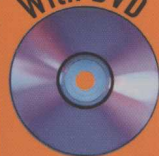


WILEY FINANCE

With DVD



Credit risk modeling using Excel and VBA

Second Edition

$$PD_i(Z) = \Phi \left(\frac{\Phi^{-1}(PD_i) - w_i Z}{\sqrt{1 - w_i^2}} \right)$$

$\lambda_{ij} =$ GUNTER LÖFFLER
PETER N. POSCH

$$\int Y(s) ds$$

Credit Risk Modeling Using Excel and VBA with DVD

Gunter Löffler
Peter N. Posch



A John Wiley and Sons, Ltd., Publication

This edition first published 2011
© 2011 John Wiley & Sons, Ltd

Registered office

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

ISBN 978-0-470-66092-8

A catalogue record for this book is available from the British Library.

Typeset in 10/12pt Times by Aptara Inc., New Delhi, India
Printed in Great Britain by CPI Antony Rowe, Chippenham, Wiltshire

Preface to the 2nd Edition

It is common to blame the inadequacy of risk models for the fact that the 2007–2008 financial crisis caught many market participants by surprise. On closer inspection, though, it often appears that it was not the models that failed. A good example is the risk contained in structured finance securities such as collateralized debt obligations (CDOs). In the first edition of this book, which was published before the crisis, we already pointed out that the rating of such products is not meant to communicate their systematic risk even though this risk component can be extremely large. This is easy to illustrate with simple, standard credit risk models, and surely we were not the first to point this out. Hence, in terms of risk, an AAA-rated bond is definitely not the same as an AAA-rated CDO. Many institutions, however, appear to have built their investment strategy on the presumption that AAA is AAA regardless of the product.

Recent events therefore do not invalidate traditional credit risk modeling as described in the first edition of the book. A second edition is timely, however, because the first edition dealt relatively briefly with the pricing of instruments that featured prominently in the crisis (CDSs and CDOs). In addition to expanding the coverage of these instruments, we devote more time to modeling aspects that were of particular relevance in the financial crisis (e.g., estimation error). We also examine the usefulness and limitations of credit risk modeling through case studies. For example, we discuss the role of scoring models in the subprime market, or show that a structural default prediction model would have assigned relatively high default probabilities to Lehman Brothers in the months before its collapse. Furthermore, we added a new chapter in which we show how to predict borrower-specific loss given default.

For university teachers, we now offer a set of powerpoint slides as well as problem sets with solutions. The material can be accessed via our homepage www.loeffler-posch.com. The hybrid character of the book – introduction to credit risk modeling as well as cookbook for model implementation – makes it a good companion to a credit risk course, at both introductory or advanced levels.

We are very grateful to Roger Bowden, Michael Kunisch and Alina Maurer for their comments on new parts of the book. One of us (Peter) benefited from discussions with a lot of people in the credit market, among them Nick Atkinson, David Kupfer and Marion Schlicker. Georg Haas taught him everything a trader needs to know, and Josef Gruber provided him with valuable insights to the practice of risk management. Several readers of the first edition pointed out errors or potential for improvement. We would like to use this opportunity to

thank them again and to encourage readers of the second edition to send us their comments (email: comment@loeffler-posch.com). Finally, special thanks to our team at Wiley: Andrew Finch, Brian Burge and our editors Aimee Dibbens and Karen Weller.

At the time of writing it is June. The weather is fine. We are looking forward to devoting more time to our families again.

Preface to the 1st Edition

This book is an introduction to modern credit risk methodology as well as a cookbook for putting credit risk models to work. We hope that the two purposes go together well. From our own experience, analytical methods are best understood by implementing them.

Credit risk literature broadly falls into two separate camps: risk measurement and pricing. We belong to the risk measurement camp. Chapters on default probability estimation and credit portfolio risk dominate chapters on pricing and credit derivatives. Our coverage of risk measurement issues is also somewhat selective. We thought it better to be selective than to include more topics with less detail, hoping that the presented material serves as a good preparation for tackling other problems not covered in the book.

We have chosen Excel as our primary tool because it is a universal and very flexible tool that offers elegant solutions to many problems. Even Excel freaks may admit that it is not their first choice for some problems. But even then, it is nonetheless great for demonstrating how to put models to work, given that implementation strategies are mostly transferable to other programming environments. While we tried to provide efficient and general solutions, this was not our single overriding goal. With the dual purpose of our book in mind, we sometimes favored a solution that appeared more simple to grasp.

Readers surely benefit from some prior Excel literacy, e.g., knowing how to use a simple function such as `AVERAGE()`, being aware of the difference between `SUM(A1:A10)` `SUM($A1:$A10)` and so forth. For less experienced readers, there is an *Excel for beginners* video on the DVD, and an introduction to VBA in the Appendix; the other videos supplied on the DVD should also be very useful as they provide a step-by-step guide more detailed than the explanations in the main text.

We also assume that the reader is somehow familiar with concepts from elementary statistics (e.g., probability distributions) and financial economics (e.g., discounting, options). Nevertheless, we explain basic concepts when we think that at least some readers might benefit from it. For example, we include appendices on maximum likelihood estimation or regressions.

We are very grateful to colleagues, friends and students who gave feedback on the manuscript: Oliver Blümke, Jürgen Bohrmann, André Güttler, Florian Kramer, Michael Kunisch, Clemens Prestele, Peter Raupach, Daniel Smith (who also did the narration of the videos with great dedication) and Thomas Verchow. An anonymous reviewer also provided a lot of helpful comments. We thank Eva Nacca for formatting work and typing video text. Finally, we thank our editors Caitlin Cornish, Emily Pears and Vivienne Wickham.

Any errors and unintentional deviations from best practice remain our own responsibility. We welcome your comments and suggestions: just send an email to comment@loeffler-posch.com or visit our homepage at www.loeffler-posch.com.

We owe a lot to our families. Before struggling to find the right words to express our gratitude we rather stop and give our families what they missed most, our time.

Some Hints for Troubleshooting

We hope that you do not encounter problems when working with the spreadsheets, macros and functions developed in this book. If you do, you may want to consider the following possible reasons for trouble:

- We repeatedly use the Excel Solver. This may cause problems if the Solver Add-in is not activated in Excel and VBA. How this can be done is described in Appendix A2. Apparently, differences in Excel versions can also lead to situations in which a macro calling the Solver does not run even though the reference to the Solver is set.
- In Chapters 10 and 11, we use functions from the *AnalysisToolpak* Add-in. Again, this has to be activated. See Chapter 10 for details.
- Some Excel 2003 functions (e.g., BINOMDIST or CRITBINOM) have been changed relative to earlier Excel versions. We've tested our programs on Excel 2003 and Excel 2010. If you're using an older Excel version, these functions might return error values in some cases.
- All functions have been tested for the demonstrated purpose only. We have not strived to make them so general that they work for most purposes one can think of. For example:
 - some functions assume that the data is sorted in some way, or arranged in columns rather than in rows;
 - some functions assume that the argument is a range, not an array. See Appendix A1 for detailed instructions on troubleshooting this issue.

A comprehensive list of all functions (Excel's and user-defined) together with full syntax and a short description can be found in Appendix A5.

Contents

Preface to the 2nd edition	xi
Preface to the 1st edition	xiii
Some Hints for Troubleshooting	xv
1 Estimating Credit Scores with Logit	1
Linking scores, default probabilities and observed default behavior	1
Estimating logit coefficients in Excel	4
Computing statistics after model estimation	8
Interpreting regression statistics	10
Prediction and scenario analysis	12
Treating outliers in input variables	16
Choosing the functional relationship between the score and explanatory variables	20
Concluding remarks	25
Appendix	25
Logit and probit	25
Marginal effects	25
Notes and literature	26
2 The Structural Approach to Default Prediction and Valuation	27
Default and valuation in a structural model	27
Implementing the Merton model with a one-year horizon	30
The iterative approach	30
A solution using equity values and equity volatilities	35
Implementing the Merton model with a T -year horizon	39
Credit spreads	43
CreditGrades	44
Appendix	50
Notes and literature	52
Assumptions	52
Literature	53

3	Transition Matrices	55
	Cohort approach	56
	Multi-period transitions	61
	Hazard rate approach	63
	Obtaining a generator matrix from a given transition matrix	69
	Confidence intervals with the binomial distribution	71
	Bootstrapped confidence intervals for the hazard approach	74
	Notes and literature	78
	Appendix	78
	Matrix functions	78
4	Prediction of Default and Transition Rates	83
	Candidate variables for prediction	83
	Predicting investment-grade default rates with linear regression	85
	Predicting investment-grade default rates with Poisson regression	88
	Backtesting the prediction models	94
	Predicting transition matrices	99
	Adjusting transition matrices	100
	Representing transition matrices with a single parameter	101
	Shifting the transition matrix	103
	Backtesting the transition forecasts	108
	Scope of application	108
	Notes and literature	110
	Appendix	110
5	Prediction of Loss Given Default	115
	Candidate variables for prediction	115
	Instrument-related variables	116
	Firm-specific variables	117
	Macroeconomic variables	118
	Industry variables	118
	Creating a data set	119
	Regression analysis of LGD	120
	Backtesting predictions	123
	Notes and literature	126
	Appendix	126
6	Modeling and Estimating Default Correlations with the Asset Value Approach	131
	Default correlation, joint default probabilities and the asset value approach	131
	Calibrating the asset value approach to default experience: the method of moments	133
	Estimating asset correlation with maximum likelihood	136
	Exploring the reliability of estimators with a Monte Carlo study	144
	Concluding remarks	147
	Notes and literature	147

7 Measuring Credit Portfolio Risk with the Asset Value Approach	149
A default-mode model implemented in the spreadsheet	149
VBA implementation of a default-mode model	152
Importance sampling	156
Quasi Monte Carlo	160
Assessing Simulation Error	162
Exploiting portfolio structure in the VBA program	165
Dealing with parameter uncertainty	168
Extensions	170
First extension: Multi-factor model	170
Second extension: t -distributed asset values	171
Third extension: Random LGDs	173
Fourth extension: Other risk measures	175
Fifth extension: Multi-state modeling	177
Notes and literature	179
8 Validation of Rating Systems	181
Cumulative accuracy profile and accuracy ratios	182
Receiver operating characteristic (ROC)	185
Bootstrapping confidence intervals for the accuracy ratio	187
Interpreting caps and ROCs	190
Brier score	191
Testing the calibration of rating-specific default probabilities	192
Validation strategies	195
Testing for missing information	198
Notes and literature	201
9 Validation of Credit Portfolio Models	203
Testing distributions with the Berkowitz test	203
Example implementation of the Berkowitz test	206
Representing the loss distribution	207
Simulating the critical chi-square value	209
Testing modeling details: Berkowitz on subportfolios	211
Assessing power	214
Scope and limits of the test	216
Notes and literature	217
10 Credit Default Swaps and Risk-Neutral Default Probabilities	219
Describing the term structure of default: PDs cumulative, marginal and seen from today	220
From bond prices to risk-neutral default probabilities	221
Concepts and formulae	221
Implementation	225
Pricing a CDS	232
Refining the PD estimation	234

Market values for a CDS	237
Example	239
Estimating upfront CDS and the 'Big Bang' protocol	240
Pricing of a pro-rata basket	241
Forward CDS spreads	242
Example	243
Pricing of swaptions	243
Notes and literature	247
Appendix	247
Deriving the hazard rate for a CDS	247
11 Risk Analysis and Pricing of Structured Credit: CDOs and First-to-Default Swaps	249
Estimating CDO risk with Monte Carlo simulation	249
The large homogeneous portfolio (LHP) approximation	253
Systemic risk of CDO tranches	256
Default times for first-to-default swaps	259
CDO pricing in the LHP framework	263
Simulation-based CDO pricing	272
Notes and literature	281
Appendix	282
Closed-form solution for the LHP model	282
Cholesky decomposition	283
Estimating PD structure from a CDS	284
12 Basel II and Internal Ratings	285
Calculating capital requirements in the Internal Ratings-Based (IRB) approach	285
Assessing a given grading structure	288
Towards an optimal grading structure	294
Notes and literature	297
Appendix A1 Visual Basics for Applications (VBA)	299
Appendix A2 Solver	307
Appendix A3 Maximum Likelihood Estimation and Newton's Method	313
Appendix A4 Testing and Goodness of Fit	319
Appendix A5 User-defined Functions	325
Index	333

Estimating Credit Scores with Logit

Typically, several factors can affect a borrower's default probability. In the retail segment, one would consider salary, occupation and other characteristics of the loan applicant; when dealing with corporate clients, one would examine the firm's leverage, profitability or cash flows, to name but a few items. A scoring model specifies how to combine the different pieces of information in order to get an accurate assessment of default probability, thus serving to automate and standardize the evaluation of default risk within a financial institution.

In this chapter, we show how to specify a scoring model using a statistical technique called *logistic regression* or simply *logit*. Essentially, this amounts to coding information into a specific value (e.g., measuring leverage as debt/assets) and then finding the combination of factors that does the best job in explaining historical default behavior.

After clarifying the link between scores and default probability, we show how to estimate and interpret a logit model. We then discuss important issues that arise in practical applications, namely the treatment of outliers and the choice of functional relationship between variables and default.

An important step in building and running a successful scoring model is its validation. Since validation techniques are applied not just to scoring models but also to agency ratings and other measures of default risk, they are described separately in Chapter 8.

LINKING SCORES, DEFAULT PROBABILITIES AND OBSERVED DEFAULT BEHAVIOR

A score summarizes the information contained in factors that affect default probability. Standard scoring models take the most straightforward approach by linearly combining those factors. Let x denote the factors (their number is K) and b the weights (or coefficients) attached to them; we can represent the score that we get in scoring instance i as

$$\text{Score}_i = b_1 x_{i1} + b_2 x_{i2} + \dots + b_K x_{iK} \quad (1.1)$$

It is convenient to have a shortcut for this expression. Collecting the b s and the x s in column vectors \mathbf{b} and \mathbf{x} we can rewrite (1.1) to

$$\text{Score}_i = b_1 x_{i1} + b_2 x_{i2} + \dots + b_K x_{iK} = \mathbf{b}' \mathbf{x}_i, \quad \mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \dots \\ x_{iK} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_K \end{bmatrix} \quad (1.2)$$

If the model is to include a constant b_1 , we set $x_{i1} = 1$ for each i .

Assume for simplicity that we have already agreed on the choice of the factors \mathbf{x} – what is then left to determine is the weight vector \mathbf{b} . Usually, it is estimated based on observed default

Table 1.1 Factor values and default behavior

Scoring instance i	Firm	Year	Default indicator for year +1	Factor values from the end of year			
			y_i	x_{i1}	x_{i2}	...	x_{iK}
1	XAX	2001	0	0.12	0.35	...	0.14
2	YOX	2001	0	0.15	0.51	...	0.04
3	TUR	2001	0	-0.10	0.63	...	0.06
4	BOK	2001	1	0.16	0.21	...	0.12
...
912	XAX	2002	0	-0.01	0.02	...	0.09
913	YOX	2002	0	0.15	0.54	...	0.08
914	TUR	2002	1	0.08	0.64	...	0.04
...
N	VRA	2005	0	0.04	0.76	...	0.03

behavior.¹ Imagine that we have collected annual data on firms with factor values and default behavior. We show such a data set in Table 1.1.²

Note that the same firm can show up more than once if there is information on this firm for several years. Upon defaulting, firms often stay in default for several years; in such cases, we would not use the observations following the year in which default occurred. If a firm moves out of default, we would again include it in the data set.

The default information is stored in the variable y_i . It takes the value 1 if the firm defaulted in the year following the one for which we have collected the factor values, and zero otherwise. N denotes the overall number of observations.

The scoring model should predict a high default probability for those observations that defaulted and a low default probability for those that did not. In order to choose the appropriate weights \mathbf{b} , we first need to link scores to default probabilities. This can be done by representing default probabilities as a function F of scores:

$$\text{Prob}(\text{Default}_i) = \text{Prob}(y_i = 1) = F(\text{Score}_i) \tag{1.3}$$

Like default probabilities, the function F should be constrained to the interval from zero to one; it should also yield a default probability for each possible score. The requirements can be fulfilled by a cumulative probability distribution function, and a distribution often considered for this purpose is the logistic distribution. The logistic distribution function $\Lambda(z)$ is defined as $\Lambda(z) = \exp(z)/(1 + \exp(z))$. Applied to (1.3) we get

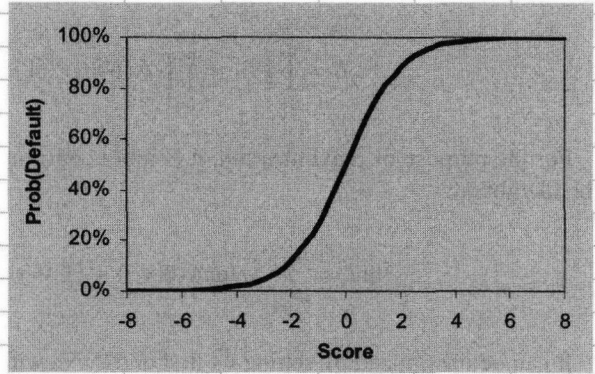
$$\text{Prob}(\text{Default}_i) = \Lambda(\text{Score}_i) = \frac{\exp(\mathbf{b}'\mathbf{x}_i)}{1 + \exp(\mathbf{b}'\mathbf{x}_i)} = \frac{1}{1 + \exp(-\mathbf{b}'\mathbf{x}_i)} \tag{1.4}$$

Models that link information to probabilities using the logistic distribution function are called *logit* models.

¹ In qualitative scoring models, however, experts determine the weights.
² Data used for scoring are usually on an annual basis, but one can also choose other frequencies for data collection as well as other horizons for the default horizon.

Table 1.2 Scores and default probabilities in the logit model

	A	B	C	D	E	F	G	H
1	Score	Prob(Default)						
2	-8	0.03%	$=1/(1+\text{EXP}(-A2))$					
3	-7	0.09%	(can be copied into B3:B18)					
4	-6	0.25%						
5	-5	0.67%						
6	-4	1.80%						
7	-3	4.74%						
8	-2	11.92%						
9	-1	26.89%						
10	0	50.00%						
11	1	73.11%						
12	2	88.08%						
13	3	95.26%						
14	4	98.20%						
15	5	99.33%						
16	6	99.75%						
17	7	99.91%						
18	8	99.97%						



In Table 1.2, we list the default probabilities associated with some score values and illustrate the relationship with a graph. As can be seen, higher scores correspond to a higher default probability. In many financial institutions, credit scores have the opposite property: they are higher for borrowers with a lower credit risk. In addition, they are often constrained to some set interval, e.g., zero to 100. Preferences for such characteristics can easily be met. If we use (1.4) to define a scoring system with scores from -9 to 1 , but want to work with scores from 0 to 100 instead (100 being the best), we could transform the original score to $myscore = -10 \times score + 10$.

Having collected the factors \mathbf{x} and chosen the distribution function F , a natural way of estimating the weights \mathbf{b} is the maximum likelihood (ML) method. According to the ML principle, the weights are chosen such that the probability (=likelihood) of observing the given default behavior is maximized (see Appendix A3 for further details on ML estimation).

The first step in maximum likelihood estimation is to set up the likelihood function. For a borrower that defaulted, the likelihood of observing this is

$$\text{Prob}(\text{Default}_i) = \text{Prob}(y_i = 1) = \Lambda(\mathbf{b}'\mathbf{x}_i) \quad (1.5)$$

For a borrower that did not default, we get the likelihood

$$\text{Prob}(\text{No default}_i) = \text{Prob}(y_i = 0) = 1 - \Lambda(\mathbf{b}'\mathbf{x}_i) \quad (1.6)$$

Using a little trick, we can combine the two formulae into one that automatically gives the correct likelihood, be it a defaulter or not. Since any number raised to the power of zero

evaluates to one, the likelihood for observation i can be written as

$$L_i = (\Lambda(\mathbf{b}'\mathbf{x}_i))^{y_i} (1 - \Lambda(\mathbf{b}'\mathbf{x}_i))^{1-y_i} \quad (1.7)$$

Assuming that defaults are independent, the likelihood of a set of observations is just the product of the individual likelihoods:³

$$L = \prod_{i=1}^N L_i = \prod_{i=1}^N (\Lambda(\mathbf{b}'\mathbf{x}_i))^{y_i} (1 - \Lambda(\mathbf{b}'\mathbf{x}_i))^{1-y_i} \quad (1.8)$$

For the purpose of maximization, it is more convenient to examine $\ln L$, the logarithm of the likelihood:

$$\ln L = \sum_{i=1}^N y_i \ln(\Lambda(\mathbf{b}'\mathbf{x}_i)) + (1 - y_i) \ln(1 - \Lambda(\mathbf{b}'\mathbf{x}_i)) \quad (1.9)$$

It can be maximized by setting its first derivative with respect to \mathbf{b} to zero. This derivative (like \mathbf{b} , it is a vector) is given by

$$\frac{\partial \ln L}{\partial \mathbf{b}} = \sum_{i=1}^N (y_i - \Lambda(\mathbf{b}'\mathbf{x}_i)) \mathbf{x}_i \quad (1.10)$$

Newton's method (see Appendix A3) does a very good job in solving equation (1.10) with respect to \mathbf{b} . To apply this method, we also need the second derivative, which we obtain as

$$\frac{\partial^2 \ln L}{\partial \mathbf{b} \partial \mathbf{b}'} = - \sum_{i=1}^N \Lambda(\mathbf{b}'\mathbf{x}_i) (1 - \Lambda(\mathbf{b}'\mathbf{x}_i)) \mathbf{x}_i \mathbf{x}_i' \quad (1.11)$$

ESTIMATING LOGIT COEFFICIENTS IN EXCEL

Excel does not contain a function for estimating logit models, and so we sketch how to construct a user-defined function that performs the task. Our complete function is called LOGIT. The syntax of the LOGIT command is equivalent to the LINEST command: LOGIT($y, x, [const], [statistics]$), where $[]$ denotes an optional argument.

The first argument specifies the range of the dependent variable, which in our case is the default indicator y ; the second parameter specifies the range of the explanatory variable(s). The third and fourth parameters are logical values for the inclusion of a constant (1 or omitted if a constant is included, 0 otherwise) and the calculation of regression statistics (1 if statistics are to be computed, 0 or omitted otherwise). The function returns an array, therefore, it has to be executed on a range of cells and entered by [ctrl]+[shift]+[enter].

³ Given that there are years in which default rates are high, and others in which they are low, one may wonder whether the independence assumption is appropriate. It will be if the factors that we input into the score capture fluctuations in average default risk. In many applications, this is a reasonable assumption.

Table 1.3 Application of the LOGIT command to a data set with information on defaults and five financial ratios

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Firm ID	Year	De- fault	WC/ TA	RE/ TA	EBIT/ TA	ME/ TL	S/ TA		CONST	WC/ TA	RE/ TA	EBIT/ TA	ME/ TL	S/ TA
2	1	1999	0	0.50	0.31	0.04	0.96	0.33	b	-2.543	0.414	-1.454	-7.999	-1.594	0.620
3	1	2000	0	0.55	0.32	0.05	1.06	0.33		{=LOGIT(C2:C4001,D2:H4001,1,0)}					
4	1	2001	0	0.45	0.23	0.03	0.80	0.25		(applies to J2:O2)					
5	1	2002	0	0.31	0.19	0.03	0.39	0.25							
6	1	2003	0	0.45	0.22	0.03	0.79	0.28							
7	1	2004	0	0.46	0.22	0.03	1.29	0.32							
8	2	1999	0	0.01	-0.03	0.01	0.11	0.25							
9	2	2000	0	-0.11	-0.12	0.03	0.15	0.32							
...															
108	21	1996	1	0.36	0.06	0.03	3.20	0.28							
...															
4001	830	2002	1	0.07	-0.11	0.04	0.04	0.12							

Before delving into the code, let us look at how the function works on an example data set.⁴ We have collected default information and five variables for default prediction; Working Capital (WC), Retained Earnings (RE), Earnings Before Interest and Taxes (EBIT) and Sales (S), each divided by Total Assets (TA); and Market Value of Equity (ME) divided by Total Liabilities (TL). Except for the market value, all these items are found in the balance sheet and income statement of the company. The market value is given by the number of shares outstanding multiplied by the stock price. The five ratios are the ones from the widely known Z-score developed by Altman (1968). WC/TA captures the short-term liquidity of a firm, RE/TA and EBIT/TA measure historic and current profitability, respectively. S/TA further proxies for the competitive situation of the company and ME/TL is a market-based measure of leverage.

Of course, one could consider other variables as well; to mention only a few, these could be: cash flows over debt service, sales or total assets (as a proxy for size), earnings volatility, stock price volatility. In addition, there are often several ways of capturing one underlying factor. Current profits, for instance, can be measured using EBIT, EBITDA (=EBIT plus depreciation and amortization) or net income.

In Table 1.3, the data is assembled in columns A to H. Firm ID and year are not required for estimation. The LOGIT function is applied to range J2:O2. The default variable that the LOGIT function uses is in the range C2:C4001, while the factors x are in the range D2:H4001. Note that (unlike in Excel's LINEST function) coefficients are returned in the same order as the variables are entered; the constant (if included) appears as the leftmost variable. To interpret the sign of the coefficient b , recall that a higher score corresponds to a higher default probability. The negative sign of the coefficient for EBIT/TA, for example, means that default probability goes down as profitability increases.

Now let us have a close look at important parts of the LOGIT code. In the first lines of the function, we analyze the input data to define the data dimensions: the total number of observations N and the number of explanatory variables (including the constant) K . If a

⁴ The data is hypothetical, but mirrors the structure of data for listed US corporates.