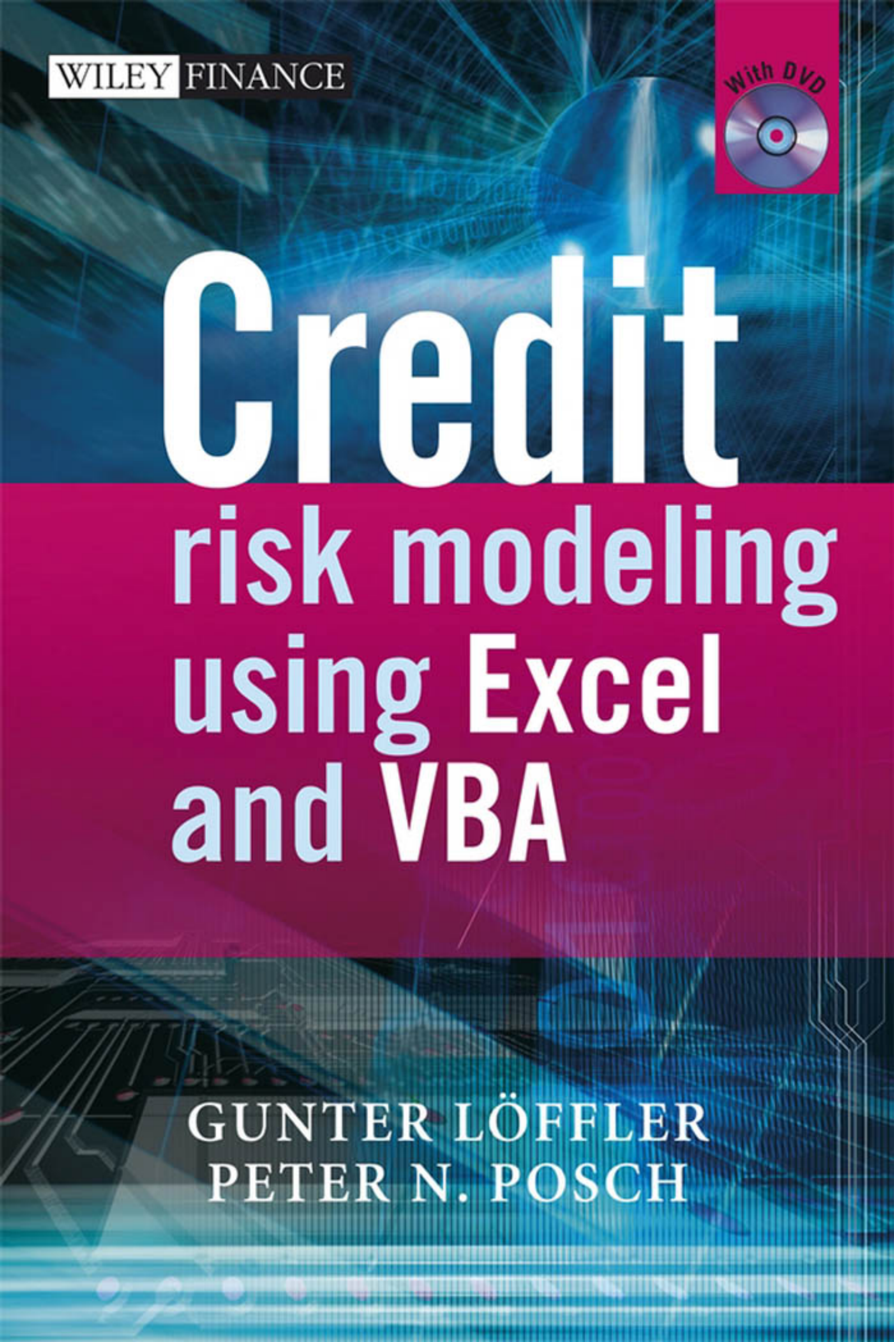
WILEY FINANCE



# Credit risk modeling using Excel and VBA

GUNTER LÖFFLER
PETER N. POSCH

# Credit risk modeling using Excel and VBA

**Gunter Löffler**
**Peter N. Posch**

John Wiley & Sons, Ltd

# Credit risk modeling
# using Excel and VBA

For other titles in the Wiley Finance series
please see www.wiley.com/finance

# Credit risk modeling using Excel and VBA

**Gunter Löffler**
**Peter N. Posch**

BICENTENNIAL 1807 WILEY 2007 BICENTENNIAL

John Wiley & Sons, Ltd

Copyright © 2007 John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

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, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770620.

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.

***Other Wiley Editorial Offices***

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, ONT, L5R 4J3, Canada

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

Anniversary Logo Design: Richard J. Pacifico

***Library of Congress Cataloging in Publication Data***

Löffler, Gunter.
Credit risk modeling using Excel and VBA / Gunter Löffler, Peter N. Posch.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-03157-5 (cloth : alk. paper)
1. Credit—Management 2. Risk Management 3. Microsoft Excel (Computer file)
4. Microsoft Visual Basic for applications. I. Posch, Peter N. II. Title.
HG3751.L64 2007
332.70285′554—dc22 2007002347

***British Library Cataloguing in Publication Data***

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

ISBN 978-0-470-03157-5 (HB)

Typeset in 10/12pt Times by Integra Software Services Pvt. Ltd, Pondicherry, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire
This book is printed on acid-free paper responsibly manufactured from sustainable forestry
in which at least two trees are planted for each one used for paper production.

Mundus est is qui constat ex caelo, et terra et mare cunctisque sideribus.
*Isidoro de Sevilla*

# Contents

**Preface** xi

**Some Hints for Troubleshooting** xiii

**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 13
Treating outliers in input variables 15
Choosing the functional relationship between the score and explanatory variables 19
Concluding remarks 23
Notes and literature 24
Appendix 24

**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 34
Implementing the Merton model with a $T$-year horizon 39
Credit spreads 44
Notes and literature 44

**3 Transition Matrices** 45
Cohort approach 46
Multi-period transitions 51
Hazard rate approach 53
Obtaining a generator matrix from a given transition matrix 58
Confidence intervals with the Binomial distribution 59
Bootstrapped confidence intervals for the hazard approach 63
Notes and literature 67
Appendix 67

4 **Prediction of Default and Transition Rates** **73**
Candidate variables for prediction 73
Predicting investment-grade default rates with linear regression 75
Predicting investment-grade default rates with Poisson regression 78
Backtesting the prediction models 83
Predicting transition matrices 87
Adjusting transition matrices 88
Representing transition matrices with a single parameter 89
Shifting the transition matrix 91
Backtesting the transition forecasts 96
Scope of application 98
Notes and literature 98
Appendix 99

5 **Modeling and Estimating Default Correlations with the Asset Value Approach** **103**
Default correlation, joint default probabilities and the asset value approach 103
Calibrating the asset value approach to default experience: the method of moments 105
Estimating asset correlation with maximum likelihood 108
Exploring the reliability of estimators with a Monte Carlo study 114
Concluding remarks 117
Notes and literature 117

6 **Measuring Credit Portfolio Risk with the Asset Value Approach** **119**
A default mode model implemented in the spreadsheet 119
VBA implementation of a default-mode model 122
Importance sampling 126
Quasi Monte Carlo 130
Assessing simulation error 132
Exploiting portfolio structure in the VBA program 135
Extensions 137
  First extension: Multi-factor model 137
  Second extension: *t*-distributed asset values 138
  Third extension: Random LGDs 139
  Fourth extension: Other risk measures 143
  Fifth extension: Multi-state modeling 144
Notes and literature 146

7 **Validation of Rating Systems** **147**
Cumulative accuracy profile and accuracy ratios 148
Receiver operating characteristic (ROC) 151
Bootstrapping confidence intervals for the accuracy ratio 153
Interpreting CAPs and ROCs 155
Brier Score 156
Testing the calibration of rating-specific default probabilities 157

Validation strategies 161
Notes and literature 162

**8 Validation of Credit Portfolio Models 163**
Testing distributions with the Berkowitz test 163
Example implementation of the Berkowitz test 166
Representing the loss distribution 167
Simulating the critical chi-squared value 169
Testing modeling details: Berkowitz on subportfolios 171
Assessing power 175
Scope and limits of the test 176
Notes and literature 177

**9 Risk-Neutral Default Probabilities and Credit Default Swaps 179**
Describing the term structure of default: PDs cumulative, marginal, and seen from today 180
From bond prices to risk-neutral default probabilities 181
Concepts and formulae 181
Implementation 184
Pricing a CDS 191
Refining the PD estimation 193
Notes and literature 196

**10 Risk Analysis of Structured Credit: CDOs and First-to-Default Swaps 197**
Estimating CDO risk with Monte Carlo simulation 197
The large homogeneous portfolio (LHP) approximation 201
Systematic risk of CDO tranches 203
Default times for first-to-default swaps 205
Notes and literature 209
Appendix 209

**11 Basel II and Internal Ratings 211**
Calculating capital requirements in the Internal Ratings-Based (IRB) approach 211
Assessing a given grading structure 214
Towards an optimal grading structure 220
Notes and literature 223

**Appendix A1 Visual Basics for Applications (VBA) 225**

**Appendix A2 Solver 233**

**Appendix A3 Maximum Likelihood Estimation and Newton's Method 239**

**Appendix A4 Testing and Goodness of Fit 245**

**Appendix A5 User-Defined Functions 251**

**Index 257**

# Preface

This book is an introduction to modern credit risk methodology as well 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 at 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 Chapter 10, we use functions from the *AnalysisToolpak* add-in. Again, this has to be activated. See Chapter 9 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. 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 the 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 at the end of Appendix A5.

# 1

# Estimating Credit Scores with Logit

Typically, several factors can affect a borrower's default probability. In the retail segment, one would consider salary, occupation, age 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. 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 will 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 7.

## 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:

$$\text{Score}_i = b_1 x_{i1} + b_2 x_{i2} + \ldots + b_K x_{iK} \tag{1.1}$$

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:

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

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 on the basis of the

**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 |

observed default 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. The overall number of observations is denoted by $N$.

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 **b**, we first need to link scores to default probabilities. This can be done by representing default probabilities as a function $F$ of scores:

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

Like default probabilities, the function $F$ should be constrained to the interval from 0 to 1; it should also yield a default probability for each possible score. The requirements can be fulfilled by a cumulative probability distribution function. 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.

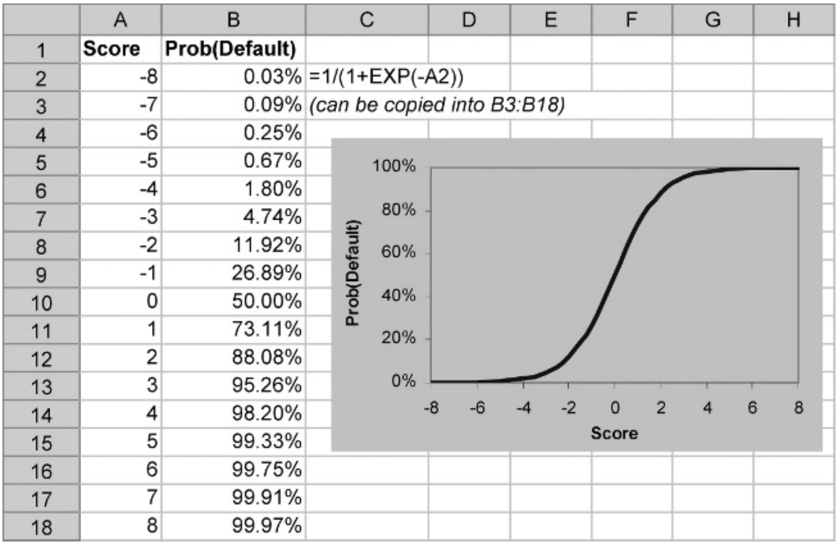
---

[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.

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. 0 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 $myscore = -10 \times score + 10$.

**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+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% | | | | | | |

Having collected the factors **x** and chosen the distribution function $F$, a natural way of estimating the weights **b** is the maximum likelihood method (ML). 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 ($Y_i = 1$), the likelihood of observing this is

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

For a borrower that did not default ($Y_i = 0$), we get the likelihood

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

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 0 evaluates to 1, the likelihood for observation $i$ can be written as:

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