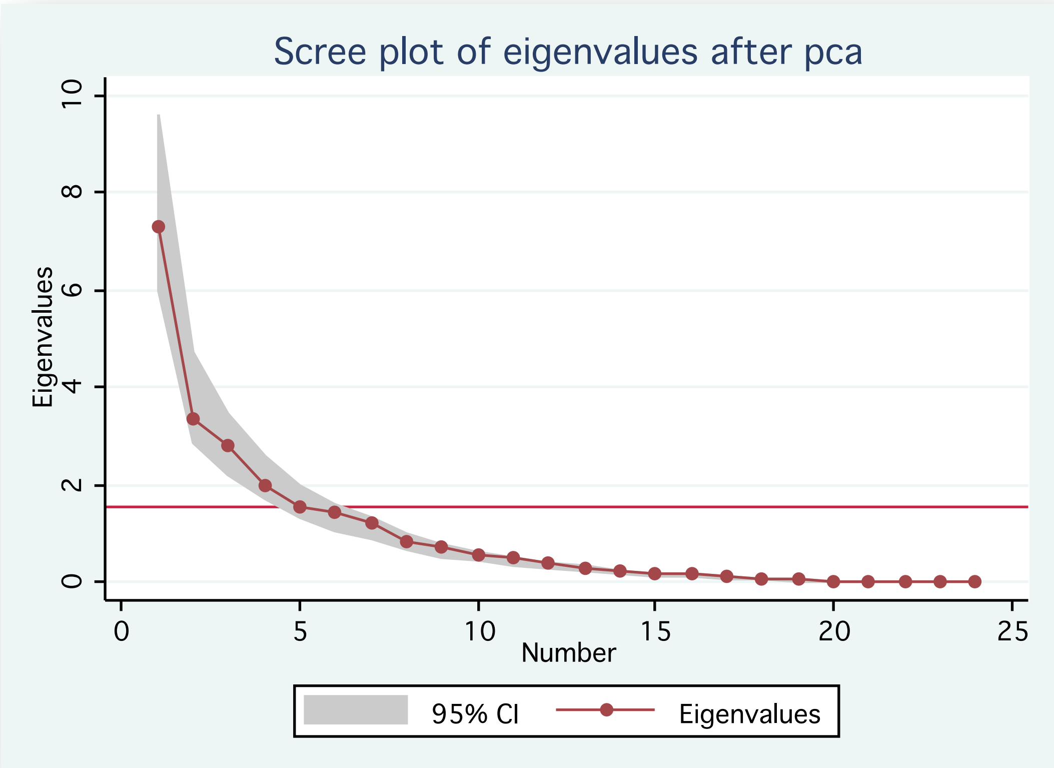

Figure 10: Plot graph of eigenvalues

On this graph you can see the principal components after the analysis. The 24 dots are ordered according to their respective eigenvalues (significance). The 19 dots below my borderline will not be included in the regression analysis.

To obtain a graph with the plotted eigenvalues, the Stata command *screeplot* is used. This command can only be used right after the analysis. Now five more variables have been added to the data set, thus it would be recommendable to label them:

```
. screeplot, yline(1) ci(het)

.
. notes factor1 : first component predicted by PCA

. notes factor2 : second component predicted by PCA

. notes factor3 : third component predicted by PCA

. notes factor4 : fourth component predicted by PCA

. notes factor5 : fifth component predicted by PCA
```

Figure 11: How to plot the PCA eigenvalue graph

Using the *scoreplot* command one can generate the typical graph of transformed observations often shown in PCA manuals. This command as well as can only be used right after the analysis has been conducted.

4.3. Logit transformation

Before entering the probabilities of default into the regression model as dependant variables, they are transformed. This process is called logit transformation. As mentioned before, the probabilities are stated in relative frequencies, not in numbers 1 or 0 (e.g. a set of data where bankruptcy equals 1, no bankruptcy equals 0).
So the question arises: How does one fit a model when the dependent variables are proportions?
One way is the logit transformation. After this transformation one can apply the OLS method on the data.
The authors Boss et al. are using an approach by Wilson (1997a and 1997b) and calculate values for the macroeconomic index y as following:

(Eq. 3) $$y_t = -\ln(\frac{1}{p_t} - 1)$$

The resulting y_t are our dependent variables that will later be part of the inputs for the regression. Pt denotes the probabilities of default.
I have created those variables in two steps using the command *generate*:

1. Calculating the term in braces $\left(\frac{1}{p} - 1\right)$
2. In the second step I took the negative natural logarithm by dividing 1 by the term in 1. as $-\ln x = \ln\left(\frac{1}{x}\right)$

```
.
. gen agr_t = ((1/PDagr)-1)

. gen prod_t = ((1/PDprod)-1)
```

```
. gen agr_logtrans = ln(1/agr_t)

. gen prod_logtrans = ln(1/prod_t)
```

Figure 12 & 13: Logit transformation

4.4. Standard regression

The least squares regression is conducted on the annual differences of the macroeconomic index:

(Eq. 7) $\Delta y_t = y_t - y_{t-4}$ [30]

This index of independent variables can be created similar to the creation of the index needed for the PCA:

```
.
. gen dfactor1 = factor1[_n] - factor1[_n-4]
(4 missing values generated)

. lab var dfactor1 "Delta Factor 1"
```

Figure 14: Creating Δy

Again, four missing values are generated for the first four quarters, so in order to use this data set, we have to cut it down by these four missing values:

```
.
. drop if dfactor1 == .
(4 observations deleted)
```

Figure 15: Drop missing values

The authors estimated the following regression model:

(Eq. 8) $$\Delta y_t = \sum_{i=\cap}^{K} \beta \Delta x_{i,t} + \varepsilon_t = X_t \beta + \varepsilon_t$$

with $\Delta x_{0,t} := 1$

Δy denotes the macroeconomic index calculated in Eq. 7 while $\Delta x_{i,t}$ denotes the year-on-year changes of macroeconomic variables and the betas are the coefficients to be estimated.[31] The index for Δx_t is calculated in the same way as Δy_t and compounds

[30] Boss et al. (2009): 92-108
[31] Boss et al. (2009): 92-108

the variables listed in chapter 2.1. Three new variables have to be created (using again command *generate*):

1. private consumption divided by the GDP
2. investment equipment divided by GDP
3. exports divided by GDP

```
.
. gen pcrgdp = pcr/yer

. gen iergdp = ier/yer

. gen xtrgdp = xtr/yer
```

Figure 16: Generate GDP variables

Having calculated both indices, the ordinary least squares optimization can be conducted using the command *mvreg* for more than one dependant variable.

```
. mvreg agr_logtrans prod_logtrans srv_logtrans cstr_logtrans trad_logtrans tour_logtrans
trsp_logtrans oth_logtrans = dyer dpcr dpcrgdp durx
> dpro dpyr ditr dier diergdp dula dxtr dxtrgdp dsti dlti dstireal dltireal dipexe doil dhic
dfactor1 dfactor2 dfactor3 dfactor4 dfactor5
```

Figure 17: Conducting the regression

The dependant variables are separated by an “=” from the independent variables in the command. For a single dependent variable the Stata command would be *regress*. In the results window the estimates for each dependant variable is shown:

Equation	Obs	Parms	RMSE	"R-sq"	F	P
agr_logtrans	86	25	.2945771	0.7477	7.530356	0.0000
prod_logtr~s	86	25	.1504687	0.6443	4.60441	0.0000
srv_logtrans	86	25	.1587689	0.7219	6.596254	0.0000
cstr_logtr~s	86	25	.1788863	0.8099	10.83165	0.0000
trad_logtr~s	86	25	.1418492	0.6975	5.861176	0.0000
tour_logtr~s	86	25	.1550704	0.7068	6.128517	0.0000
trsp_logtr~s	86	25	.1655162	0.7025	6.002822	0.0000
oth_logtrans	86	25	.1135823	0.7398	7.225494	0.0000
overall_lo~s	86	25	.1158798	0.7278	6.796578	0.0000

Table 2: Summary of estimates

In this table, the goodness of fit and the statistical significance for the regressions are summed up. Stata has used the 86 observations of macroeconomic variables or default rates per industrial sector for this analysis. Altogether it has used 25 parameters (variables) for the estimation.

RMSE denotes the square root of the mean squared error and is the standard deviation of the error term not explained by the model.

One measure of the fit is the coefficient of determination R^2 which tells us how much better we are able to predict the dependant variable Y, our probabilities of default, with the data we have got than we could predict without any information at all.[32]

The P value estimates the probability of the estimated coefficient being zero: $\mathrm{P}\left[\hat{\beta}_i = 0\right] = p \leq 0.05$. The lower the p-value, the less likely it is for the coefiicient to be zero and the more statistically significant the result becomes. The significance level α is 0.05.

The F-test measures the probability for all coefficients to be zero: $P\left[\hat{\beta} = 0\right] = p$ for $\hat{\beta}_i \ldots \hat{\beta}_k$.

The F-test indicates that the model is statistically significant.

[32] Menard (2002): 3-4

	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
agr_logtrans						
dyer	.0007685	.0009652	0.80	0.429	-.0011616	.0026986
dpcr	-.0014573	.0019327	-0.75	0.454	-.005322	.0024073
dpcrgdp	76.71049	88.5995	0.87	0.390	-100.4551	253.8761
durx	-.1615985	.1580942	-1.02	0.311	-.4777274	.1545304
dpro	3.408126	.9467534	3.60	0.001	1.514976	5.301277
dpyr	-.0001059	.0001434	-0.74	0.463	-.0003927	.0001809
ditr	.0009072	.0004773	1.90	0.062	-.0000472	.0018615
dier	.0021273	.0031467	0.68	0.502	-.0041649	.0084194
diergdp	-97.96079	152.5778	-0.64	0.523	-403.0589	207.1373
dula	5.25846	8.620218	0.61	0.544	-11.97873	22.49565
dxtr	-.0008762	.0003584	-2.44	0.017	-.0015929	-.0001595
dxtrgdp	49.49865	23.30388	2.12	0.038	2.899665	96.09764
dsti	-1.537143	.7639131	-2.01	0.049	-3.064682	-.0096046
dlti	1.21431	.763962	1.59	0.117	-.3133269	2.741946
dstireal	1.31409	.7632332	1.72	0.090	-.2120886	2.84027
dltireal	-1.245972	.7625594	-1.63	0.107	-2.770804	.2788597
dipexe	.0027437	.0249592	0.11	0.913	-.0471653	.0526528
doil	-.0036394	.0076345	-0.48	0.635	-.0189056	.0116268
dhic	.1657262	.089438	1.85	0.069	-.0131162	.3445686
dfactor1	-.097524	.0463038	-2.11	0.039	-.1901142	-.0049338
dfactor2	-.0108998	.0265636	-0.41	0.683	-.0640169	.0422173
dfactor3	-.0006041	.0397759	-0.02	0.988	-.080141	.0789327
dfactor4	-.0451572	.0643751	-0.70	0.486	-.1738831	.0835687
dfactor5	-.0055038	.0360021	-0.15	0.879	-.0774945	.0664869
_cons	-7.196281	.3037241	-23.69	0.000	-7.803615	-6.588947

Table 3: Regression estimates for the agricultural sector

Shown in the table above are the estimates for the agricultural sector. The t-values show that five variables are statistically significant (not counting the constant): The average labor productivity (PRO), the exports (XTR), the exports in relation to the GDP (XTRGDP), the nominal short-term interest rate (STI) and the first PCA factor. The R-squared value in table 1 indicates that about 74% of the probability of default's variance is explained by the model (the independent variables).

The single coefficients denote how much the probability of default in the agricultural sector would rise if the coefficient rises one unit in standard deviation holding all other variables constant. However, it is hard to find an economic interpretation of them without taking the predictor's standard deviation into account. Thus, I will multiply the statistically significant variables' coefficients with the predictors' standard deviation. I will calculate the economic significance for all statistically significant variables. Economic significance levels are not strictly defined in literature. In my thesis I will count values that are larger than 1 or smaller than -1 as economically significant, values in the interval -1 < x < 1 will not be regarded significant.

Variable	**Coefficient β_i**	**Standard deviation σ_i**	$\beta_i*\sigma_i$
Labor productivity	3.408126	1.60257345	5.461772242
Exports	-0.0008762	9133.39166	-8.002677772
Exports/GDP	49.49865	0.10447437	5.171340275
Short-term interest	-1.537143	2.11792016	-3.255546149
1st PCA factor	-0.097524	2.69167304	-0.262502722

Table 4: Economic significance in the agricultural sector

All variables except the PCA factor are economically significant in the agricultural sector.

overall_lo~s						
dyer	-.0007636	.0003797	-2.01	0.049	-.0015228	-4.31e-06
dpcr	.0004769	.0007603	0.63	0.533	-.0010433	.0019972
dpcrgdp	-35.69036	34.853	-1.02	0.310	-105.3832	34.00251
durx	-.1854817	.0621906	-2.98	0.004	-.3098395	-.0611238
dpro	1.894256	.372431	5.09	0.000	1.149534	2.638978
dpyr	.0000115	.0000564	0.20	0.839	-.0001013	.0001243
ditr	.0002125	.0001877	1.13	0.262	-.0001629	.0005879
dier	.0010113	.0012378	0.82	0.417	-.0014639	.0034864
diergdp	-40.03566	60.02057	-0.67	0.507	-160.0542	79.98288
dula	-2.126636	3.390995	-0.63	0.533	-8.90735	4.654078
dxtr	.0001768	.000141	1.25	0.215	-.0001051	.0004587
dxtrgdp	-7.310287	9.167209	-0.80	0.428	-25.64125	11.02068
dsti	-.0872652	.3005058	-0.29	0.772	-.6881637	.5136332
dlti	-.0722362	.300525	-0.24	0.811	-.6731732	.5287007
dstireal	-.0243693	.3002383	-0.08	0.936	-.6247329	.5759943
dltireal	.0551737	.2999733	0.18	0.855	-.5446599	.6550074
dipexe	.0135541	.0098184	1.38	0.172	-.0060789	.0331872
doil	.0010291	.0030033	0.34	0.733	-.0049763	.0070345
dhic	.0847602	.0351829	2.41	0.019	.0144077	.1551127
dfactor1	-.0681931	.0182149	-3.74	0.000	-.1046159	-.0317702
dfactor2	.002961	.0104495	0.28	0.778	-.017934	.0238561
dfactor3	.0329201	.0156469	2.10	0.040	.0016321	.064208
dfactor4	.006175	.0253237	0.24	0.808	-.0444628	.0568128
dfactor5	-.0192514	.0141624	-1.36	0.179	-.0475709	.0090681
_cons	-5.557499	.1194781	-46.51	0.000	-5.796411	-5.318588

Table 5: Regression table for overall probabilities of default

Table 5 shows the regression estimates for the overall probabilities of default. The t-values indicate that the following six variables are statistically significant in the overall sector: the GDP (YER), which is significant in five out of nine sectors, the unemployment rate (URX), which is significant in seven sectors, the average labor productivity (PRO), which is significant in all sectors except the construction sector, the harmonized consumer price index (HIC), which is significant in five sectors and PCA factor one and three. The first PCA factor is significant in seven, the third in four sectors.

The other sectors' regression tables can be found in the appendix. I will brievly sum up the results here: In all estimation tables, the constant term β_0 (_cons) differs significantly from zero.

Variable	Coefficient β_i	Standard deviation σ_i	β_i*σ_i
GDP	-0.0007636	8768.73881	-6.695808955
Unemployment rate	-0.1854817	0.63857648	-0.118444251
Labor productivity	1.894256	1.60257345	3.035684373
CPI	0.0847602	11.7520547	0.996106507
1st PCA factor	-0.0681931	2.69167304	-0.183553529
3rd PCA factor	0.0329201	1.67016128	0.054981876

Table 6: Overall economic significance

Overall, the GDP and the labor productivity are economically significant here. If statistically significant, these variables also prove to be economically significant in all sectors. The consumer price index is only marginally not significant here with a value of 0.996. The unemployment rate is not economically significant in any sectors even if it is statistically significant, except the service sector. None of the outliers (short-term interest rate or industrial production) are economically significant.

Production:
In this sector only three variables are statistically significant: The labor productivity, and PCA factors one and five.

Variable	Coefficient β_i	Standard deviation σ_i	β_i*σ_i
Labor productivity	2.105256	1.60257345	3.373827371
1st PCA factor	-0.0699686	2.69167304	-0.188332594
5th PCA factor	-0.0410859	1.24556743	-0.051175259

Table 7: Economic significance in the production sector

In the production sector, only one variable shows an economic significance: the labor productivity with a value of 3.37. PCA factors one and five have values close to zero and thus are not economically significant.

Construction: The construction sector is the only one where labor productivity is (by far) not significant. Here, five other variables show low t-values: the GDP, the consumer price index, the exports and the exports in relation to the GDP, and the third PCA factor.

Variable	**Coefficient β_i**	**Standard deviation σ_i**	**β_i*σ_i**
GDP	-0.0014869	8768.73881	-13.03823774
Unemployment rate	-0.2475023	0.63857648	-0.158049148
Exports	0.0009048	9133.39166	8.263892774
Exports/GDP	-44.42602	0.10447437	-4.641380451
3rd PCA factor	0.0661315	1.67016128	0.110450271

Table 8: Economic significance in the construction

In the construction sector, three out of five statistically significant variables are also economically significant: the GDP, the exports and the exports in relation to the GDP with values -13.03, 8.23 and -4.64, the signs indicating the direction of the relationship.

Trading: In the trade sector, the labor productivity is statistically highly significant, as well as the consumer price index and the first PCA factor. Also the unemployment rate shows a low t-value.

Variable	**Coefficient β_i**	**Standard deviation σ_i**	**β_i*σ_i**
Unemployment rate	-0.1582224	0.63857648	-0.101037103
Labor productivity	2.538579	1.60257345	4.068259306
CPI	0.1550849	11.7520547	1.822566228

1st PCA factor	-0.0784635	2.69167304	-0.211198088

Table 9: Economic significance in the trade sector

In the trade sector labor productivity shows the highest significance with a value of 4.06. Also, the consumer price index is significant.

Tourism: Here, six factors can be found statistically significant: The GDP, the unemployment rate, the labor productivity, PCA factor one and five, and one outlier which is only significant in this sector: the industrial production (IPEXE).

Variable	Coefficient β_i	Standard deviation σ_i	$\beta_i*\sigma_i$
GDP	-0.0012511	8768.73881	-10.97056913
Unemployment rate	-0.2182199	0.63857648	-0.139350096
Labor productivity	2.752687	1.60257345	4.411383102
Industrial production	0.0345029	18.7853606	0.648149418
1st PCA factor	-0.0963286	2.69167304	-0.259285096
5th PCA factor	-0.047555	1.24556743	-0.059232959

Table 10: Economic significance in the tourism sector

Here, two variables are significant, the GDP strongly so with a value of -10.97 and the labor productivity with a value of 4.41.

Transport: In the transport sector we can find the highest amount of significant variables: As in most sectors, the GDP is statistically significant, as well as the consumer price index, the labor productivity, the exports and the exports in relation to the GDP, the unemployment rate and the PCA factor three. Also, the consumer price index is highly significant with a t-value of 0.000.

Variable	Coefficient β_i	Standard deviation σ_i	$\beta_i*\sigma_i$
GDP	-0.0012549	8768.73881	-11.00389033
Unemployment rate	-0.2795028	0.63857648	-0.178483914
Labor productivity	1.164778	1.60257345	1.866642298
Exports	0.0005843	9133.39166	5.336640747
Exports/GDP	-27.18503	0.10447437	-2.840138883
CPI	0.19844	11.7520547	2.332077735
3rd PCA factor	0.0618969	1.67016128	0.103377806

Table 11: Economic significance in the transport sector

In the transport sector, only two out of seven statistically significant variables are not economically significant: the PCA factor and the unemployment rate. GDP, labor productivity, CPI and the export variables are significant.

Others: The labor productivity and the PCA factor one both show t-values of 0.000 and thus are highly statistically significant. Furthermore, the GDP, the unemployment rate and the consumer price index are significant. With a t-value of 0.04 the PCA factor three is also significant at the 0.05 level.

Variable	Coefficient β_i	Standard deviation σ_i	$\beta_i*\sigma_i$
GDP	-0.0008531	8768.73881	-7.480611079
Unemployment rate	-0.1807732	0.63857648	-0.115437514
Labor productivity	1.901183	1.60257345	3.046785399
CPI	0.0817216	11.7520547	0.960396713
1st PCA factor	-0.0648637	2.69167304	-0.174591873
3rd PCA factor	0.0322763	1.67016128	0.053906627

Table 12: Economic significance in the other sectors

Here, only the GDP and the labor productivity are economically significant. Again, no PCA factor has a value significally different from zero.

Services: In the service sector we can only find four significant variables: the unemployment rate, the labor productivity, the consumer price index and the PCA factor one. The labor productivity and the consumer price index short the smallest t-values of 0.000.

Variable	**Coefficient β_i**	**Standard deviation σ_i**	$\beta_i * \sigma_i$
Labor productivity	-0.184155	0.63857648	-0.117597052
CPI	2.485435	1.60257345	3.983092143
Unemployment rate	0.1791365	11.7520547	2.105221947
1st PCA factor	-0.0752793	2.69167304	-0.202627262

Table 13: Economic significance in the service sector

In the service sector we find two values that are economically significant: labor productivity and consumer price index. Again, PCA factor one is not significant, neither is the unemployment rate.

Conclusion

For the overall results (appendix) we find that six macroeconomic variables are statistically significant: the GDP, the unemployment rate, the labor productivity, the consumer price index and PCA factors one and three.
The regression tables show that the GDP is statistically significant in five out of nine sectors. The exceptions are production, service, agriculture and trade sector. In the agriculture and trade sectors the t-values for the GDP have been much higher than 0.05, but in the production and service sector, this variable has just narrowly missed the significance level.

The unemployment rate is statistically significant in seven out of nine sectors, the exceptions being the production sector and the agriculture sector.
The next variable overall statistically significant is the consumer price index in five out of nine sectors.
While the second and the fourth PCA factor do not prove to be significant at all, the first, thrid and fifth factor are significant in different business sectors, the 1st factor in seven sectors, the third in four sectors.
Exports are significant in some sectors, but not in all. If they are significant, so are the exports in relation to the GDP, even if the GDP itself is not significant.
The industrial production is significant only in the tourism sector.
Another outlier is the short-term interest rate which is only significant in the agricultural sector.

From an economic viewpoint, very few variables are both statistically and economically significant.
It is remarkable that there is not one PCA factor that is economically significant in any sector although there is at least one PCA factor statistically significant in each. Overall, the GDP and the labor productivity are economically significant here. If statistically significant, these variables also prove to be economically significant in all sectors. The consumer price index is only marginally not significant here with a value of 0.996. The unemployment rate is not economically significant in any sectors even if it is statistically significant, except the service sector. None of the outliers (short-term interest rate or industrial production) are economically significant.

One shortcoming in this model is the choice of transformation. In order to find out which transformation of the variables' distribution would be the most suitable, the log-transformation is not always the best option.
A way to examine the transformations is the Stata command *ladder*:

Transformation	formula	chi2(2)	P(chi2)
cubic	CAC^3	13.86	0.001
square	CAC^2	12.44	0.002
identity	CAC	11.21	0.004
square root	sqrt(CAC)	10.68	0.005
log	log(CAC)	10.19	0.006
1/(square root)	1/sqrt(CAC)	9.76	0.008
inverse	1/CAC	9.39	0.009
1/square	1/(CAC^2)	8.79	0.012
1/cubic	1/(CAC^3)	8.41	0.015

Table 14: Ladder command

```
. ladder CAC
```

Figure 18: Stata command *ladder*

This table shows normality tests for various transformations to find out which is the most suitable to make the variable more normally distributed. Thus, the log-transformation is not the best transformation for the variable CAC, not having the smallest chi-square. To verify this result one can plot a graph showing all transformations here using the Stata command *gladder*.

Although my data set compounds fifteen years less than the original data set, the results are more or less the same: strong GDP-ratio coefficients and the same significant variables in the different business sectors. To assess the flexibility of this model, a stress test similar to the one conducted by Boss et al. should be performed on the data set. The authors simulate an economic recession in order to test the models.

However, this model still does not offer a solution how to link the credit cycle and the business cycle. Given the importance due to the recent economic development, further research is recommended to examine the impact of macroeconomic variables on default probabilities.

References

Methodical References

Boss, M. et al. (2009). Modeling Credit Risk through the Austrian Business Cycle: An Update of the OeNB Model. *Financial Stability Report, 17*. 92-108.

Draper, N. & Smith, H. (1981). *Applied Regression Analysis*. New York: Wiley & Sons.

Kleinbaum, D. (1996). *Logistic Regression. A Self-Learning Text.* New York: Springer-Verlag.

Mc Dowell, A. & Cox, N.J. (2004). Logit transformation. Retrieved June 12, 2010, from http://www.stata.com/support/faqs/stat/logit.html

Menard, S. (2002). *Logistic Regression Analysis*. London: Sage.

Papke, L. & Wooldridge, J. 1996. Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. *Journal of Applied Econometrics, 11*. 619-663.

Smith, L. (2002). Retrieved June 16, 2010, from http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

Stata References

Baum, C.F. (2006). *An Introduction to Modern Econometrics Using Stata*. Texas: Stata Press.

Kohler, U. & Kreuter, F. (2009). *Data Analysis Using Stata*. Texas: Stata Press.

Stata Corp. (2007). *Stata Statistical Software: Release 10. Getting Started With Stata for Macintosh*. Texas: Stata Press.

Stata Corp. (2007). *Stata Time-Series Refrence Manual: Release 10*. Texas: Stata Press.

Stata Corp. (n.d.) In Stata 11 help. Retrieved June 13, 2010, from http://www.stata.com/help.cgi?clear

StataCorp. (n.d.) In *Stata 11 help*. Retrieved June 13, 2010, from http://www.stata.com/help.cgi?label

StataCorp. (n.d.) In *Stata 11 help*. Retrieved June 13, 2010, from http://www.stata.com/help.cgi?egen

StataCorp.(n.d.) In *Stata 11 help*. Retrieved June 22, 2010, from http://www.stata.com/help.cgi?tsset

StataCorp. (n.d.) In *Stata 11 help*. Retrieved June 22, 2010, from http://www.stata.com/help.cgi?format

StataCorp. (n.d.) In *Stata 11 help*. Retrieved June 12, 2010, from http://www.stata.com/help.cgi?glm

StataCorp. (n.d.) In *Stata 11 help*. Retrieved June 14, 2010, from http://www.stata.com/help.cgi?vce_option

Appendix I

Table 1: Result window for PCA:

Principal components/correlation

Number of obs = 94
Number of comp. = 24
Trace = 24
Rotation: (unrotated = principal)
Rho = 1.0000

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	7.32301	3.93913	0.3051	0.3051
Comp2	3.38387	.564442	0.1410	0.4461
Comp3	2.81943	.852231	0.1175	0.5636
Comp4	1.9672	.399081	0.0820	0.6456
Comp5	1.56812	.122947	0.0653	0.7109
Comp6	1.44517	.224194	0.0602	0.7711
Comp7	1.22098	.390365	0.0509	0.8220
Comp8	.830614	.122622	0.0346	0.8566
Comp9	.707992	.141713	0.0295	0.8861
Comp10	.566279	.059604	0.0236	0.9097
Comp11	.506675	.0990298	0.0211	0.9308
Comp12	.407646	.133446	0.0170	0.9478
Comp13	.274199	.0130861	0.0114	0.9592
Comp14	.261113	.0639378	0.0109	0.9701
Comp15	.197175	.0284379	0.0082	0.9783
Comp16	.168737	.0260544	0.0070	0.9853
Comp17	.142683	.0722509	0.0059	0.9913
Comp18	.070432	.0141165	0.0029	0.9942
Comp19	.0563155	.0236621	0.0023	0.9966
Comp20	.0326534	.0108427	0.0014	0.9979
Comp21	.0218107	.00713528	0.0009	0.9988
Comp22	.0146754	.00564053	0.0006	0.9994
Comp23	.00903488	.00486069	0.0004	0.9998
Comp24	.00417419	.	0.0002	1.0000

Table 2: Summary of estimates

Equation	Obs	Parms	RMSE	"R-sq"	F	P
agr_logtrans	86	25	.2945771	0.7477	7.530356	0.0000
prod_logtr~s	86	25	.1504687	0.6443	4.60441	0.0000
srv_logtrans	86	25	.1587689	0.7219	6.596254	0.0000
cstr_logtr~s	86	25	.1788863	0.8099	10.83165	0.0000
trad_logtr~s	86	25	.1418492	0.6975	5.861176	0.0000
tour_logtr~s	86	25	.1550704	0.7068	6.128517	0.0000
trsp_logtr~s	86	25	.1655162	0.7025	6.002822	0.0000
oth_logtrans	86	25	.1135823	0.7398	7.225494	0.0000
overall_lo~s	86	25	.1158798	0.7278	6.796578	0.0000

Table 3: Regression estimates for the agricultural sector

	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
agr_logtrans						
dyer	.0007685	.0009652	0.80	0.429	-.0011616	.0026986
dpcr	-.0014573	.0019327	-0.75	0.454	-.005322	.0024073
dpcrgdp	76.71049	88.5995	0.87	0.390	-100.4551	253.8761
durx	-.1615985	.1580942	-1.02	0.311	-.4777274	.1545304
dpro	3.408126	.9467534	3.60	0.001	1.514976	5.301277
dpyr	-.0001059	.0001434	-0.74	0.463	-.0003927	.0001809
ditr	.0009072	.0004773	1.90	0.062	-.0000472	.0018615
dier	.0021273	.0031467	0.68	0.502	-.0041649	.0084194
diergdp	-97.96079	152.5778	-0.64	0.523	-403.0589	207.1373
dula	5.25846	8.620218	0.61	0.544	-11.97873	22.49565
dxtr	-.0008762	.0003584	-2.44	0.017	-.0015929	-.0001595
dxtrgdp	49.49865	23.30388	2.12	0.038	2.899665	96.09764
dsti	-1.537143	.7639131	-2.01	0.049	-3.064682	-.0096046
dlti	1.21431	.763962	1.59	0.117	-.3133269	2.741946
dstireal	1.31409	.7632332	1.72	0.090	-.2120886	2.84027
dltireal	-1.245972	.7625594	-1.63	0.107	-2.770804	.2788597
dipexe	.0027437	.0249592	0.11	0.913	-.0471653	.0526528
doil	-.0036394	.0076345	-0.48	0.635	-.0189056	.0116268
dhic	.1657262	.089438	1.85	0.069	-.0131162	.3445686
dfactor1	-.097524	.0463038	-2.11	0.039	-.1901142	-.0049338
dfactor2	-.0108998	.0265636	-0.41	0.683	-.0640169	.0422173
dfactor3	-.0006041	.0397759	-0.02	0.988	-.080141	.0789327
dfactor4	-.0451572	.0643751	-0.70	0.486	-.1738831	.0835687
dfactor5	-.0055038	.0360021	-0.15	0.879	-.0774945	.0664869
_cons	-7.196281	.3037241	-23.69	0.000	-7.803615	-6.588947

Table 4: Economic significance in the agricultural sector

Variable	Coefficient β_i	Standard deviation σ_i	$\beta_i*\sigma_i$

Labor productivity	3.408126	1.60257345	5.461772242
Exports	-0.0008762	9133.39166	-8.002677772
Exports/GDP	49.49865	0.10447437	5.171340275
Short-term interest	-1.537143	2.11792016	-3.255546149
1st PCA factor	-0.097524	2.69167304	-0.262502722

Table 5: Regression table for overall probabilities of default

```
overall_lo~s |
        dyer |  -.0007636   .0003797   -2.01   0.049    -.0015228   -4.31e-06
        dpcr |   .0004769   .0007603    0.63   0.533    -.0010433    .0019972
     dpcrgdp |  -35.69036     34.853   -1.02   0.310    -105.3832    34.00251
        durx |  -.1854817   .0621906   -2.98   0.004    -.3098395   -.0611238
        dpro |   1.894256    .372431    5.09   0.000     1.149534    2.638978
        dpyr |   .0000115   .0000564    0.20   0.839    -.0001013    .0001243
        ditr |   .0002125   .0001877    1.13   0.262    -.0001629    .0005879
        dier |   .0010113   .0012378    0.82   0.417    -.0014639    .0034864
     diergdp |  -40.03566   60.02057   -0.67   0.507    -160.0542    79.98288
        dula |  -2.126636   3.390995   -0.63   0.533     -8.90735    4.654078
        dxtr |   .0001768    .000141    1.25   0.215    -.0001051    .0004587
     dxtrgdp |  -7.310287   9.167209   -0.80   0.428    -25.64125    11.02068
        dsti |  -.0872652   .3005058   -0.29   0.772    -.6881637    .5136332
        dlti |  -.0722362    .300525   -0.24   0.811    -.6731732    .5287007
    dstireal |  -.0243693   .3002383   -0.08   0.936    -.6247329    .5759943
    dltireal |   .0551737   .2999733    0.18   0.855    -.5446599    .6550074
      dipexe |   .0135541   .0098184    1.38   0.172    -.0060789    .0331872
        doil |   .0010291   .0030033    0.34   0.733    -.0049763    .0070345
        dhic |   .0847602   .0351829    2.41   0.019     .0144077    .1551127
    dfactor1 |  -.0681931   .0182149   -3.74   0.000    -.1046159   -.0317702
    dfactor2 |    .002961   .0104495    0.28   0.778     -.017934    .0238561
    dfactor3 |   .0329201   .0156469    2.10   0.040     .0016321     .064208
    dfactor4 |    .006175   .0253237    0.24   0.808    -.0444628    .0568128
    dfactor5 |  -.0192514   .0141624   -1.36   0.179    -.0475709    .0090681
       _cons |  -5.557499   .1194781  -46.51   0.000    -5.796411   -5.318588
-----------------------------------------------------------------------------
```

Table 6: Overall economic significance

Variable	Coefficient β_i	Standard deviation σ_i	$\beta_i*\sigma_i$
GDP	-0.0007636	8768.73881	-6.695808955
Unemployment rate	-0.1854817	0.63857648	-0.118444251

Labor productivity	1.894256	1.60257345	3.035684373
CPI	0.0847602	11.7520547	0.996106507
1st PCA factor	-0.0681931	2.69167304	-0.183553529
3rd PCA factor	0.0329201	1.67016128	0.054981876

Table 7: Economic significance in the production sector

Variable	Coefficient β_i	Standard deviation σ_i	$\beta_i*\sigma_i$
Labor productivity	2.105256	1.60257345	3.373827371
1st PCA factor	-0.0699686	2.69167304	-0.188332594
5th PCA factor	-0.0410859	1.24556743	-0.051175259

Table 8: Economic significance in the construction

Variable	Coefficient β_i	Standard deviation σ_i	$\beta_i*\sigma_i$
GDP	-0.0014869	8768.73881	-13.03823774
Unemployment rate	-0.2475023	0.63857648	-0.158049148
Exports	0.0009048	9133.39166	8.263892774
Exports/GDP	-44.42602	0.10447437	-4.641380451
3rd PCA factor	0.0661315	1.67016128	0.110450271

Table 9: Economic significance in the trade sector

Variable	Coefficient β_i	Standard deviation σ_i	$\beta_i*\sigma_i$
Unemployment rate	-0.1582224	0.63857648	-0.101037103
Labor productivity	2.538579	1.60257345	4.068259306
CPI	0.1550849	11.7520547	1.822566228
1st PCA factor	-0.0784635	2.69167304	-0.211198088

Table 10: Economic significance in the tourism sector

Variable	Coefficient β_i	Standard	$\beta_i*\sigma_i$

		deviation σ_i	
GDP	-0.0012511	8768.73881	-10.97056913
Unemployment rate	-0.2182199	0.63857648	-0.139350096
Labor productivity	2.752687	1.60257345	4.411383102
Industrial production	0.0345029	18.7853606	0.648149418
1st PCA factor	-0.0963286	2.69167304	-0.259285096
5th PCA factor	-0.047555	1.24556743	-0.059232959

Table 11: Economic significance in the transport sector

Variable	**Coefficient β_i**	**Standard deviation σ_i**	$\beta_i * \sigma_i$
GDP	-0.0012549	8768.73881	-11.00389033
Unemployment rate	-0.2795028	0.63857648	-0.178483914
Labor productivity	1.164778	1.60257345	1.866642298
Exports	0.0005843	9133.39166	5.336640747
Exports/GDP	-27.18503	0.10447437	-2.840138883
CPI	0.19844	11.7520547	2.332077735
3rd PCA factor	0.0618969	1.67016128	0.103377806

Table 12: Economic significance in the other sectors

Variable	**Coefficient β_i**	**Standard deviation σ_i**	$\beta_i * \sigma_i$
GDP	-0.0008531	8768.73881	-7.480611079
Unemployment rate	-0.1807732	0.63857648	-0.115437514
Labor productivity	1.901183	1.60257345	3.046785399
CPI	0.0817216	11.7520547	0.960396713
1st PCA factor	-0.0648637	2.69167304	-0.174591873
3rd PCA factor	0.0322763	1.67016128	0.053906627

Table 13: Economic significance in the service sector

Variable	**Coefficient β_i**	**Standard deviation σ_i**	$\beta_i*\sigma_i$
Labor productivity	-0.184155	0.63857648	-0.117597052
CPI	2.485435	1.60257345	3.983092143
Unemployment rate	0.1791365	11.7520547	2.105221947
1st PCA factor	-0.0752793	2.69167304	-0.202627262

Table 14: Ladder command

Transformation	formula	chi2(2)	P(chi2)
cubic	CAC^3	13.86	0.001
square	CAC^2	12.44	0.002
identity	CAC	11.21	0.004
square root	sqrt(CAC)	10.68	0.005
log	log(CAC)	10.19	0.006
1/(square root)	1/sqrt(CAC)	9.76	0.008
inverse	1/CAC	9.39	0.009
1/square	1/(CAC^2)	8.79	0.012
1/cubic	1/(CAC^3)	8.41	0.015

Full PCA list

Principal components (eigenvectors)

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9	Comp10	Comp11	Comp12
yerSTAND	0.3403	0.0218	-0.0830	0.1232	0.0338	-0.0870	0.0457	0.0582	0.2955	-0.0460	-0.0679	-0.0057
xtrSTAND	0.3193	-0.0295	-0.2256	-0.0855	0.0449	0.0164	0.1351	-0.0059	0.0834	0.0516	0.0700	0.1735
wurydSTAND	-0.1105	0.1369	0.3174	0.1456	0.1416	-0.1323	-0.0599	0.5853	0.3575	-0.1483	-0.1265	0.3049
vatSTAND	0.0537	0.2594	0.1035	-0.0295	-0.4516	0.0384	-0.3315	-0.2296	0.3893	0.2902	0.1733	0.0159
urxSTAND	-0.1728	-0.0165	0.2250	-0.2780	0.0599	0.3504	0.3052	0.1510	0.2521	0.2521	0.1358	0.1651
ulaSTAND	-0.2187	-0.0648	0.0833	0.3875	-0.1647	0.2033	-0.1046	0.2065	-0.3164	0.0223	-0.1854	0.0113
totrevSTAND	-0.2028	0.3433	-0.0125	0.1260	0.2242	0.1475	0.0084	-0.0125	-0.0794	0.2590	0.3205	0.0242
pyrSTAND	0.1634	-0.1158	0.2624	0.2567	0.2694	-0.2655	-0.1790	-0.1499	0.0660	0.0836	0.1511	-0.3908
proSTAND	0.3354	0.0542	0.0897	-0.1249	0.0676	0.0383	0.0351	0.0841	0.1614	-0.0863	-0.2573	0.0471
pdnSTAND	0.0262	0.1353	-0.3996	0.2633	0.0825	0.1555	-0.1665	0.2024	0.2374	0.4197	-0.0155	-0.2241
pcrSTAND	0.1837	0.1312	0.3216	0.2250	-0.2317	-0.0300	0.2284	-0.1221	-0.1836	0.0671	0.2460	0.0873
nfnSTAND	0.1332	-0.2540	-0.1647	-0.0122	0.3416	-0.0458	-0.3067	0.2061	-0.2760	0.3303	0.1291	0.3202
nfaSTAND	-0.0481	-0.4048	-0.0506	0.1130	-0.1983	-0.2737	0.2143	0.2136	0.1040	0.2243	-0.1127	-0.3152
mtrSTAND	0.3263	0.0084	-0.1200	-0.0988	0.0722	0.0077	0.2189	-0.0737	-0.1813	0.0333	0.0504	0.1577
itrSTAND	0.2987	0.1054	0.1381	0.1697	-0.0796	0.0123	0.0333	0.1986	-0.2525	0.1752	-0.0113	0.0728
innSTAND	0.0476	-0.1917	0.2378	0.1069	0.3414	0.3781	0.3903	-0.0928	0.1127	0.1197	0.0367	-0.2792
hicSTAND	0.0143	-0.1221	-0.2402	0.3547	-0.1883	0.4692	0.0591	-0.1793	0.0507	-0.0633	-0.3037	0.1929
gynSTAND	-0.0580	0.4371	-0.1137	0.1883	0.1902	-0.0189	0.1167	0.0209	-0.0441	-0.2535	-0.1786	-0.2541
gdnSTAND	0.1545	-0.0423	0.2392	0.0101	0.2582	0.3535	-0.4724	-0.1285	0.0004	-0.2616	0.0247	-0.0457
gbSTAND	-0.0151	0.4641	-0.1922	-0.0974	0.1537	-0.1109	0.1796	0.0296	-0.1310	0.0542	-0.0595	-0.0672
ddrSTAND	0.2718	0.1396	0.2339	0.2568	-0.1697	-0.0313	0.0954	0.0988	-0.1366	0.0566	0.0251	0.0742
cpnSTAND	-0.0741	-0.1213	-0.2229	0.4448	0.1369	-0.1726	0.1539	-0.1760	0.2822	-0.2767	0.3979	0.3303
cacSTAND	0.3488	0.0541	-0.0590	-0.0457	0.0282	0.0841	-0.0383	-0.0020	0.1289	0.0328	-0.2043	-0.0541
mpcSTAND	0.1672	-0.0421	-0.2015	-0.1050	-0.2446	0.2729	-0.0271	0.4669	-0.0539	-0.3699	0.5261	-0.3162

Variable	Comp13	Comp14	Comp15	Comp16	Comp17	Comp18	Comp19	Comp20	Comp21	Comp22	Comp23	Comp24
yerSTAND	0.0457	0.1070	-0.0842	0.1149	-0.0494	-0.1158	-0.1577	-0.1425	-0.2477	-0.4963	-0.3641	-0.4672
xtrSTAND	-0.2310	0.1986	-0.0814	0.1776	0.1242	0.1256	0.1476	0.0111	-0.5432	0.2017	0.5026	-0.0374
wurydSTAND	-0.1587	0.1780	-0.0935	-0.1397	0.1482	0.1802	0.0834	-0.0612	0.0145	0.1516	-0.1209	0.1287
vatSTAND	0.1040	0.1865	0.3339	0.0823	0.2924	0.1221	-0.0922	-0.0383	0.0706	0.0706	-0.0036	0.0014
urxSTAND	0.2586	-0.0068	0.1223	0.3196	-0.4370	-0.0390	0.2106	0.0048	-0.0192	-0.0731	0.0146	0.0008
ulaSTAND	0.0248	0.1437	-0.0747	0.6289	0.1615	-0.0007	-0.2156	0.1687	-0.0393	-0.0440	0.0604	-0.0563
totrevSTAND	-0.0202	0.2071	-0.3410	-0.1357	0.2325	-0.4614	0.0729	-0.2358	0.1076	-0.1429	0.1831	-0.0681
pyrSTAND	-0.0189	0.4645	-0.0514	0.1036	-0.3742	0.0567	0.1205	0.1608	0.1117	0.0832	0.0698	0.0568
proSTAND	0.2737	-0.0740	-0.1402	-0.0175	0.0881	-0.1152	-0.1620	0.2767	0.4741	0.1017	0.4299	-0.3171
pdnSTAND	-0.1820	-0.4197	-0.1918	0.0238	-0.1972	0.2455	-0.0135	0.0423	0.1021	0.1043	-0.0109	-0.0128
pcrSTAND	0.3250	-0.1921	-0.4207	-0.1009	0.0100	0.2274	0.0539	0.0582	-0.1757	0.2945	-0.2100	-0.1320
nfnSTAND	0.4491	0.0970	0.1959	-0.1598	0.0534	0.1414	-0.1593	-0.0285	-0.0644	-0.0017	-0.0500	-0.0014
nfaSTAND	0.2537	-0.0631	0.0299	0.0021	0.3467	-0.0974	0.4531	-0.1507	0.0397	-0.0648	0.0827	0.0024
mtrSTAND	-0.1895	0.1364	-0.0517	0.3355	0.1228	0.2799	0.1301	-0.3799	0.5204	0.0156	-0.2138	0.0904
itrSTAND	-0.3063	-0.1073	0.4612	-0.0550	-0.1052	-0.4317	0.1815	0.0646	0.0131	0.2809	-0.1332	-0.2290
innSTAND	-0.1392	-0.0016	0.2157	-0.1554	0.3602	0.1362	-0.3323	0.0752	-0.0571	0.0350	-0.0791	0.0313
hicSTAND	0.0906	0.3729	-0.0087	-0.3551	-0.1503	0.0439	0.2553	0.0381	0.0853	0.0048	-0.0470	0.0055
gynSTAND	0.3753	0.0016	0.3104	0.0764	-0.0863	0.0978	-0.0280	-0.4457	-0.1137	0.2257	0.1349	-0.0491
gdnSTAND	0.0337	-0.3017	0.0027	0.1200	0.2077	0.0536	0.4838	-0.0420	-0.0776	-0.1334	-0.0086	-0.0486
gbSTAND	0.0945	0.1174	0.1189	0.0197	0.2011	0.1583	0.2975	0.5998	-0.0221	-0.2248	-0.1859	0.0644
ddrSTAND	-0.0333	-0.1663	0.1305	-0.1472	-0.1390	0.1380	-0.0631	-0.0788	0.0156	-0.5615	0.3769	0.3684
cpnSTAND	0.0804	-0.2123	0.1851	0.1779	0.0527	-0.1555	-0.0000	0.1780	0.1365	0.0349	0.0279	0.0935
cacSTAND	0.1988	-0.0132	-0.1733	0.1416	0.0440	-0.4303	-0.1117	0.0256	-0.1271	0.1483	-0.2267	0.6470
mpcSTAND	0.0726	0.1568	0.0128	-0.0761	-0.0126	0.0073	-0.0445	0.0700	0.0682	0.0108	-0.0236	0.0307

Regression table for the production sector:

prod_logtr~s						
dyer	-.0009201	.000493	-1.87	0.067	-.001906	.0000658
dpcr	.0005267	.0009872	0.53	0.596	-.0014474	.0025007
dpcrgdp	-41.75814	45.25624	-0.92	0.360	-132.2536	48.7373
durx	-.1195582	.0807538	-1.48	0.144	-.2810355	.0419191
dpro	2.105256	.4835975	4.35	0.000	1.138243	3.072269
dpyr	.0000922	.0000733	1.26	0.213	-.0000543	.0002387
ditr	.0004158	.0002438	1.71	0.093	-.0000716	.0009033
dier	.0014179	.0016073	0.88	0.381	-.0017961	.0046319
diergdp	-97.0493	77.93605	-1.25	0.218	-252.8921	58.79346
dula	-2.307385	4.40317	-0.52	0.602	-11.11207	6.497298
dxtr	-.0000125	.0001831	-0.07	0.946	-.0003786	.0003536
dxtrgdp	5.262995	11.90352	0.44	0.660	-18.53956	29.06555
dsti	-.0948446	.3902035	-0.24	0.809	-.8751047	.6854155
dlti	-.0090889	.3902285	-0.02	0.981	-.789399	.7712211
dstireal	.0304191	.3898562	0.08	0.938	-.7491465	.8099848
dltireal	-.0376661	.389512	-0.10	0.923	-.8165435	.7412114
dipexe	.0094548	.0127491	0.74	0.461	-.0160385	.0349482
doil	.001211	.0038997	0.31	0.757	-.0065869	.0090089
dhic	.0202059	.0456846	0.44	0.660	-.0711461	.1115578
dfactor1	-.0699686	.0236518	-2.96	0.004	-.1172633	-.0226739
dfactor2	.0128865	.0135686	0.95	0.346	-.0142455	.0400185
dfactor3	.0308486	.0203174	1.52	0.134	-.0097785	.0714757
dfactor4	.0122565	.0328825	0.37	0.711	-.0534961	.0780092
dfactor5	-.0410859	.0183897	-2.23	0.029	-.0778584	-.0043133
_cons	-5.532787	.1551409	-35.66	0.000	-5.84301	-5.222563

Regression table for the service sector:

srv_logtrans						
dyer	-.0009222	.0005202	-1.77	0.081	-.0019624	.0001181
dpcr	.0002592	.0010417	0.25	0.804	-.0018237	.0023421
dpcrgdp	-26.92389	47.75268	-0.56	0.575	-122.4113	68.56351
durx	-.184155	.0852084	-2.16	0.035	-.3545397	-.0137702
dpro	2.485435	.510274	4.87	0.000	1.465079	3.50579
dpyr	.0001269	.0000773	1.64	0.106	-.0000277	.0002814
ditr	.0002366	.0002572	0.92	0.361	-.0002777	.000751
dier	.0020985	.001696	1.24	0.221	-.0012928	.0054898
diergdp	-93.91623	82.2352	-1.14	0.258	-258.3557	70.52321
dula	-1.792936	4.64606	-0.39	0.701	-11.08331	7.497435
dxtr	.0001885	.0001932	0.98	0.333	-.0001978	.0005748
dxtrgdp	-7.127521	12.56015	-0.57	0.572	-32.24309	17.98805
dsti	-.1153512	.4117281	-0.28	0.780	-.9386524	.70795
dlti	-.1111809	.4117545	-0.27	0.788	-.9345349	.712173
dstireal	.0178897	.4113616	0.04	0.965	-.8046787	.8404581
dltireal	.0452422	.4109985	0.11	0.913	-.7766	.8670845
dipexe	.0165122	.0134523	1.23	0.224	-.0103874	.0434118
doil	.0005457	.0041148	0.13	0.895	-.0076823	.0087738
dhic	.1791365	.0482046	3.72	0.000	.0827454	.2755276
dfactor1	-.0752793	.0249565	-3.02	0.004	-.1251829	-.0253757
dfactor2	.0073968	.014317	0.52	0.607	-.0212318	.0360255
dfactor3	.0405952	.0214381	1.89	0.063	-.0022729	.0834634
dfactor4	-.0249087	.0346964	-0.72	0.476	-.0942884	.044471
dfactor5	-.0236836	.0194042	-1.22	0.227	-.0624846	.0151174
_cons	-6.042504	.1636989	-36.91	0.000	-6.36984	-5.715168

Regression table for the construction sector:

cstr_logtr~s						
dyer	-.0014869	.0005862	-2.54	0.014	-.002659	-.0003148
dpcr	.0017823	.0011737	1.52	0.134	-.0005645	.0041292
dpcrgdp	-103.3455	53.80337	-1.92	0.059	-210.932	4.240965
durx	-.2475023	.0960051	-2.58	0.012	-.4394763	-.0555283
dpro	-.2646853	.5749302	-0.46	0.647	-1.414329	.8849586
dpyr	-.0001428	.0000871	-1.64	0.106	-.000317	.0000314
ditr	-.0001441	.0002898	-0.50	0.621	-.0007236	.0004355
dier	.000416	.0019109	0.22	0.828	-.003405	.004237
diergdp	-18.84311	92.65512	-0.20	0.840	-204.1185	166.4323
dula	-9.313071	5.234757	-1.78	0.080	-19.78061	1.154472
dxtr	.0009048	.0002176	4.16	0.000	.0004695	.00134
dxtrgdp	-44.42602	14.15163	-3.14	0.003	-72.72396	-16.12809
dsti	.014947	.4638977	0.03	0.974	-.9126737	.9425677
dlti	-.0804544	.4639274	-0.17	0.863	-1.008135	.8472257
dstireal	-.1543191	.4634848	-0.33	0.740	-1.081114	.772476
dltireal	.1153127	.4630756	0.25	0.804	-.8106642	1.04129
dipexe	.0056028	.0151569	0.37	0.713	-.0247052	.0359109
doil	.0003468	.0046362	0.07	0.941	-.0089239	.0096174
dhic	-.0519943	.0543126	-0.96	0.342	-.160599	.0566104
dfactor1	-.0239532	.0281187	-0.85	0.398	-.08018	.0322736
dfactor2	.0182328	.0161311	1.13	0.263	-.0140234	.050489
dfactor3	.0661315	.0241545	2.74	0.008	.0178315	.1144314
dfactor4	.0408523	.0390927	1.05	0.300	-.0373184	.119023
dfactor5	.0029548	.0218628	0.14	0.893	-.0407626	.0466723
_cons	-4.593109	.184441	-24.90	0.000	-4.961922	-4.224296

Regression table for other:

oth_logtrans						
dyer	-.0008531	.0003722	-2.29	0.025	-.0015973	-.0001089
dpcr	.0005803	.0007452	0.78	0.439	-.0009098	.0020705
dpcrgdp	-40.38454	34.16197	-1.18	0.242	-108.6956	27.92654
durx	-.1807732	.0609576	-2.97	0.004	-.3026654	-.058881
dpro	1.901183	.3650468	5.21	0.000	1.171227	2.631139
dpyr	.0000199	.0000553	0.36	0.721	-.0000907	.0001304
ditr	.0001766	.000184	0.96	0.341	-.0001913	.0005446
dier	.0012501	.0012133	1.03	0.307	-.001176	.0036762
diergdp	-51.12947	58.83055	-0.87	0.388	-168.7684	66.50948
dula	-2.499457	3.323762	-0.75	0.455	-9.14573	4.146816
dxtr	.0001829	.0001382	1.32	0.191	-.0000934	.0004593
dxtrgdp	-7.965545	8.985451	-0.89	0.379	-25.93307	10.00198
dsti	-.1190178	.2945477	-0.40	0.688	-.7080023	.4699667
dlti	-.0309678	.2945666	-0.11	0.917	-.61999	.5580545
dstireal	.0107593	.2942855	0.04	0.971	-.577701	.5992196
dltireal	.0122127	.2940257	0.04	0.967	-.5757281	.6001535
dipexe	.0147959	.0096237	1.54	0.129	-.0044479	.0340398
doil	.0011827	.0029437	0.40	0.689	-.0047036	.007069
dhic	.0817216	.0344853	2.37	0.021	.012764	.1506792
dfactor1	-.0648637	.0178537	-3.63	0.001	-.1005644	-.0291629
dfactor2	.0028295	.0102423	0.28	0.783	-.0176513	.0233103
dfactor3	.0322763	.0153367	2.10	0.039	.0016087	.0629439
dfactor4	.0033407	.0248216	0.13	0.893	-.0462931	.0529745
dfactor5	-.0202483	.0138816	-1.46	0.150	-.0480063	.0075097
_cons	-5.524412	.1171092	-47.17	0.000	-5.758586	-5.290238

Regression table for trading sector:

trad_logtr~s						
dyer	-.0001171	.0004648	-0.25	0.802	-.0010466	.0008123
dpcr	-.0004554	.0009307	-0.49	0.626	-.0023164	.0014056
dpcrgdp	16.01368	42.66375	0.38	0.709	-69.29776	101.3251
durx	-.1582224	.0761279	-2.08	0.042	-.3104495	-.0059952
dpro	2.538579	.4558948	5.57	0.000	1.626961	3.450197
dpyr	.0000293	.0000691	0.42	0.673	-.0001089	.0001674
ditr	.0003087	.0002298	1.34	0.184	-.0001509	.0007682
dier	.0009843	.0015152	0.65	0.518	-.0020456	.0040141
diergdp	-23.98241	73.47152	-0.33	0.745	-170.8978	122.933
dula	.4457004	4.150936	0.11	0.915	-7.854609	8.74601
dxtr	-.0001299	.0001726	-0.75	0.455	-.000475	.0002152
dxtrgdp	6.928074	11.22163	0.62	0.539	-15.51097	29.36711
dsti	-.1994632	.3678508	-0.54	0.590	-.9350265	.5361
dlti	.0240583	.3678744	0.07	0.948	-.711552	.7596686
dstireal	.0961024	.3675234	0.26	0.795	-.6388061	.831011
dltireal	-.0310316	.367199	-0.08	0.933	-.7652913	.7032282
dipexe	.0105666	.0120187	0.88	0.383	-.0134664	.0345996
doil	.000734	.0036763	0.20	0.842	-.0066172	.0080852
dhic	.1550849	.0430675	3.60	0.001	.0689661	.2412038
dfactor1	-.0784635	.0222969	-3.52	0.001	-.123049	-.0338781
dfactor2	-.0034218	.0127913	-0.27	0.790	-.0289996	.0221559
dfactor3	.017927	.0191535	0.94	0.353	-.0203727	.0562268
dfactor4	-.0155912	.0309988	-0.50	0.617	-.0775772	.0463949
dfactor5	-.0147933	.0173363	-0.85	0.397	-.0494594	.0198727
_cons	-5.910945	.1462538	-40.42	0.000	-6.203397	-5.618492

Regression table for transport sector:

trsp_logtr~s						
dyer	-.0012549	.0005423	-2.31	0.024	-.0023394	-.0001704
dpcr	.0017377	.0010859	1.60	0.115	-.0004337	.0039092
dpcrgdp	-92.29176	49.78206	-1.85	0.069	-191.8371	7.25361
durx	-.2795028	.0888296	-3.15	0.003	-.4571286	-.1018771
dpro	1.164778	.5319594	2.19	0.032	.1010599	2.228497
dpyr	-.0000871	.0000806	-1.08	0.284	-.0002483	.000074
ditr	-.0003677	.0002682	-1.37	0.175	-.0009039	.0001685
dier	.0001581	.001768	0.09	0.929	-.0033773	.0036935
diergdp	59.04271	85.72999	0.69	0.494	-112.385	230.4704
dula	3.962038	4.843506	0.82	0.417	-5.723152	13.64723
dxtr	.0005843	.0002014	2.90	0.005	.0001816	.000987
dxtrgdp	-27.18503	13.09392	-2.08	0.042	-53.36795	-1.002113
dsti	.2863799	.4292255	0.67	0.507	-.5719096	1.144669
dlti	-.5098536	.429253	-1.19	0.240	-1.368198	.3484909
dstireal	-.4621202	.4288435	-1.08	0.285	-1.319646	.3954054
dltireal	.5082171	.4284649	1.19	0.240	-.3485514	1.364986
dipexe	.0075256	.014024	0.54	0.593	-.0205172	.0355684
doil	.0015015	.0042897	0.35	0.728	-.0070762	.0100793
dhic	.19844	.0502532	3.95	0.000	.0979525	.2989275
dfactor1	-.0366558	.0260171	-1.41	0.164	-.0886802	.0153686
dfactor2	.0001485	.0149255	0.01	0.992	-.0296968	.0299938
dfactor3	.0618969	.0223492	2.77	0.007	.0172069	.1065869
dfactor4	.0194766	.0361709	0.54	0.592	-.0528516	.0918048
dfactor5	.0204178	.0202288	1.01	0.317	-.0200321	.0608678
_cons	-5.289894	.1706557	-31.00	0.000	-5.631141	-4.948647

Regression table for tourism sector:

tour_logtr~s						
dyer	-.0012511	.0005081	-2.46	0.017	-.0022671	-.000235
dpcr	.00104	.0010174	1.02	0.311	-.0009944	.0030744
dpcrgdp	-70.23119	46.6403	-1.51	0.137	-163.4942	23.03185
durx	-.2182199	.0832235	-2.62	0.011	-.3846356	-.0518042
dpro	2.752687	.4983873	5.52	0.000	1.7561	3.749274
dpyr	-5.99e-07	.0000755	-0.01	0.994	-.0001516	.0001504
ditr	.0002925	.0002512	1.16	0.249	-.0002099	.0007949
dier	.0013678	.0016565	0.83	0.412	-.0019445	.0046801
diergdp	-59.86971	80.31955	-0.75	0.459	-220.4786	100.7392
dula	-4.258627	4.537831	-0.94	0.352	-13.33258	4.815327
dxtr	.0001465	.0001887	0.78	0.441	-.0002308	.0005238
dxtrgdp	-9.271832	12.26756	-0.76	0.453	-33.80234	15.25868
dsti	-.213719	.402137	-0.53	0.597	-1.017842	.5904036
dlti	.0875129	.4021628	0.22	0.828	-.7166612	.891687
dstireal	.0811906	.4017791	0.20	0.841	-.7222163	.8845975
dltireal	-.0790314	.4014244	-0.20	0.845	-.8817291	.7236662
dipexe	.0345029	.013139	2.63	0.011	.0082299	.0607759
doil	.0031297	.004019	0.78	0.439	-.0049067	.0111661
dhic	.0251018	.0470817	0.53	0.596	-.0690439	.1192475
dfactor1	-.0963286	.0243751	-3.95	0.000	-.1450697	-.0475875
dfactor2	-.0009161	.0139835	-0.07	0.948	-.0288779	.0270456
dfactor3	.0259718	.0209387	1.24	0.220	-.0158978	.0678413
dfactor4	.0204791	.0338881	0.60	0.548	-.0472845	.0882426
dfactor5	-.047555	.0189521	-2.51	0.015	-.0854521	-.0096578
_cons	-5.037381	.1598856	-31.51	0.000	-5.357092	-4.71767

Appendix II

Abstract in German

In dieser Diplomarbeit geht es um die Messung des Einflusses von makroökonomischen Variablen auf Kreditausfallsraten, wie zum Beispiel das BIP oder der Preisindex für die Lebenserhaltung, in neun Industriesektoren.
Die Software, die hier zum Rechnen der Modelle verwendet wird, ist Stata. Die Diplomarbeit enthält auch die Softwarebefehle, die für das Programmieren des Modells notwendig sind.

Die statistischen Abläufe, die hier verwendet werden, sind bereits in einem Artikel von Boss et al. von der österreichischen Nationalbank veröffentlicht worden. Diese Abläufe sollen Informationen aus großen Datenmengen filtern und die Unterschiede zwischen dem Zeithorizont des Konjunkturzyklus' und des Kreditzyklus' überbrücken. Es wird nur das erste der zwei Modelle hier implementiert.
Dabei werden 24 makroökonomische Variablen in einer Hauptfaktorenanalyse in ein kleineres Set von fünf linearen Kombinationen transformiert, das ca. 71% der Varianz des Modells erklärt. Diese fünf Faktoren werden im Anschluss in die Regressionsanalyse miteinbezogen und auf ihre statistische und ökonomische Signifikanz untersucht.

Die Ergebnisse zeigen, dass entweder Hauptfaktor eins, drei oder fünf in jeden Sektor statistisch signifikant sind, mindestens einer der drei, meistens zwei davon. Sie sind jedoch in keinem Sektor ökonomisch signifikant.
Das BIP ist in mehreren Sektoren statistisch signifikant und eine der wenigen Variablen, die auch häufig ökonomisch signifikant sind.
Der Preisindex für die Lebenserhaltung ist statistisch signifikant in vielen Sektoren, jedoch ökonomisch insignifikant. Die Arbeitsproduktivität ist, wie das BIP, eine der wenigen Variablen, die sowohl statistisch als auch ökonomisch signifikant in der Mehrzahl der Sektoren sind.

Die statistischen Ausreißer wie zum Beispiel der Zinssatz für kurzzeitige Anleihen, Hauptfaktor fünf oder die Industrieproduktivität sind nur in einzelnen Sektoren statistisch signifikant und nicht ökonomisch signifikant, außer im Dienstleistungssektor.

Curriculum Vitae

Personal Profile

Date of Birth	born on August 1st 1986 in Salzburg
Nationality	Austria

Internships

03/2009 – 04/2009 **Austrian Federal Economic Chamber, Tokyo**

Internship at the Austrian Embassy Commercial Section

- Market research (Japan and Austria)
- Writing of industry reports
- Preparation of graphs, data and presentations
- Support of Austrian corporations at Japanese events (e.g. Austria Fair at Isetan Shinjuku)

09/2007 **Salzburger Sparkasse Bank AG (part of the Erste Bank group), Salzburg**

Finance internship

- Assistance of account managers (corporate and non-corporate clients)
- Back office
- Assistance in upcoming day-to-day activities

Education

03/2010 – 09/2010 Diploma thesis financed by and in cooperation with ÖVAG, Vienna: *"Modlling through the Austrian Business Cycle: An Update of the OeNB Model"*, grade: 1

10/2004 – now **International Business Administration**, University of Vienna

Magister (Master) course, degree expected in spring 2010

1st field of specialization: Corporate Finance

2nd field of specialization: International Management

01/2008 – 05/2008 Semester abroad at the London City University CASS business school

10/2005 – now **Japanese Studies**, University of Vienna Bachelor course

05/2004	General qualification for university entrance

Additional skills

Computer literacy	Stata Statistical Software, SPSS, MS-Office (good)	
Languages	German	native
	English	fluent
	Japanese	JLPT 3-2 level
	French	basic
Driving license (category B)		

References Available on request

Table of figures

Figure 1: Ordinary least squares, taken from Draper, N. & Smith H. (1981). *Applied Regression Analysis*. New York: Wiley, 10.

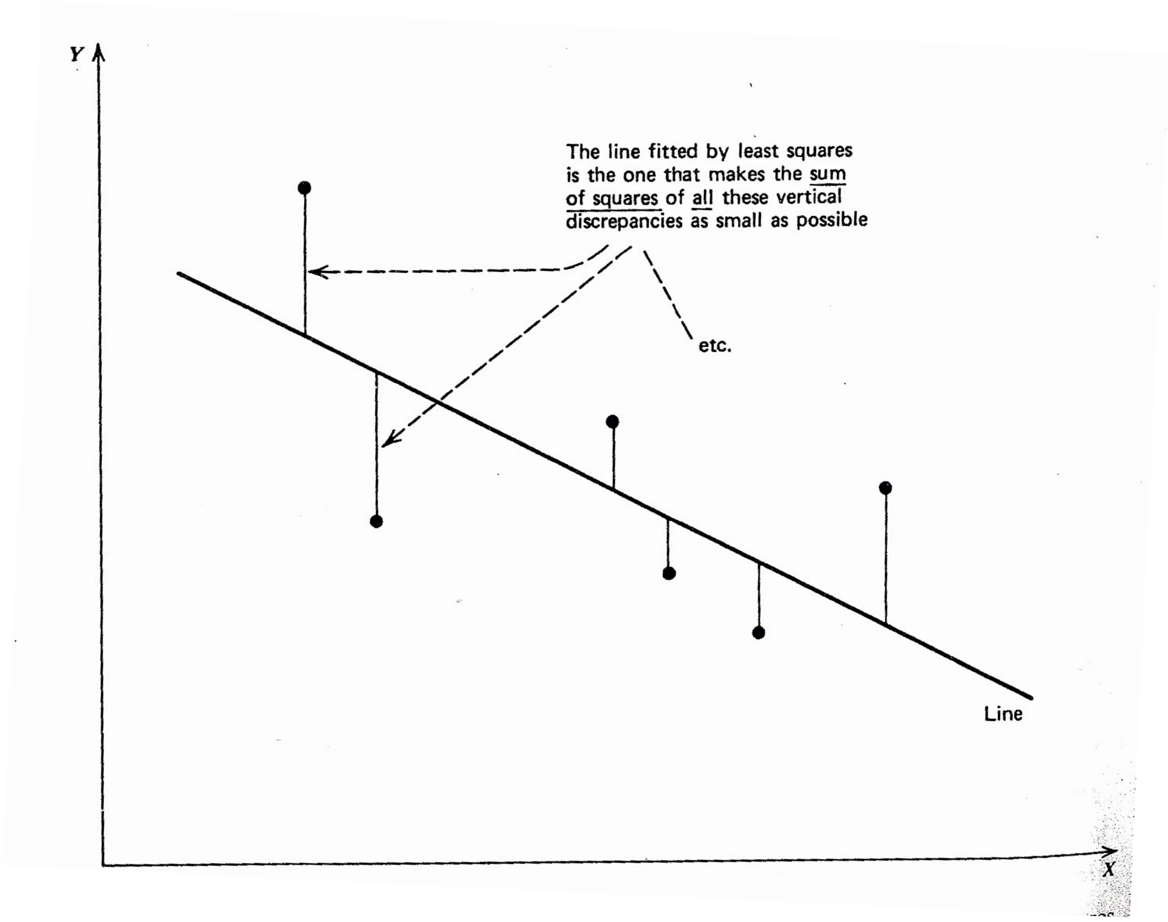

Figure 2: Stata result window

```
.
. generate time = q(1985q1) + _n -1

. format time %tq

.
. tsset time
        time variable:  time, 1985q1 to 2009q2
                delta:  1 quarter
```

Figure 3: Creating the deltas

```
. gen dcac = cac[_n] - cac[_n-4]
(4 missing values generated)

. lab var dcac "Delta CAC"

. gen dcpn = cpn[_n] - cpn[_n-4]
(4 missing values generated)

. lab var dcpn "Delta CPN"
```

Figure 4: Drop observations with missing values

```
.
. drop if dcac == .
(4 observations deleted)
```

Figure 5: Egen command for calculating the mean

```
.
. egen cacMEAN = mean(dcac)

. egen cpnMEAN = mean(dcpn)
```

Figure 6: Egen command for calculating the standard deviation

```
. egen cacSD = sd(dcac)

. egen cpnSD = sd(dcpn)
```

Figure 7: Generating the standardized variables

```
. generate cacSTAND = (dcac - cacMEAN) / cacSD

. generate cpnSTAND = (dcpn - cpnMEAN) / cpnSD
```

Figure 8: PCA

```
pca yerSTAND xtrSTAND wurydSTAND vatSTAND urxSTAND ulaSTAND totrevSTAND
pyrSTAND proSTAND pdnSTAND pcrSTAND nfnSTAND nfaSTAND mtrSTAND itrSTAND
innSTAND hicSTAND gynSTAND gdnSTAND gbSTAND ddrSTAND cpnSTAND cacSTAND
mpcSTAND
```

Figure 9: Predict PCA factors in Stata

```
. predict factor1 factor2 factor3 factor4 factor5, score
(19 components skipped)

Scoring coefficients
    sum of squares(column-loading) = 1
```

Figure 10: Graph of eigenvalues

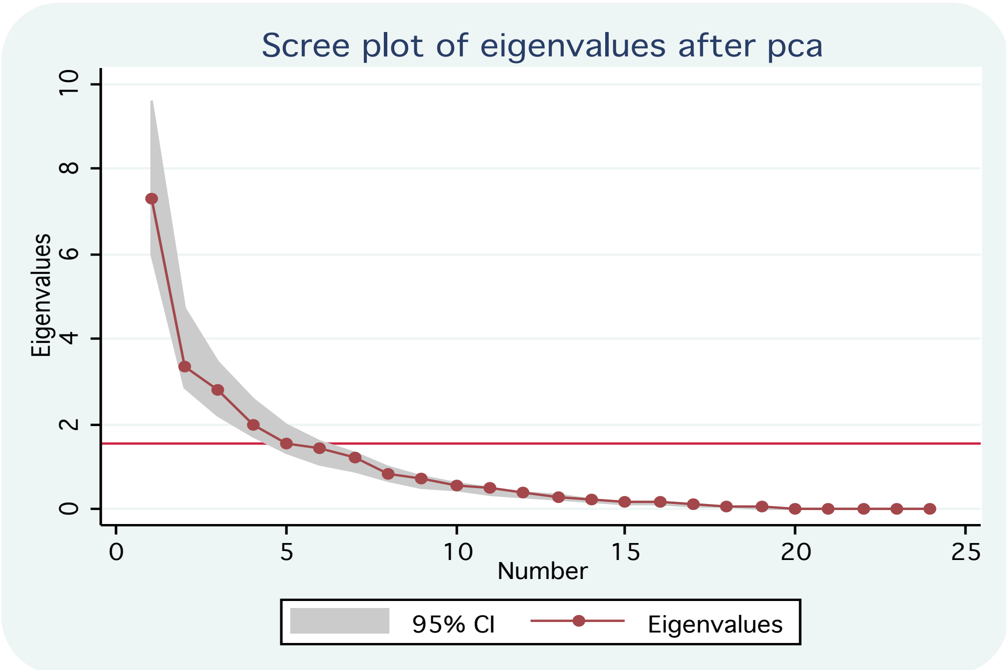

Figure 11: How to plot the PCA eigenvalue graph

```
. screeplot, yline(1) ci(het)

.
. notes factor1 : first component predicted by PCA

. notes factor2 : second component predicted by PCA

. notes factor3 : third component predicted by PCA

. notes factor4 : fourth component predicted by PCA

. notes factor5 : fifth component predicted by PCA
```

Figure 12 and 13: Logit transformation

```
.
. gen agr_t = ((1/PDagr)-1)

. gen prod_t = ((1/PDprod)-1)
```

```
. gen agr_logtrans = ln(1/agr_t)

. gen prod_logtrans = ln(1/prod_t)
```

Figure 14: Creating Δy

```
.
. gen dfactor1 = factor1[_n] - factor1[_n-4]
(4 missing values generated)

. lab var dfactor1 "Delta Factor 1"
```

Figure 15: Drop missing values

```
.
. drop if dfactor1 == .
(4 observations deleted)
```

Figure 16: Generate GDP variables

```
. gen pcrgdp = pcr/yer

. gen iergdp = ier/yer

. gen xtrgdp = xtr/yer
```

Figure 17: Conducting the regression

```
. mvreg agr_logtrans prod_logtrans srv_logtrans cstr_logtrans trad_logtrans tour_logtrans
trsp_logtrans oth_logtrans = dyer dpcr dpcrgdp durx
> dpro dpyr ditr dier diergdp dula dxtr dxtrgdp dsti dlti dstireal dltireal dipexe doil dhic
dfactor1 dfactor2 dfactor3 dfactor4 dfactor5
```

Figure 18: Stata command *ladder*

```
. ladder CAC

Transformation         formula              chi2(2)       P(chi2)
------------------------------------------------------------------
cubic                  CAC^3                 13.86         0.001
square                 CAC^2                 12.44         0.002
identity               CAC                   11.21         0.004
square root            sqrt(CAC)             10.68         0.005
log                    log(CAC)              10.19         0.006
1/(square root)        1/sqrt(CAC)            9.76         0.008
inverse                1/CAC                  9.39         0.009
1/square               1/(CAC^2)              8.79         0.012
1/cubic                1/(CAC^3)              8.41         0.015
```