Chapter 11
Exposure at Default (EAD) and Adverse Selection

INTRODUCTION

In this chapter, you learn what exposure at default (EAD) modeling is, the Basel requirements for EAD, and the various methods for building EAD models. Remember, just as with LGD, EAD also has a linear impact on both the expected loss and the Basel capital. It is thus of key importance to model EAD as accurately as possible.

Let us first start by defining the concept of EAD modeling. For on-balance-sheet exposures, such as term loans, installment loans, and mortgages, the EAD is defined as the nominal outstanding balance, net of specific provisions. In other words, it represents the net outstanding debt.

Examples of studies on EAD include Jacobs (2011), who develops linear regression models for conversion measures for a corporate revolving credit facility with a number of determinants such as credit rating, utilization, tenor, industry, and macro-economic factors. The significant finding is that utilization is the strongest factor influencing EAD. Barakova and Parthasarathy (2012) analyze syndicated corporate credit lines. The authors find that risk rating, line utilization, size, and sudden turns in the economic cycle have the most significant impact on EAD. Agarwal, Ambrose, and Liu (2006) use private bank data and analyze home equity lines of credit utilization. Tong et al. (2016) analyze EAD for credit card loans. We will describe this study in more detail later. Another example is Valvonis (2008), who discusses the estimation of exposure at default for regulatory capital.

Fixed versus Variable Outstanding

The EAD is deterministic for some loans (e.g., many corporate bonds) and variable for others. Within the loans with variable amount outstanding, two types may be distinguished: credit lines and loans with flexible payment schedules.

Credit lines (e.g., credit card loans) generally have a limit and a drawn amount. The borrower can draw on the line up to the limit. At the end of a period, the borrower is required to repay the drawn amount, after which it is set back to zero.

Loans with flexible payment schedules are often loans with a prepayment option and a redraw option. Amortizing loans (e.g., mortgage loans) generally require borrowers to make interest and principal payments prior to maturity, and the loan balance reduces with the principal amount that is repaid. A prepayment option allows the borrower to pay down more than the scheduled amount, whereas a redraw option allows the borrower to draw on prepayments and/or principal repayments.

It is common in some countries (e.g., Australia) to arrange for two connected accounts: a loan account (with a negative balance) and an offset account (with a positive balance). The outstanding loan amount is the net of the two balances, and borrowers can draw on the offset account at any time while principal repayments are credited to this account. This product in essence allows borrowers to redraw the prepayments. It is also common in other countries (e.g., United States) to arrange for a home equity line of credit next to a mortgage loan. The idea here is to redraw already-made principal payments on a separate account with the value of the house less the primary mortgage being used as collateral to determine the limit. Other home equity extraction contracts include reverse mortgages, which may also have redraw features.

Off-Balance-Sheet Exposures

As already mentioned, banks are exposed to off-balance-sheet exposures. Off-balance-sheet exposures include contingent credit exposures (e.g., credit guarantees and loan commitments) and counterparty credit risk in relation to over-the-counter (OTC), i.e., bilateral, derivative agreements on future deliveries. The exposure at default depends on the loss in relation to a guarantee or an adverse movement of an underlying measure in conjunction with the credit event of a counterparty in derivative contracts.

For off-balance-sheet exposures, you need to take into account what portion of the undrawn amount is likely to be converted into credit upon the default of a reference firm or person (in the case of a guarantee) or the counterparty (in the case of a derivative).

Conversion Measures

Exposures at default can be modeled directly by building a model for the exposure amount or a monotone transformation thereof. Alternatively, you may relate the EAD to a scaling variable and derive conversion measures. Yang and Tkachenko (2012) find that models for conversion measures are generally more robust than EAD models, as they scale all observations to a common denomination.

Exhibit 11.1 shows four conversion measures that are common. Other measures toconvert drawn amounts and limits, as well as other quantities, may also be considered. Examples of such quantities include sales, number of staff, profitability, liquidity, or age.

Conversion Measure	**Formula**
Credit Conversion Factor (CCF)	$EAD = Drawn + CCF*(Limit\text{-}Drawn)$
Credit Equivalent (CEQ)	$EAD = Drawn + CEQ*Limit$
Limit Conversion Factor (LCF)/Loan Equivalent (LEQ)	$EAD = LCF*Limit$
Used Amount Conversion Factor (UACF)	$EAD = UACF*Drawn$

Exhibit 11.1 Conversion of Limits and Drawn Amounts to EAD for Off-Balance-Sheet Exposures

The credit conversion factor (CCF) is defined as the portion of the undrawn amount that will be converted into credit. Note that the undrawn amount is equal to the limit minus the drawn amount. The EAD thus becomes the drawn amount plus the CCF times the limit minus the drawn amount. The credit equivalent (CEQ), is defined as the portion of the limit likely to be converted into credit. The EAD is then defined as the drawn amount plus the CEQ times the limit. The limit conversion factor (LCF), or loan equivalent (LEQ), is defined as a fraction of the limit representing the total exposure. The EAD is then defined as the LCF or LEQ times the limit. Finally, the used amount conversion factor (UACF) is defined using the drawn amount as the reference. Hence, the EAD is then computed as the UACF times the drawn amount.

Note that these measures need to be carefully chosen. For example, there is some controversy on the use of conversion measures as they may be too restrictive (compare Taplin, To, and Hee 2007) or volatile (compare Qi 2009).

REGULATORY PERSPECTIVE ON EAD

In the Basel Accord, generally the credit conversion factor (CCF) approach is used. The CCF ranges between zero and one, corresponding with an EAD equal to the drawn amount and limit, respectively. EAD modeling now comes down to estimating the CCF. Some financial institutions do not develop CCF models and use a conservative approach by consistently setting the CCF to one. In other words, they assume that the EAD will always be equal to the credit limit upon default of the exposure. In what follows, we will see what Basel says about EAD modeling. For corporates, sovereigns, and bank exposures, paragraph 310 of the Basel II Accord reads (see Basel Committee on Banking Supervision 2006):

> For off-balance sheet items, exposure is calculated as the committed but undrawn amount multiplied by a CCF. There are two approaches for the estimation of CCFs: a foundation approach and an advanced approach.

Remember, the IRB approach has two subapproaches, the foundation IRB approach and the advanced IRB approach. In the foundation IRB approach, banks can estimate the PD themselves, but rely on the Accord or local regulators for reference values of the LGD and EAD. In the advanced IRB approach, the bank can estimate all three risk parameters: the PD, LGD, and EAD. Also remember that the foundation IRB approach is not allowed for retail exposures. The reference values for EAD or thus CCF for corporate, sovereign, and bank exposures are provided for loan commitments in paragraph 83 of the Basel Committee on Banking Supervision 2006 as follows:

> Commitments with an original maturity up to one year and commitments with an original maturity over one year will receive a CCF of 20% and 50%, respectively. However, any commitments that are unconditionally cancelable at any time by the bank without prior notice will receive a 0% CCF.

For retail exposures, paragraph 334 of the Accord reads:

> Both on- and off-balance sheet retail exposures are measured gross of specific provisions or partial write-offs. The EAD on drawn amounts should not be less than the sum of (i) the amount by which a bank's regulatory capital would be reduced if the exposure were written-off fully, and (ii) any specific provisions and partial write-offs.

This paragraph is then followed by paragraph 335 as follows:

> For retail off-balance sheet items, banks must use their own estimates of CCFs.

Paragraph 336 then considers the issue of a consistent definition of both EAD and LGD as follows:

> For retail exposures with uncertain future drawdown such as credit cards, banks must take into account their history and/or expectation of additional drawings prior to default in their overall calibration of loss estimates. In particular, where a bank does not reflect conversion factors for undrawn lines in its EAD estimates, it must reflect in its LGD estimates the likelihood of additional drawings prior to default. Conversely, if the bank does not incorporate the possibility of additional drawings in its LGD estimates, it must do so in its EAD estimates.

To summarize, additional drawings prior to default can be included in either LGD or EAD. However, it is common practice to include those in the EAD definition by using credit conversion factors, as discussed earlier.

The EU and U.S. regulations also introduce the idea of a margin of conservatism and economic downturn EAD if it turns out that the EAD is volatile over the economic cycle. Note that this is actually very similar to LGD.

For OTC derivatives, the exposure is a combination of the current exposure and the potential future exposure. In the past, the EAD for OTC derivatives was the sum of the replacement cost (RC, also called current exposure) and the potential future exposure (PFE). PFE was determined by a multiplication of the notional amount and a look-up CCF (depending on the maturity and the underlying class). These rules will be updated in the standardized approach for measuring counterparty credit risk exposures (SA-CCR) (compare Basel Committee on Banking Supervision 2014). Under these rules, the exposure is calculated as follows:

$$EAD = 1.4(RC + PFE)$$

The replacement costs consider margining (i.e., provision of cash or collateral by the counterparty), and the potential future exposure considers a number of characteristics that describe the risk profile of the derivatives.

A practical problem with EAD models (regardless of whether they are modeling EAD, a transformation thereof, or a conversion measure) is that they could lead up to an estimate of the EAD that is lower than the current outstanding or drawn amount. Hence, a flooring operation is generally applied for determining EAD for regulatory purposes. EAD is always at least the drawn amount, or, in other words, the EAD equals the maximum of the drawn amount and the LCF times the limit. This was also confirmed by the PRA as follows:

> The PRA expects that EAD estimates should not be less than current drawings.

In terms of CCF, a negative value may occur when the obligor has paid back a portion of the debt prior to default. For estimation purposes, it is, however, recommended to floor a negative CCF to zero.

The CCF can also exceed one. This can be due to credit limit changes or off-line transactions that allow borrowers to exceptionally exceed the credit limit. In other words, the limit communicated to the customer is a soft limit that the customer can occasionally exceed. If the CCF is above one, then the exposure decreases as the drawn balance increases.

Let us illustrate this with a brief example. Suppose that the credit limit is 2,500 and the CCF equals 110%. If the drawn balance is 1,000, then the EAD becomes 2,650. However, if the drawn balance equals 1,500, then the exposure decreases to 2,600. It is recommended to always have the CCF limited between zero and one. One way to ensure it is always below one is to work with a hard credit limit that absolutely cannot be exceeded. This hard credit limit can be set based on historical data or by using a confidence level, if needed.

EAD MODELING

Models for EAD and conversion measures can be created using similar techniques as the ones presented in our LGD chapter. Since EAD modeling is also exposed to default selection, models like the Tobit model and the Heckman selection model may provide interesting approaches.

Data Preprocessing

Let's now have a look at how we can create the development sample for EAD. For defaulted exposures, the EAD at the moment of default can be determined. We now need to consider a period δ_t before the time of default to determine the risk factors and drawn amount. The risk factors are the variables that will be used as predictors in the CCF model. Once the drawn amount is known, the CCF can be calculated as the ratio of the EAD minus the drawn amount, and the limit minus the drawn amount.

A key problem is the determination of the time lag between the observed exposure amount at default and the observed drawn amount and limit. The Prudential Regulation Authority (PRA) in the United Kingdom says the following about this:

> The PRA expects firms to use a time horizon of one year for EAD estimates; unless they can demonstrate that another period would be more conservative.

Various methods to determine this time horizon exist, and their impact on the composition of the development sample will be discussed in the following.

Cohort Method

A first way is the cohort approach. This approach groups defaulted facilities into discrete calendar periods, for example, 12 months unless another time period would be more conservative and appropriate. It collects information about the risk factors and drawn/undrawn amounts at the beginning of the calendar period and the drawn amount at the date of default. Data of different calendar periods can then be pooled for estimation.

For example, if the calendar period is defined as the 1st of November 2015 to the 31st of October 2016, then information about the risk factors and drawn/undrawn amounts is collected on November 1, 2015, and the drawn amounts of the defaulted facilities upon default. Facilities can then go into default in January 2016, February 2016, March 2016, until October 2016. In other words, the time lag between the risk factors and the EAD is different for each defaulter depending upon the time of default. This is also what will happen as we will start using our model for predicting the EAD. Some exposures may default in the next three months, four months, five months, and so on.

Fixed Time Horizon Method

The fixed horizon method starts by defining a fixed time horizon, usually 12 months, unless another time period would be more conservative and appropriate. It then collects information about the risk factors and drawn/undrawn amounts 12 months prior to the date of default, and the drawn amount on the date of default, regardless of the actual calendar date on which the default occurred. For example, if a default occurred on July 15, 2016, then information about the risk factors and the drawn/undrawn amounts of the defaulted facility on July 16, 2015, is used. Hence, the time frame between measuring the risk factors and the CCF is always 12 months. When we use the model to predict EAD, it generally assumes that if default occurs, it will occur exactly 12 months from now. About the cohort and fixed time horizon approach, the Prudential Regulation Authority says the following:

> EAD estimates can be undertaken on the basis that default occurs at any time during the time horizon (the cohort approach), or at the end of the time horizon (the fixed horizon approach). The PRA considers that either approach is acceptable in principle.

Variable Time Horizon Method

The variable time horizon approach is a variant of the fixed time horizon approach whereby several reference times within a chosen time horizon are used to determine the drawn/undrawn amounts and risk factors. For example, the drawn amount upon default is compared to the drawn/undrawn amounts and risk factors one, two, three, or more months before default.

We now showcase the data preparation for our mortgage data set. Note that these are loan exposures with flexible payment schedules and that low-risk mortgage borrowers often make prepayments, while high-risk borrowers usually do not, even if contractually allowed.

In a first step, we set up the data by computing the lagged exposure amounts (variable balance_time) for mortgage borrowers as a measure for the drawn amount at the beginning of the reference period. Consistent with the variable time horizon method, we consider one to four lags under the assumption that one period equals one quarter (i.e., a lag of four periods is equal to one year).

The ARRAY command in SAS allows us to set the first observations for which no lagged value is available (and otherwise would be taken from the previous borrower) to a missing value for multiple lags. The data are first sorted by the BY variable "id." We prefer to apply the calculations on all observations and not on default observations only, as we will later estimate regression models that control for the selecting default event.

```
PROC SORT DATA=data.mortgage;
BY id;
RUN;
DATA mortgage(drop=i count);
 SET data.mortgage;
 BY id;
 /* Create lagged variables*/
 ARRAY x(*) lag1-lag4;
 lag1=LAG1(balance_time);
 lag2=LAG2(balance_time);
 lag3=LAG3(balance_time);
 lag4=LAG4(balance_time);
 /* Reset count at the start of each new BY-Group */
 IF FIRST.id THEN COUNT=1;
 /* assign missing values to first observation/s of a BY group*/
 DO i=count TO DIM(x);
  x(i)=.;
 END;
 count + 1;
RUN;
```

A selection of variables of the resulting data set is shown in Exhibit 11.2.

id	time	default_ time	balance_ orig_ time	balance_ time	lag1	lag2	lag3	lag4
46	25	0	88000	86884.83	.	.	.	.
46	26	0	88000	86718.33	86884.83	.	.	.
46	27	0	88000	86379.71	86718.33	86884.83	.	.
46	28	0	88000	86207.57	86379.71	86718.33	86884.83	.
46	29	1	88000	86033.51	86207.57	86379.71	86718.33	86884.83
47	25	0	132000	130286.14	.	.	.	.
47	26	0	132000	129925.24	130286.14	.	.	.
47	27	0	132000	129558.06	129925.24	130286.14	.	.

Exhibit 11.2 Panel Data

Next, we generate the dependent variable based on the four concepts: CCF, CEQ, LCF, and UACF. (See Exhibit 11.3.) We derive theoretical limits for mortgages that allow prepayments and redraws prior to principal repayments, and suggest transformation functions that may be used as dependent variables in later regression models. The transformations ensure that the predicted conversion measures and hence EAD are within our expectations (i.e., do not exceed the limit or become negative).

Measure	Formula	Lower Bound	Upper Bound	Transformation
CCF	(EAD-Drawn)/(Limit-Drawn)	$-\infty$	1	−ln(1-CCF)
CEQ	(EAD-Drawn)/Limit	-1	1	ln((1+CEQ)/(1-CEQ))
LCF	EAD/Limit	0	1	ln(LCF/(1-LCF))
UACF	EAD/Drawn	0	∞	ln(UACF)

Exhibit 11.3 Definitions, Boundaries, and Transformations for Credit Conversion Measures

We define the limit as the outstanding loan amount at origination (variable “balance_orig_time”). The drawn amount is the outstanding loan amount prior to the observation period. We will focus on the four-period lag (variable “lag4”), which is equal to one year if one period equals a quarter. The exposure amount is the outstanding loan amount in the observation period (variable “balance_time”).

Furthermore, we cap the variables' exposure and drawn amount by the credit limit. CCF is set to zero if the drawn amount is equal to the limit. As we are interested in economic EAD and mortgages are exposed to prepayments, we do not impose the regulatory floor of the exposure by the drawn amount. Note that the latter assumption is required by the Basel Committee on Banking Supervision for regulatory EAD and may be implemented by an additional line (IF exposure<drawn THEN exposure=drawn;). We compute the transformations as highlighted in Exhibit 11.3. We add one restriction (IF drawn=limit THEN CCF=0;) as the data has observations where the limit is completely drawn.

```
DATA mortgage1(WHERE=(drawn NE . AND limit NE . AND exposure NE .
AND exposure NE 0));
SET mortgage;
/*Definitions*/
drawn=lag4;
limit=balance_orig_time;
exposure=balance_time;
/*Caps for exposure and draw*/
if exposure>limit then exposure=limit;
if drawn>limit then drawn=limit;
/*Conversion measures*/
CCF=(exposure-drawn)/(limit-drawn);
if drawn=limit then CCF=0;
CEQ=(exposure-drawn)/limit;
LCF=exposure/limit;
UACF=exposure/drawn;
RUN;
```

We compute the 5th and 95th percentiles of the conversion measures (see Exhibit 11.4):

The MEANS Procedure

Variable	1st Pctl	99th Pctl
CCF	−18.0502849	1.0000000
CEQ	−0.1297378	0.0102912
LCF	0.3724166	1.0000000
UACF	0.7492269	1.0105358

Exhibit 11.4 Percentiles for Conversion Measures

```
PROC MEANS DATA=mortgage1 P1 P99;
VAR CCF CEQ LCF UACF;
RUN;
```

We winsorize the conversion measures by the 1st and 99th percentile (i.e., floor the measures at their 1st percentile and cap the measures at their 99th percentile). We further confirm that the boundary values are not hit, as the numerator and the natural logarithm are not defined for zero. As a result, for CCF and LCF, we substitute an observation of one with 0.9999999.

We then transform the variables CCF, CEQ, LCF, and UACF by using the following transformation functions to match the range of normally distributed residuals in our regressions (see Exhibit 11.5:

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
CCF	11673	−0.6358660	2.0406573	−18.0502849	0.9999999
CEQ	11673	−0.0044892	0.0095413	−0.1297378	0.0102912
LCF	11673	0.9781223	0.0505083	0.3724166	0.9999999
UACF	11673	0.9949136	0.0146982	0.7492269	1.0105358
CCF_t	11673	0.1218104	2.6773378	−2.9470821	16.1180957
CEQ_t	11673	−0.0089835	0.0191379	−0.2609463	0.0205831
LCF_t	11673	8.1114383	5.5361960	−0.5218635	16.1180956
UACF_t	11673	−0.0052234	0.0162798	−0.2887134	0.0104807

Exhibit 11.5 Percentiles for Conversion Measures

```
DATA mortgage2;
SET mortgage1;
/*Floors*/
IF CCF<=-18.0502849 THEN CCF=-18.0502849;
IF CEQ<=-0.1297378 THEN CEQ=-0.1297378;
IF LCF<=0.3724166 THEN LCF=0.3724166;
IF UACF<=0.7492269 THEN UACF=0.7492269;
/*Caps*/
IF CCF>=0.9999999 THEN CCF=0.9999999;
IF CEQ>=0.0102912 THEN CEQ=0.0102912;
IF LCF>=0.9999999 THEN LCF=0.9999999;
IF UACF>=1.0105358 THEN UACF=1.0105358;
/*Transformations*/
```

```
CCF_t=-LOG(1-CCF);
CEQ_t=LOG((1+CEQ)/(1-CEQ));
LCF_t=LOG(LCF/(1-LCF));
UACF_t=LOG(UACF);
RUN;
```

We generate moments of the four dependent variables for the complete data set using PROC MEANS for CCF, CEQ, LCF, and UACF, as well as their transforms for all observations and by the default indicator:

```
PROC MEANS DATA=mortgage2(where=(default_time=1));
VAR CCF CEQ LCF UACF CCF_t CEQ_t LCF_t UACF_t;
RUN;
```

Furthermore, we generate the histograms using PROC UNIVARIATE for CCF, CEQ, LCF, and UACF, as well as their transforms (see Exhibits 11.6 through 11.13):

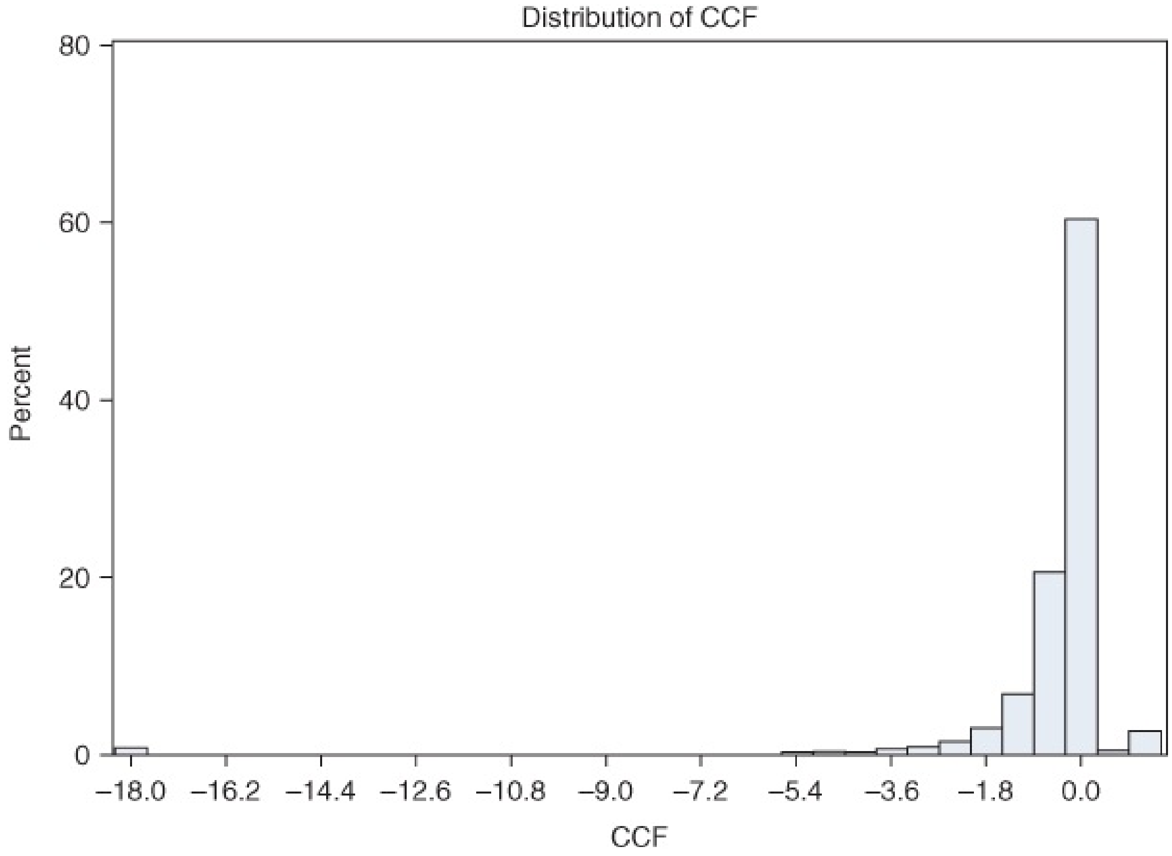


Exhibit 11.6 Histogram Credit Conversion Factor (CCF)

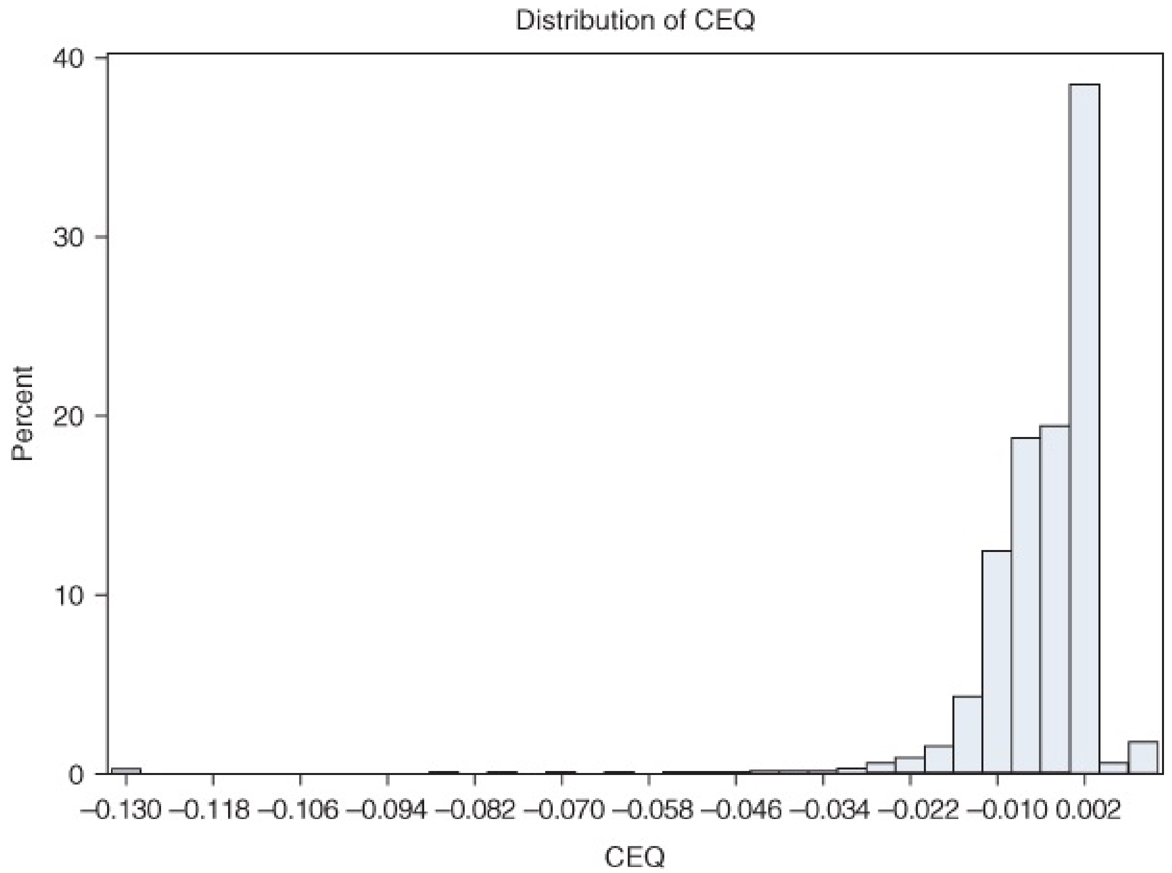


Exhibit 11.7 Histogram Credit Equivalent (CEQ)

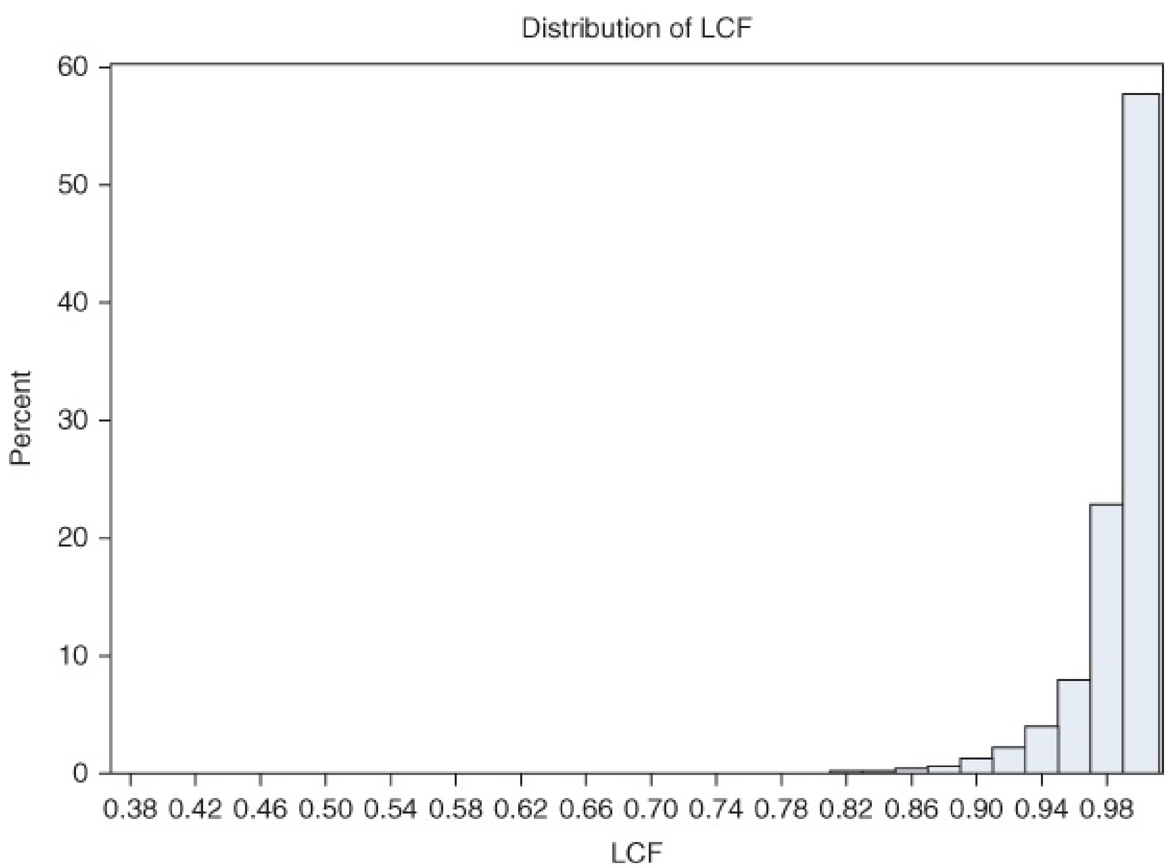


Exhibit 11.8 Histogram Limit Conversion Factor (LCF)

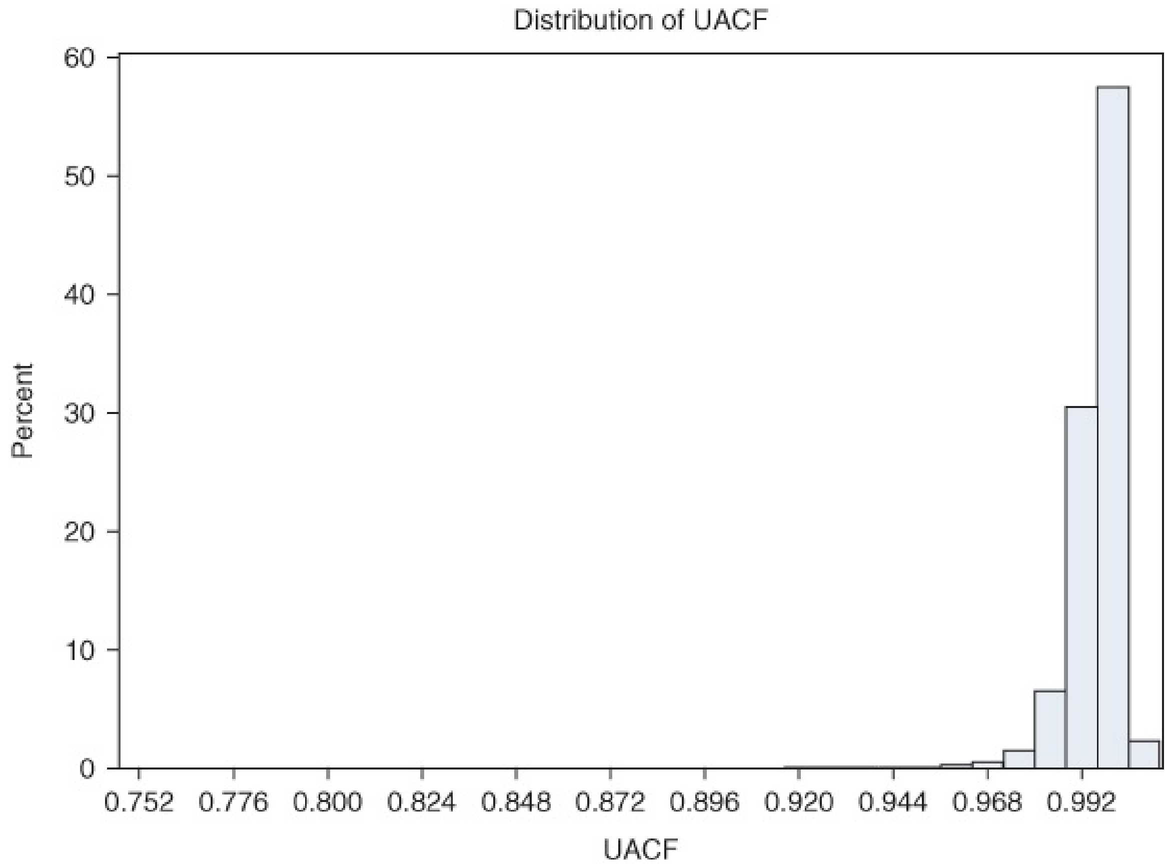


Exhibit 11.9 Histogram Used Amount Conversion Factor (UACF)

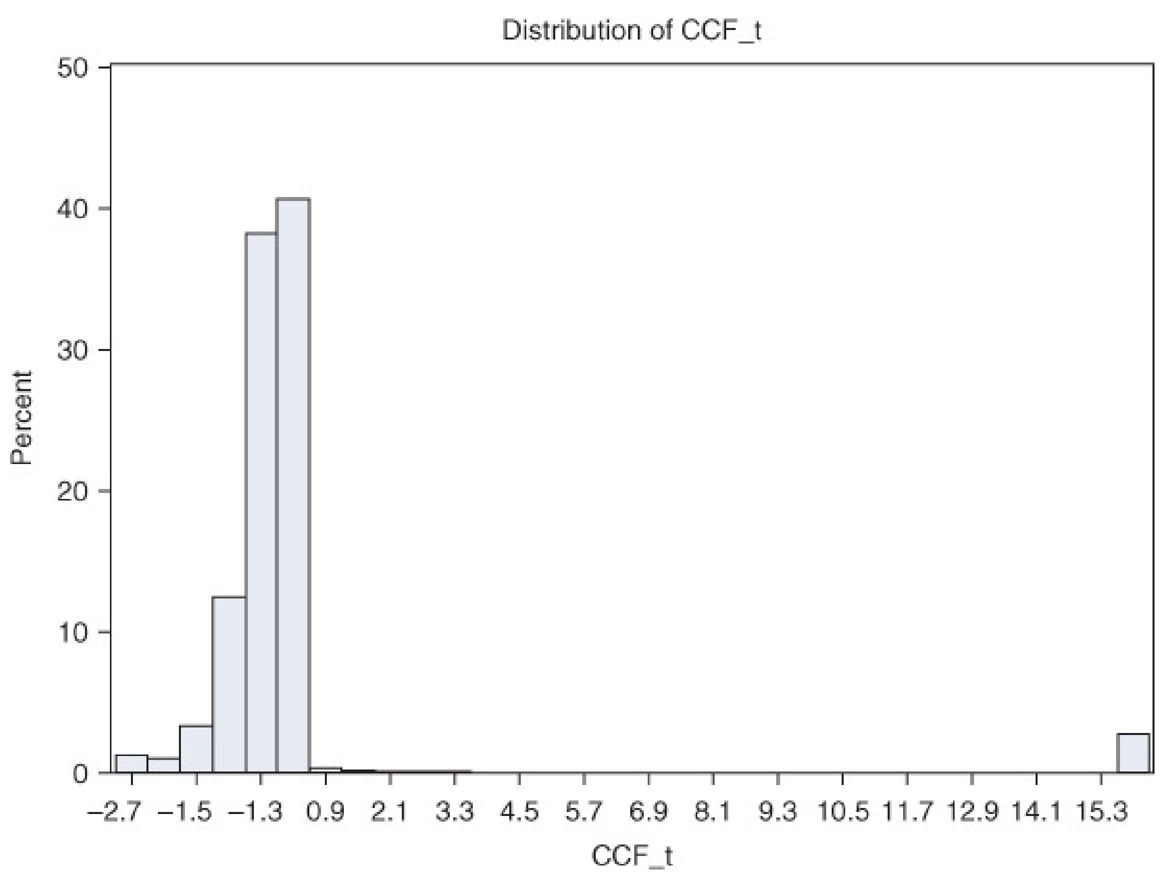


Exhibit 11.10 Histogram Transformed Credit Conversion Factor (CCF_t)

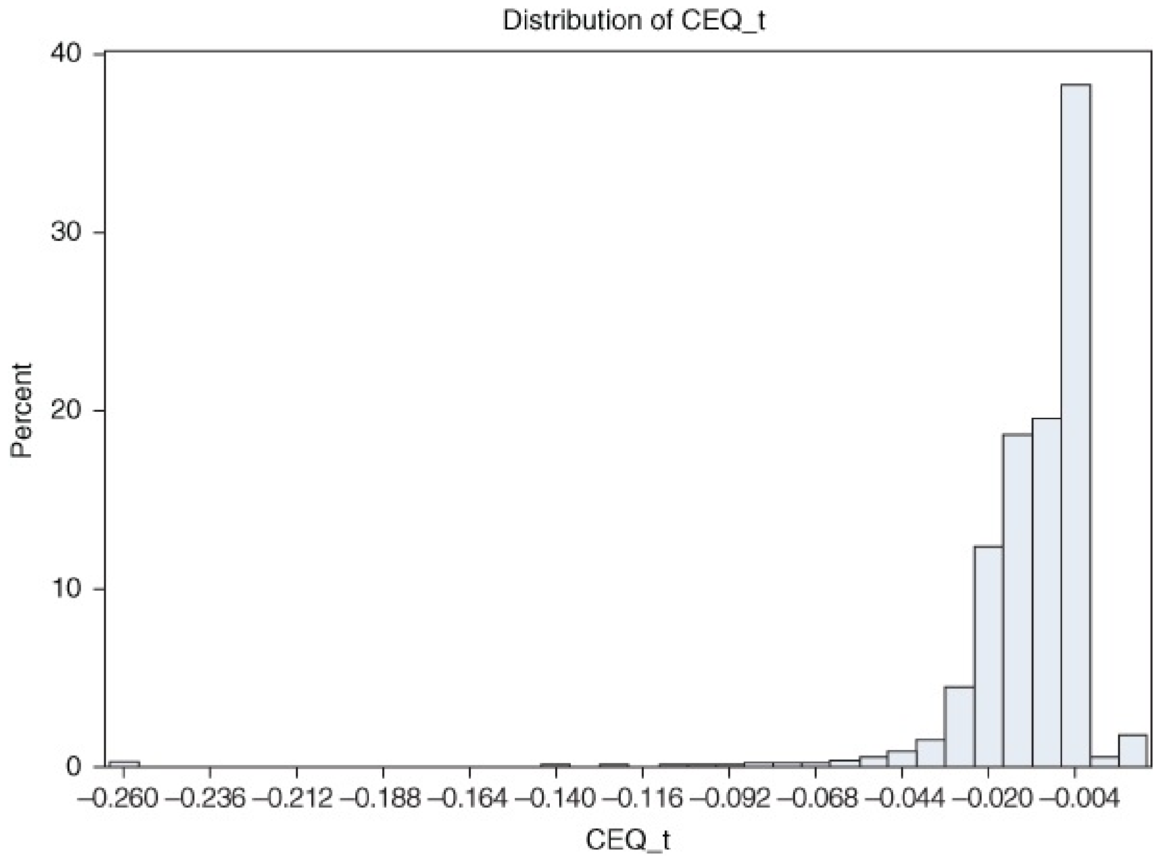


Exhibit 11.11 Histogram Transformed Credit Equivalent (CEQ_t)

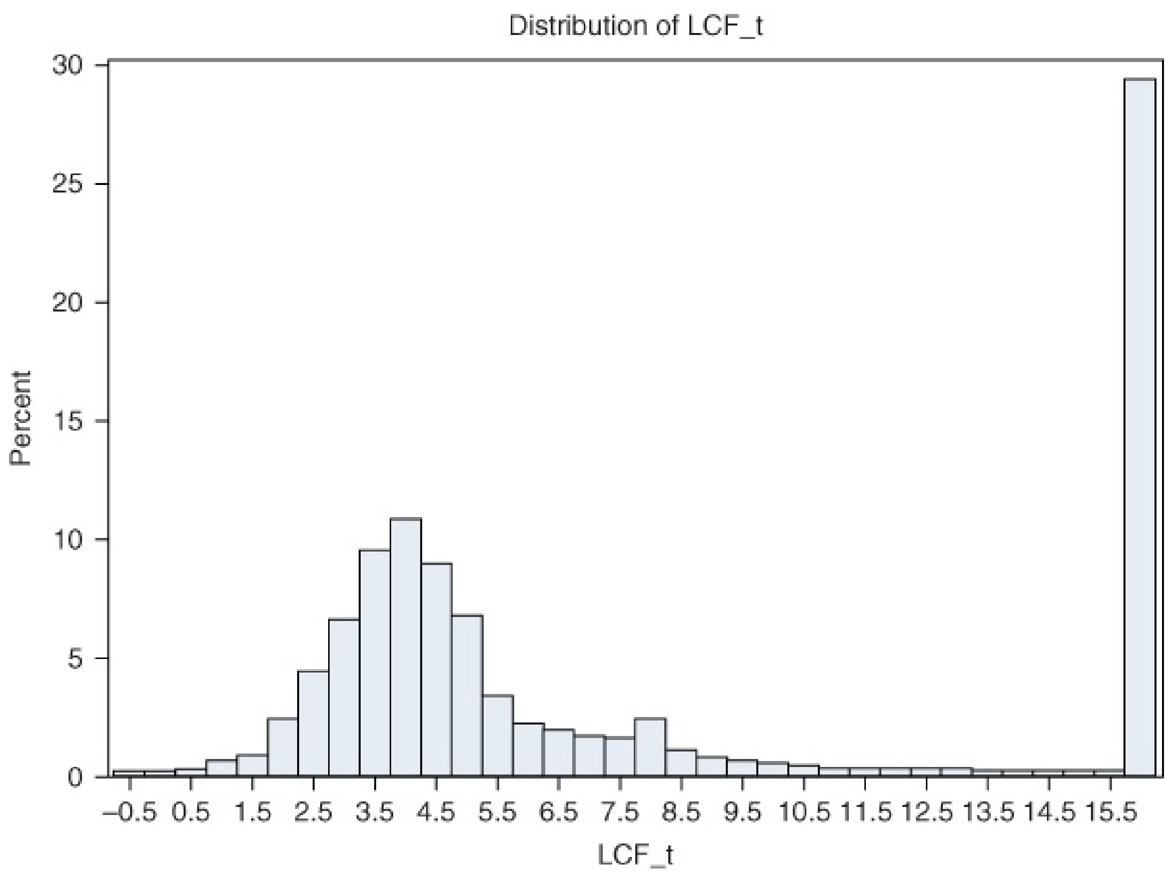


Exhibit 11.12 Histogram Transformed Limit Conversion Factor (LCF_t)

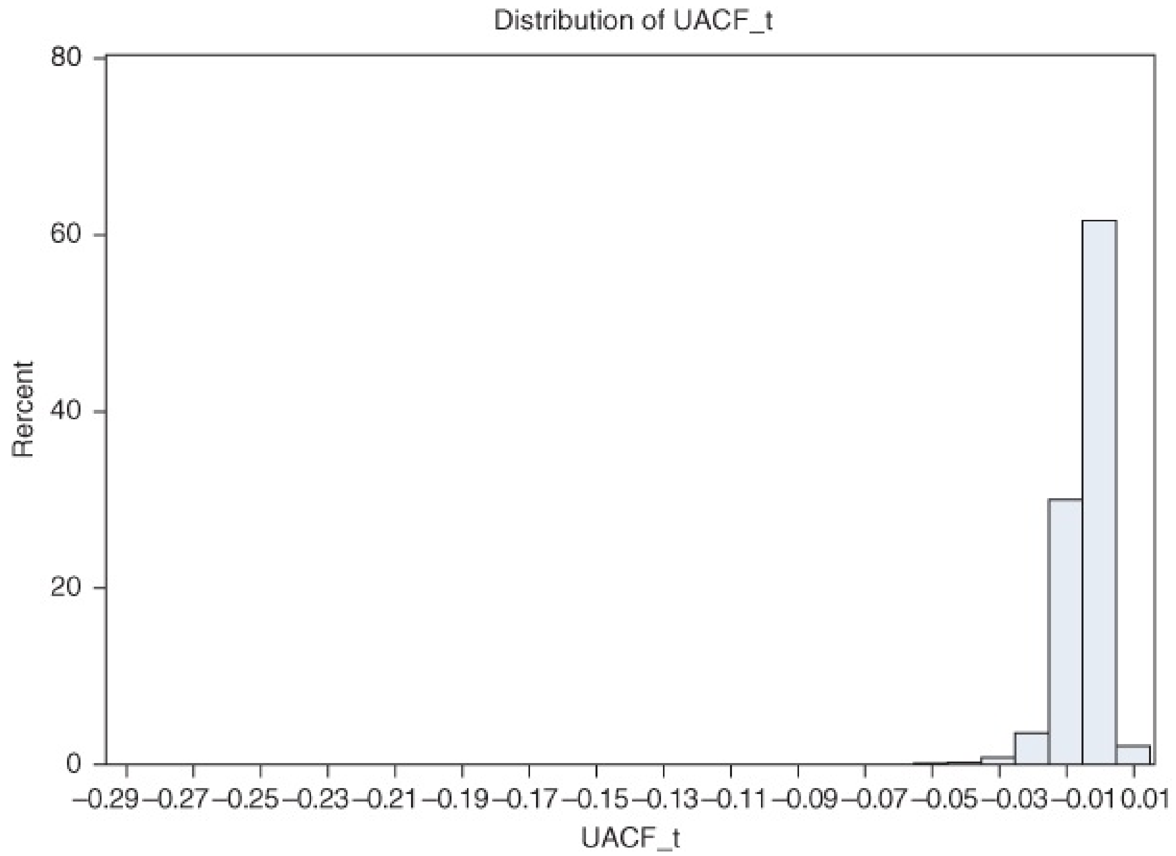


Exhibit 11.13 Histogram Transformed Used Amount Conversion Factor (UACF_t)

```
ODS GRAPHICS ON;
PROC UNIVARIATE DATA=mortgage2(where=(default_time=1));
VAR CCF CEQ LCF UACF CCF_t CEQ_t LCF_t UACF_t;
HISTOGRAM;
RUN;
ODS GRAPHICS OFF;
```

Credit Line Models

A number of risk factors can generally be considered in EAD models:

- Facility-specific: covenant protection, utilization, or LTV
- Borrower-specific: industry, geographical region, PD, credit rating, recent new loans/payoff, change in frequency of payment, or prepayment
- Time density: time to default, time since origination
- Macroeconomic variables

Furthermore, Moral (2011) recommends using the following risk factors:

- The committed amount, which is the advised credit limit at the start of the cohort
- The drawn amount, which is the exposure at the start of the cohort
- The undrawn amount, which is the advised limit minus the exposure at the start of the cohort

- The credit percentage usage, which is the exposure at the start of the cohort divided by the advised credit limit at the start of the cohort
- The time to default, which is the number of months between the start of the cohort and the default date
- The rating class, which is the behavioral score at the start of the cohort, binned into four categories

Let us now have a look at the risk factors that can be used to model CCF for credit lines. Exposure modeling for credit card loans has been the focus in the existing literature. For example, Tong et al. (2016) find a bimodal distribution for the CCF with a peak at zero and one, similar to LGDs. The authors include the following variables next to the variables from Moral (2011):

- The average number of days delinquent during the previous 3, 6, 9, and 12 months prior to the start of the cohort
- The increase in committed amount, which is a binary variable indicating whether there has been an increase in the committed amount since 12 months prior to the start of the cohort
- The undrawn percentage, which is the undrawn amount at the start of the cohort divided by the advised credit limit at the start of the cohort
- The relative change in drawn, undrawn, and committed amounts
- The absolute change in drawn, undrawn, and committed amounts

Tong et al. (2016) find the following significant variables: the credit percentage usage, the committed amount, the undrawn amount, the time-to-default, the rating class, and the average number of days delinquent in the past six months. Of the additional variables tested, only the average number of days delinquent in the past six months turned out to be significant. Note that the maximum performance in terms of R -squared equals about 0.10, which is rather low but in line with LGD modeling.

Loans with Flexible Payment Schedules

Linear Regression

To showcase regression models for the conversion measures, we estimate a linear regression model for each transform of CCF, CEQ, LCF, and UACF. We include LTV_time as the sole covariate and encourage the reader to go beyond this specification and enhance the model's accuracy:

```
   ODS GRAPHICS ON;
PROC REG DATA=mortgage2(WHERE=(default_time=1))
PLOTS(MAXPOINTS=20000 STATS= ALL)= DIAGNOSTICS;
   MODEL CCF = LTV_time;
   RUN;
   ODS GRAPHICS OFF;
```

```
   ODS GRAPHICS ON;
PROC REG DATA=mortgage2(WHERE=(default_time=1))
PLOTS(MAXPOINTS=20000 STATS= ALL)= DIAGNOSTICS;
   MODEL CEQ = LTV_time;
   RUN;
   ODS GRAPHICS OFF;

   ODS GRAPHICS ON;
PROC REG DATA=mortgage2(WHERE=(default_time=1))
PLOTS(MAXPOINTS=20000 STATS= ALL)= DIAGNOSTICS;
   MODEL LCF = LTV_time;
   RUN;
   ODS GRAPHICS OFF;

   ODS GRAPHICS ON;
PROC REG DATA=mortgage2(WHERE=(default_time=1))
PLOTS(MAXPOINTS=20000 STATS= ALL)= DIAGNOSTICS;
   MODEL UACF = LTV_time;
   RUN;
   ODS GRAPHICS OFF;
```

Exhibits 11.14 through 11.17 show that the R-squared is highest for LCF.

The REG Procedure

Model: MODEL1

Dependent Variable: CCF

Root MSE	2.03501	***R*-Squared**	0.0056
Dependent Mean	−0.63587	**Adj *R*-Sq**	0.0055
Coeff Var	−320.03702		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	***t* Value**	**Pr > |*t*|**
Intercept	1	−1.35217	0.09022	−14.99	<.0001
LTV_time	1	0.00710	0.00087437	8.12	<.0001

Exhibit 11.14 Linear Regression Fit for CCF

The REG Procedure
Model: MODEL1
Dependent Variable: CEQ

Root MSE	0.00906	*R*-Squared	0.0983
Dependent Mean	−0.00449	Adj *R*-Sq	0.0982
Coeff Var	−201.83151		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	*t* Value	Pr > |*t*|
Intercept	1	−0.01850	0.00040169	−46.06	<.0001
LTV_time	1	0.00013886	0.00000389	35.67	<.0001

Exhibit 11.15 Linear Regression Fit for CEQ

The REG Procedure
Model: MODEL1
Dependent Variable: LCF

Root MSE	0.04605	*R*-Squared	0.1688
Dependent Mean	0.97812	Adj *R*-Sq	0.1687
Coeff Var	4.70817		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	*t* Value	Pr > |*t*|
Intercept	1	0.88093	0.00204	431.48	<.0001
LTV_time	1	0.00096316	0.00001979	48.68	<.0001

Exhibit 11.16 Linear Regression Fit for LCF

The REG Procedure
Model: MODEL1
Dependent Variable: UACF

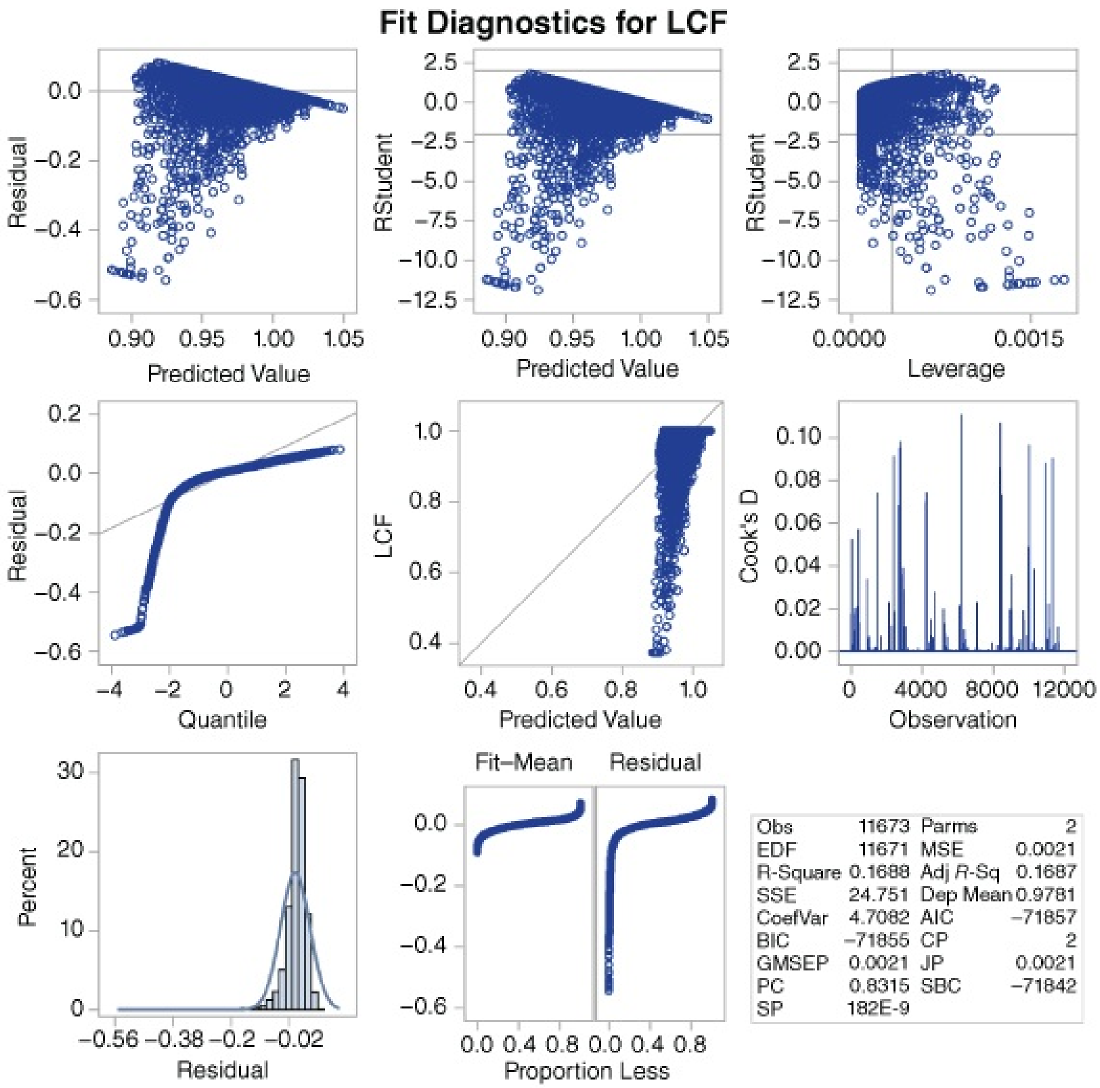
Root MSE	0.01411	*R*-Squared	0.0781
Dependent Mean	0.99491	Adj *R*-Sq	0.0780
Coeff Var	1.41856		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	*t* Value	Pr > |*t*|
Intercept	1	0.97568	0.00062570	1559.34	<.0001
LTV_time	1	0.00019063	0.00000606	31.44	<.0001

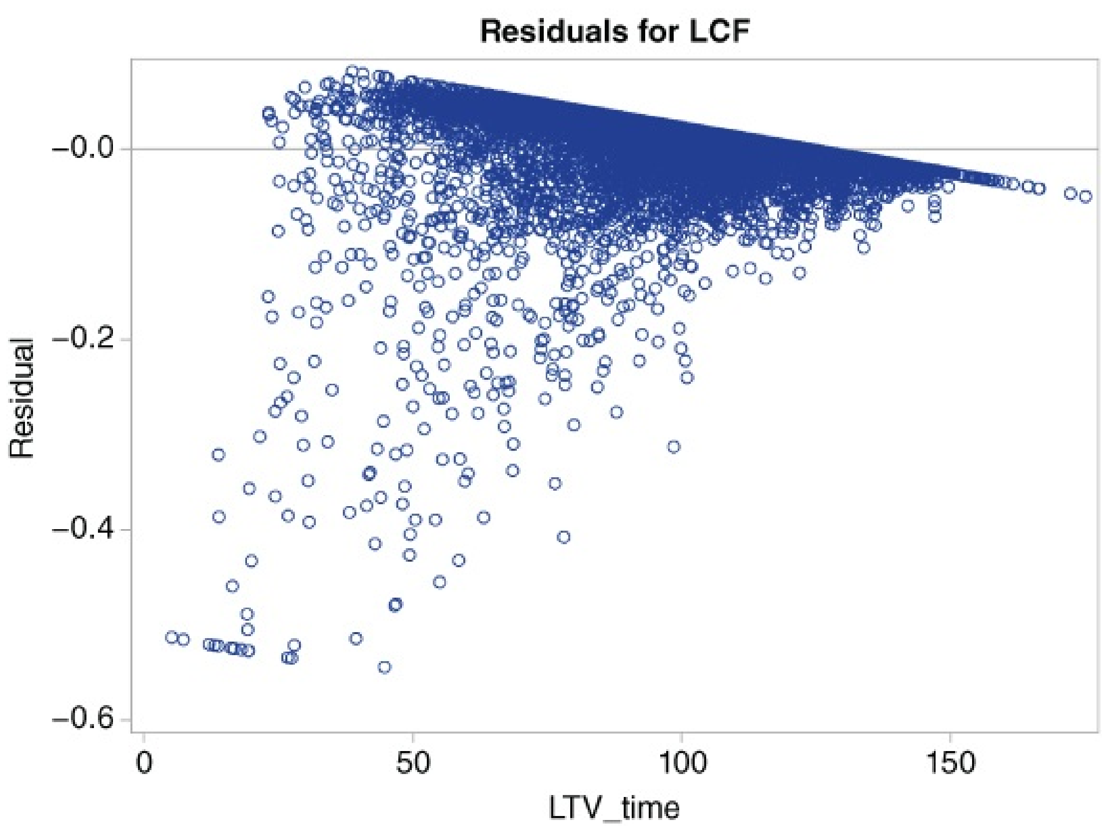
Exhibit 11.17 Linear Regression Fit for UACF

The fit diagnostics and residual plots for a linear model for the best-fitting (with regard to R -

squared) measure LCF are shown in Exhibit 11.18.



Fit Diagnostics for LCF
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
LCF
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Obs 11673
Parms 2
EDF 11671
MSE 0.0021
R-Square 0.1688
Adj R-Sq 0.1687
SSE 24.751
Dep Mean 0.9781
CoefVar 4.7082
AIC −71857
BIC −71855
CP 2
GMSEP 0.0021
JP 0.0021
PC 0.8315
SBC −71842
SP 182E-9



Residuals for LCF
Residual
LTV_time

Exhibit 11.18 Linear Regression Fit and Residuals for LCF

The comparison of the fitted normal distribution and the histogram of the residuals shows that the fit is not that good. Thus, we try the transformations of the conversion measures as a next step.

Transformed Linear Regression

We regress our transformed conversion measures on the covariate LTV_time.

```
   ODS GRAPHICS ON;
PROC REG DATA=mortgage2(WHERE=(default_time=1))
PLOTS(MAXPOINTS=20000 STATS= ALL)= DIAGNOSTICS;
   MODEL CCF_t = LTV_time;
   RUN;
   ODS GRAPHICS OFF;

   ODS GRAPHICS ON;
PROC REG DATA=mortgage2(WHERE=(default_time=1))
PLOTS(MAXPOINTS=20000 STATS= ALL)= DIAGNOSTICS;
   MODEL CEQ_T = LTV_time;
   RUN;
   ODS GRAPHICS OFF;

   ODS GRAPHICS ON;
PROC REG DATA=mortgage2(WHERE=(default_time=1))
PLOTS(MAXPOINTS=20000 STATS= ALL)= DIAGNOSTICS;
   MODEL LCF_t = LTV_time;
   RUN;
   ODS GRAPHICS OFF;

   ODS GRAPHICS ON;
PROC REG DATA=mortgage2(WHERE=(default_time=1))
PLOTS(MAXPOINTS=20000 STATS= ALL)= DIAGNOSTICS;
   MODEL UACF_t = LTV_time;
   RUN;
   ODS GRAPHICS OFF;
```

Exhibits 11.19 through 11.22 show that the R-squared is highest for the transformed LCF.

The REG Procedure
Model: MODEL1
Dependent Variable: CCF_t

Root MSE	2.65749	*R*-Squared	0.0149
Dependent Mean	0.12181	Adj *R*-Sq	0.0148
Coeff Var	2181.66187		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	*t* Value	Pr > |*t*|
Intercept	1	−1.40669	0.11782	−11.94	<.0001
LTV_time	1	0.01515	0.00114	13.27	<.0001

Exhibit 11.19 Transformed Linear Regression Fit for CCF_t

The REG Procedure
Model: MODEL1
Dependent Variable: CEQ_t

Root MSE	0.01818	*R*-Squared	0.0979
Dependent Mean	−0.00898	Adj *R*-Sq	0.0979
Coeff Var	−202.34252		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	*t* Value	Pr > |*t*|
Intercept	1	−0.03704	0.00080587	−45.96	<.0001
LTV_time	1	0.00027803	0.00000781	35.60	<.0001

Exhibit 11.20 Transformed Linear Regression Fit for CEQ_t

The REG Procedure
Model: MODEL1
Dependent Variable: LCF_t

Root MSE	5.21231	*R*-Squared	0.1137
Dependent Mean	8.11144	Adj *R*-Sq	0.1136
Coeff Var	64.25870		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	*t* Value	Pr > |*t*|
Intercept	1	−0.63130	0.23108	−2.73	0.0063
LTV_time	1	0.08664	0.00224	38.69	<.0001

Exhibit 11.21 Transformed Linear Regression Fit for LCF_t

The REG Procedure
Model: MODEL1
Dependent Variable: UACF_t

Root MSE	0.01570	*R*-Squared	0.0706
Dependent Mean	−0.00522	Adj *R*-Sq	0.0705
Coeff Var	−300.48693		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	*t* Value	Pr > |*t*|
Intercept	1	−0.02548	0.00069584	−36.62	<.0001
LTV_time	1	0.00020073	0.00000674	29.77	<.0001

Exhibit 11.22 Transformed Linear Regression Fit for UACF_t

Generally speaking, our transformation models have a poorer fit, and we continue by estimating a beta regression model for the currently dominating model for LCF, as this might provide a better fit of the empirical distribution of the residuals.

Beta Regression

We now apply a beta regression for LCF as we suspect that, similar to LGD, this may also be an interesting approach for EAD modeling. Next to a better model fit, LCF has the benefit of the same range of values as LGD and does not require a further transformation to the range of zero to one. We encourage the reader to explore other conversion measures, covariates, and other regression models such as nonlinear regressions, fractional logit regression, and a beta regression from the previous LGD chapter. Furthermore, an application of the LGD-PD model may be able to control for the bias of the selecting default event (see Exhibit 11.23).

The NLMIXED Procedure

Specifications	
Data Set	WORK.MORTGAGE2
Dependent Variable	LCF
Distribution for Dependent Variable	General
Optimization Technique	Trust Region
Integration Method	None

Dimensions	
Observations Used	11673
Observations Not Used	0
Total Observations	11673
Parameters	4

Fit Statistics	
-2 Log-Likelihood	−124E3
AIC (smaller is better)	−124E3
AICC (smaller is better)	−124E3
BIC (smaller is better)	−124E3

Parameter Estimates								
Parameter	Estimate	Standard Error	DF	*t* Value	Pr > |*t*|	95% Confidence Limits		Gradient
b0	0.8062	0.07272	12E3	11.09	<.0001	0.6637	0.9487	0.000013
b1	0.03272	0.000757	12E3	43.22	<.0001	0.03123	0.03420	0.001216
c0	0.6939	0.08126	12E3	8.54	<.0001	0.5346	0.8531	−0.00001
c1	0.01769	0.000833	12E3	21.22	<.0001	0.01605	0.01932	−0.00160

Exhibit 11.23 Beta Regression for LCF

```
ODS GRAPHICS ON;
PROC NLMIXED DATA=mortgage2(WHERE=(default_time=1)) TECH =TRUREG;
PARMS   b0 = 0 b1 = 0.001
c0 = 0 c1 = 0.001 ;
*Linear predictors;
Xb = b0 + b1 * LTV_time ;
Wc = c0 + c1 * LTV_time ;
mu = 1 / (1 + exp(-xb));
delta = EXP(Wc);
*transform to standard parameterization;
alpha = mu * delta;
beta = (1-mu) * delta;
*log-likelihood;
lh = (GAMMA(alpha + beta) / (GAMMA(alpha) * GAMMA(beta))
```

```
* (LCF ** (alpha - 1)) * ((1 - LCF) ** (beta - 1)));
ll = LOG(lh);
MODEL LCF ~ GENERAL(ll);
PREDICT mu OUT = out_mu;
PREDICT delta OUT = out_delta;
RUN;
ODS GRAPHICS OFF;
```

As for modeling LGDs, we use the predicted values for μ and produce a real-fit plot using PROC GPLOT of the realized LGDs versus the predicted values as illustrated next. We also run a linear regression using PROC REG of the realizations against the predictions and compute the R^2. The regression reveals that the intercept is statistically different from zero and the slope is statistically different from one. (See Exhibits 11.24 and 11.25.)

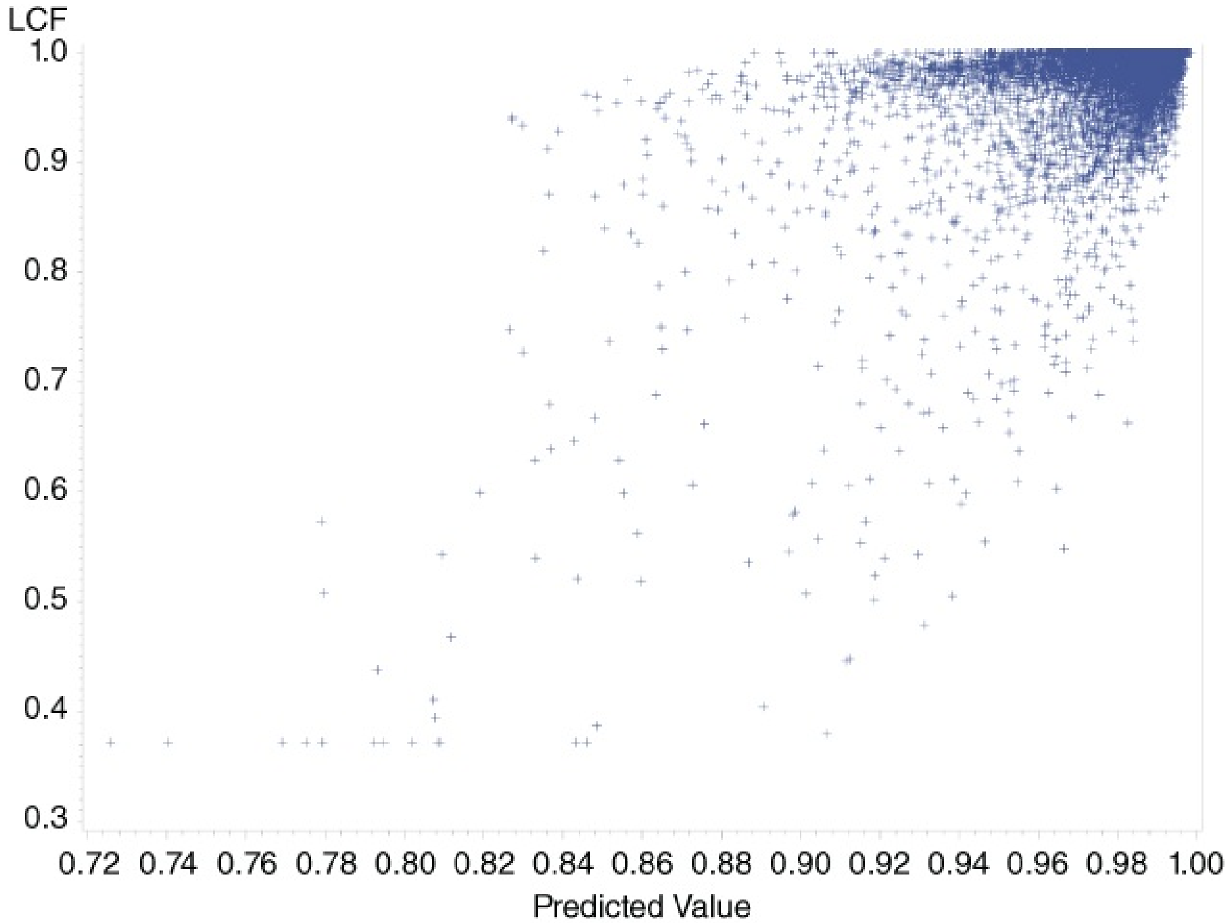


Exhibit 11.24 Real-Fit Plot of Beta Regression for LCF

Root MSE	0.04250	***R*-Squared**	0.2922
Dependent Mean	0.97812	**Adj *R*-Sq**	0.2921
Coeff Var	4.34461		

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error Error**	***t* Value**	**Pr > |*t*|**
Intercept	Intercept	1	−0.32001	0.01871	−17.11	<.0001
Pred	Predicted Value	1	1.32615	0.01911	69.41	<.0001

Exhibit 11.25 Real-Fit Regression of Beta Regression for LCF

```
ODS GRAPHICS ON;
PROC GPLOT DATA=out_mu;
PLOT LCF * PRED;
RUN;
PROC REG DATA = out_mu;
MODEL LCF = PRED;
RUN;
ODS GRAPHICS OFF;
```

Interestingly, the R-squared is now much higher than in the case of linear regression and transformed linear regression models.

Controlling for Adverse Selection in PD Models

Interaction of PD and EAD

Generally speaking, low-risk borrowers are more likely to reduce the loan balance below the expected or scheduled balance, while high-risk borrowers are more likely to increase the loan balance above the expected balance. Consider, for example, a revolving credit line or credit card. If the financial distress worsens, the obligor might try to draw down as much as possible on his or her existing unutilized credit lines in order to avoid default, thereby significantly increasing the EAD risk.

In other words, exposures generally increase in the case of default. Another important aspect for mortgage lending is that borrowers typically have a prepayment option, which they generally exercise if they are low risk and have excess cash to repay the mortgage loan. A prepayment has a limited impact and can be controlled for by a separate risk factor. However, after the complete mortgage has been paid off, low-risk borrowers leave the observed population, while high-risk mortgage borrowers remain. It is common in such situations to model the competing states default, payoff, and nondefault/nonpayoff by the following status variable:

$$S_{it} = \begin{cases} 1 & \text{borrower } i \text{ defaults at time } t \\ 2 & \text{borrower } i \text{ pays loan off at time } t \\ 0 & \text{otherwise} \end{cases}$$

To account for this, you may consider a discrete-time hazard model such as the multinomial logit or probit model, or a continuous-time hazard model (compare Deng, Quigley, and Van Order 2000).

Discrete Time Hazard Models

The first option is a multinomial logit (or probit) model that is a representative of the class of discrete-time hazard models as follows:

$$P(S_{it} = s|\boldsymbol{x}_{it-1}) = \frac{exp\ (\boldsymbol{\beta}_s'\boldsymbol{x}_{it-1})}{1 + \sum_{s=1}^{2} exp\ (\boldsymbol{\beta}_s'\boldsymbol{x}_{it-1})}$$

We can estimate a multinomial logit model in SAS using PROC LOGISTIC (see Exhibit 11.26):

The LOGISTIC Procedure

Model Information	
Data Set	DATA.MORTGAGE
Response Variable	status_time
Number of Response Levels	3
Model	Generalized logit
Optimization Technique	Newton-Raphson

Number of Observations Read	622489
Number of Observations Used	622219

Response Profile		
Ordered Value	status_time	Total Frequency
1	0	580484
2	1	15153
3	2	26582

Note	Logits modeled use status_time=0 as the reference category.

Model Convergence Status
Convergence criterion (GCONV=1E−8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	360828.66	346768.56
SC	360851.35	346859.28
-2 Log L	360824.66	346752.56

R-Squared	0.0224	Max-Rescaled *R*-Squared	0.0508

Testing Global Null Hypothesis: Beta = 0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	14072.1083	6	<.0001
Score	14159.6297	6	<.0001
Wald	13072.9694	6	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter	status_time	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1	−1.5286	0.0833	337.0219	<.0001
Intercept	2	1	−1.2835	0.0607	447.4465	<.0001
FICO_orig_time	1	1	−0.00545	0.000116	2195.5817	<.0001
FICO_orig_time	2	1	−0.00122	0.000084	208.9571	<.0001
LTV_time	1	1	0.0178	0.000373	2281.1240	<.0001
LTV_time	2	1	−0.0159	0.000271	3444.4963	<.0001
gdp_time	1	1	−0.1059	0.00375	797.6120	<.0001
gdp_time	2	1	0.1460	0.00452	1042.9821	<.0001

Exhibit 11.26 Multinomial Logit Model

```
PROC LOGISTIC DATA=data.mortgage;
CLASS status_time (REF='0');
MODEL status_time = FICO_orig_time LTV_time gdp_time/ LINK=GLOGIT RSQUARE;
OUTPUT OUT=probabilities PREDICTED=p;
RUN;
```

We obtain two parameter estimates for every covariate. FICO decreases the probabilities of default and payoff. LTV increases the probability of default but decreases the probability of payoff. GDP growth decreases the probability of default and increases the probability of payoff.

To compute performance measures, we generally include the estimated default (payoff) probabilities in a discrete-time hazard model with the default (payoff) indicator as the dependent variable, and obtain the standard pseudo-R-squared, AUROC, and accuracy ratio measures. For details, refer to the chapter on discrete-time hazard models.

Estimation of Default Probabilities

The probabilities of default can be estimated with the OUTPUT statement, which has evaluated the model equation and estimated parameters (indicated by a hat) as follows:

$$\hat{P}(S_{it} = 1 | \boldsymbol{x}_{it-1}) = \frac{\exp(\hat{\boldsymbol{\beta}}_1' \boldsymbol{x}_{it-1})}{1 + \exp(\hat{\boldsymbol{\beta}}_1' \boldsymbol{x}_{it-1}) + \exp(\hat{\boldsymbol{\beta}}_2' \boldsymbol{x}_{it-1})}$$

with

$$\hat{\boldsymbol{\beta}}_1' \boldsymbol{x}_{it-1} = \hat{\beta}_{0,1} + \hat{\beta}_{1,1} * \text{FICO\_orig\_time} + \hat{\beta}_{2,1} * \text{LTV\_time} + \hat{\beta}_{3,1} * \text{gdp\_time}$$
$$\hat{\boldsymbol{\beta}}_2' \boldsymbol{x}_{it-1} = \hat{\beta}_{0,2} + \hat{\beta}_{1,2} * \text{FICO\_orig\_time} + \hat{\beta}_{2,2} * \text{LTV\_time} + \hat{\beta}_{3,2} * \text{gdp\_time}$$

with $\hat{\beta}_{0,1}$ to $\hat{\beta}_{3,2}$ being the estimated default parameters. A PROC MEANS calculates the mean for the default indicators and the estimated PDs (see Exhibit 11.27).

The MEANS Procedure

Variable	Mean	Variable	Mean
default_time	0.0243506	p	0.0243532

Exhibit 11.27 Calibration of Multinomial Logit Models: Comparison of Default Indicators and Estimated Default Probabilities

```
PROC MEANS DATA=probabilities(WHERE=(_LEVEL_=1)) MEAN NOLABELS;
VAR default_time p;
RUN;
```

The calibration is clear, as the mean of the default event almost matches the mean of the estimated PDs. The minor difference that we observe is due to the estimation algorithm that iteratively maximizes the likelihood and stops if a target function indicates a low model improvement.

We now compute the default rates and average estimated default probabilities by time:

```
PROC SORT DATA=probabilities;
BY time;
RUN;
PROC MEANS DATA=probabilities(WHERE=(_LEVEL_=1));
BY time;
OUTPUT OUT=means MEAN(default_time p)=default_time p;
RUN;
```

The chart in Exhibit 11.28 compares the observed default rate (DR) and the average of the estimated default probabilities (PD):

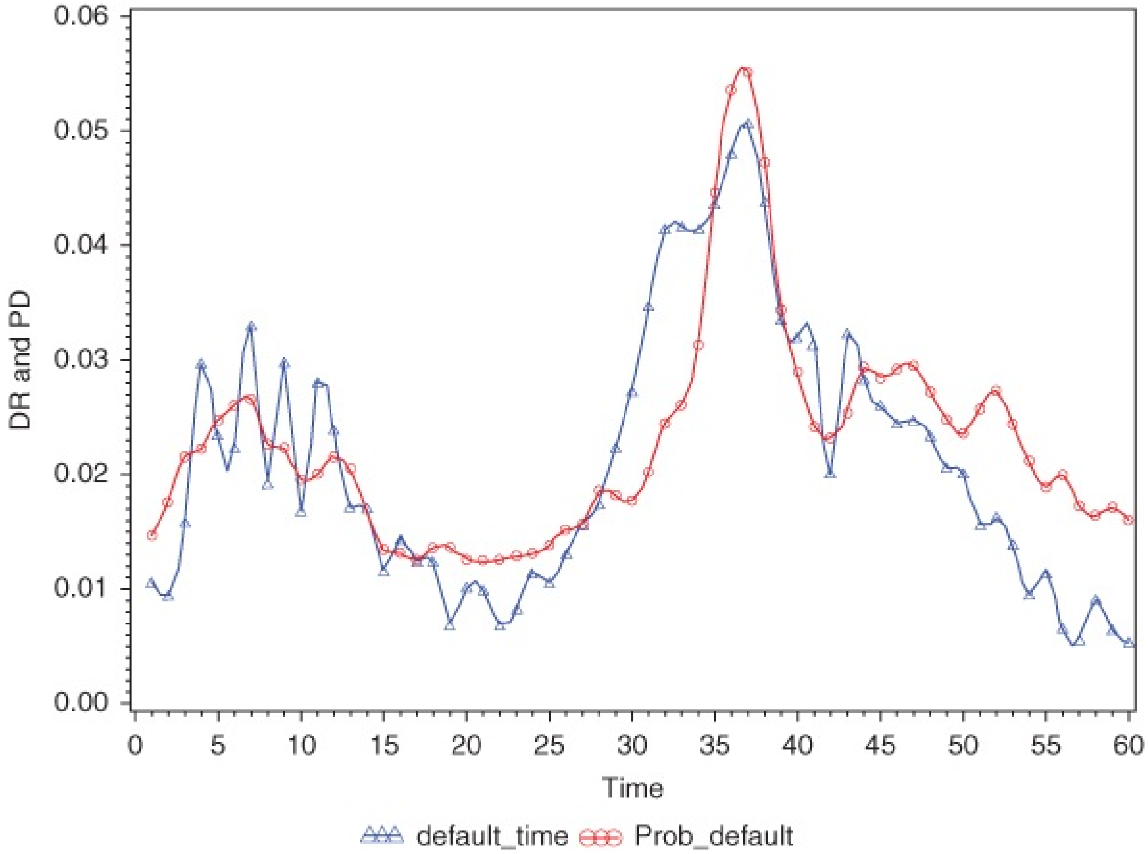


Exhibit 11.28 Real-Fit Diagram for the Default Probabilities

```
DATA means;
SET means;
LABEL p="Prob_default";
RUN;
ODS GRAPHICS ON;
AXIS1 ORDER=(0 to 60 by 5) LABEL=('Time');
AXIS2 order=(0 to 0.06 by 0.01) LABEL=('DR and PD');
SYMBOL1 INTERPOL=SPLINE WIDTH=2 VALUE=TRIANGLE C=BLUE;
SYMBOL2 INTERPOL=SPLINE WIDTH=2 VALUE=CIRCLE C=RED;
LEGEND1 LABEL=NONE SHAPE=SYMBOL(4,2) POSITION=(bottom outside);
PROC GPLOT DATA=means;
PLOT (default_time p)*time
/ OVERLAY HAXIS=AXIS1 VAXIS=AXIS2 LEGEND=LEGEND1;
RUN;
ODS GRAPHICS OFF;
```

Estimation of Payoff Probabilities

Similarly, the probabilities of payoff can be estimated with the OUTPUT statement, which has evaluated the model equation and estimated parameters (indicated by a hat) as follows:

$$\hat{P}(S_{it} = 2|\boldsymbol{x}_{it-1}) = \frac{\exp(\hat{\boldsymbol{\beta}}_2'\boldsymbol{x}_{it-1})}{1 + \exp\,(\hat{\boldsymbol{\beta}}_1'\boldsymbol{x}_{it-1}) + \exp(\hat{\boldsymbol{\beta}}_2'\boldsymbol{x}_{it-1})}$$

with

$$\hat{\boldsymbol{\beta}}_1' \boldsymbol{x}_{it-1} = \hat{\beta}_{0,1} + \hat{\beta}_{1,1} * \text{FICO\_orig\_time} + \hat{\beta}_{2,1} * \text{LTV\_time} + \hat{\beta}_{3,1} * \text{gdp\_time}$$

$$\hat{\boldsymbol{\beta}}_2' \boldsymbol{x}_{it-1} = \hat{\beta}_{0,2} + \hat{\beta}_{1,2} * \text{FICO\_orig\_time} + \hat{\beta}_{2,2} * \text{LTV\_time} + \hat{\beta}_{3,2} * \text{gdp\_time}$$

with $\hat{\beta}_{0,1}$ to $\hat{\beta}_{3,2}$ being the estimated payoff parameters. A PROC MEANS calculates the mean for the payoff indicators and the estimated payoff probabilities (see Exhibit 11.29).

The MEANS Procedure

Variable	Mean	Variable	Mean
payoff_time	0.0427140	p	0.0427213

Exhibit 11.29 Calibration of Multinomial Logit Models: Comparison of Payoff Indicators and Estimated Payoff Probabilities

```
PROC MEANS DATA=probabilities(WHERE=(_LEVEL_=2)) MEAN NOLABELS;
VAR payoff_time p;
RUN;
```

The calibration is clear, as the mean of the payoff event almost matches the mean of the estimated payoff probabilities.

We now compute the payoff rates and average estimated payoff probabilities by time:

```
PROC SORT DATA=probabilities;
BY time;
RUN;
PROC MEANS DATA=probabilities(WHERE=(_LEVEL_=2));
BY time;
OUTPUT OUT=means MEAN(payoff_time p)=default_time p;
RUN;
```

The chart in Exhibit 11.30 compares the observed payoff rate (PR) and the average of the estimated payoff probabilities (PP):

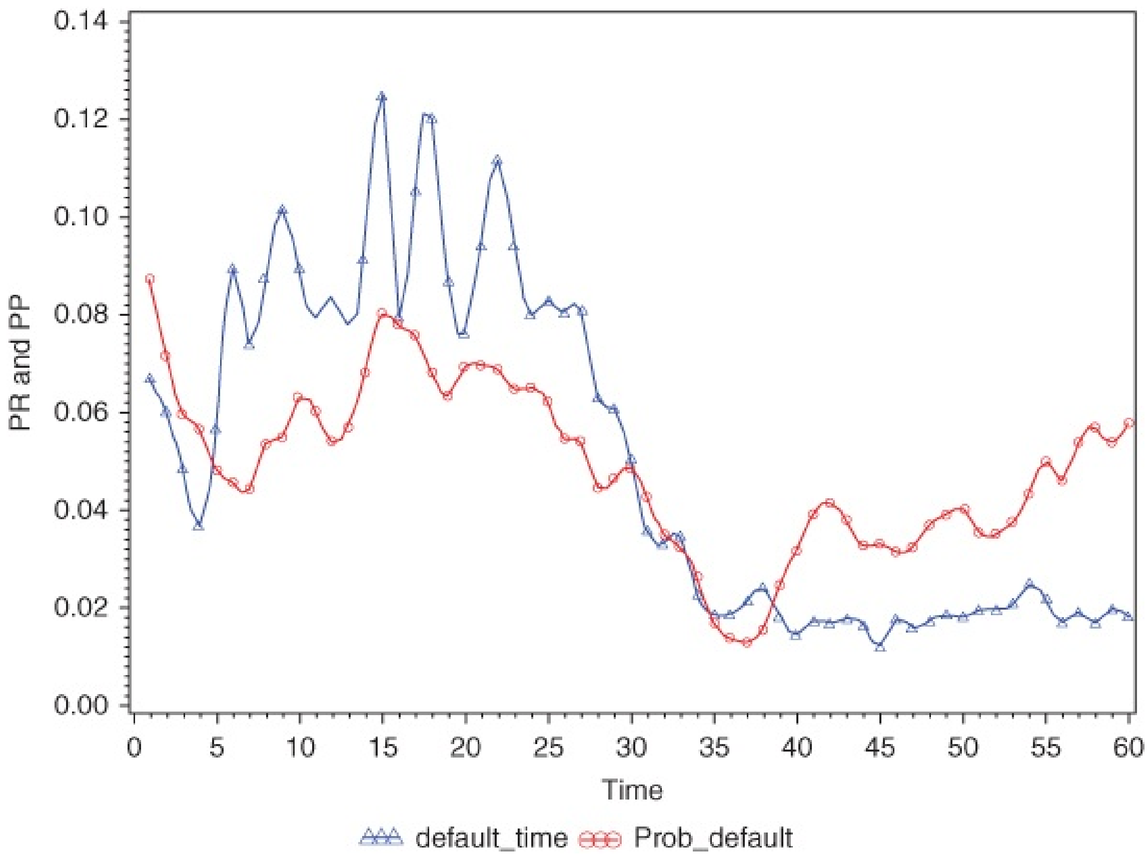


Exhibit 11.30 Real-Fit Diagram for the Payoff Probabilities

```
DATA means;
SET means;
LABEL p="Prob_payoff";
RUN;
ODS GRAPHICS ON;
AXIS1 ORDER=(0 to 60 by 5) LABEL=('Time');
AXIS2 order=(0 to 0.14 by 0.02) LABEL=('PR and PP');
SYMBOL1 INTERPOL=SPLINE WIDTH=2 VALUE=TRIANGLE C=BLUE;
SYMBOL2 INTERPOL=SPLINE WIDTH=2 VALUE=CIRCLE C=RED;
LEGEND1 LABEL=NONE SHAPE=SYMBOL(4,2) POSITION=(bottom outside);
PROC GPLOT DATA=means;
PLOT (default_time p)*time
/ OVERLAY HAXIS=AXIS1 VAXIS=AXIS2 LEGEND=LEGEND1;
RUN;
ODS GRAPHICS OFF;
```

Generally speaking, default rates and payoff rates move in opposite directions, which explains the opposite signs of the covariates, in particular the macroeconomic factors.

Continuous-Time Hazard Models

Reshaping the Data

We first replicate the data steps from our chapter on continuous-time hazard models as follows:

```
PROC SORT DATA=data.mortgage;
BY id;
RUN;
DATA lifetest_temp1;
SET data.mortgage;
time2 = time-first_time+1;
BY id;
RETAIN id;
IF LAST.id THEN indicator=1;
RUN;
DATA lifetest_temp2;
SET lifetest_temp1;
IF indicator = 1 OR status_time =1 OR status_time =2;
RUN;
DATA lifetest;
SET lifetest_temp2;
BY id;
RETAIN id;
IF FIRST.id THEN output;
RUN;
```

Exhibit 11.31 shows three borrower observations as an example. The time stamp time2 shows the time to default since the first observation period. This is contrary to the panel data set applied for discrete-time hazard models, where the time indicates the absolute time and loans were generally originated at different times. The first mortgage results in a default event, the second mortgage in a payoff event, and the third mortgage in neither. The second and third mortgages are considered to be censored.

id	first_ time	time2	default_ time	payoff_ time	status_ time
46	25	5	1	0	1
47	25	3	0	1	2
56	25	36	0	0	0

Exhibit 11.31 Cross-Sectional Data

A competing risk analysis is also available for continuous-time hazard models. One way to compute such a model in SAS is by using PROC PHREG:

```
PROC PHREG data=lifetest;
MODEL time2*status_time(0,2)=FICO_orig_time LTV_orig_time / TIES=EFRON;
RUN;
```

The structure of PROC PHREG is very similar to PROC LOGISTIC. The dependent variable in the model is the (survival) observation time and an indicator variable, which indicates default (coded by one), or whether the borrower is no longer observed (i.e., censored, coded by zero and two). The censoring state is specified between parentheses and the two variables connected by “*”.

The model output is presented in Exhibit 11.32.

The PHREG Procedure

Model Information	
Data Set	WORK.LIFETEST
Dependent Variable	time2
Censoring Variable	status_time
Censoring Value(s)	0 2
Ties Handling	EFRON

Number of Observations Read	50000	Number of Observations Used	50000

Convergence Status
Convergence criterion (GCONV=1E−8) satisfied.

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	303232.75	301068.64
AIC	303232.75	301072.64
SBC	303232.75	301087.89

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2164.1072	2	<.0001
Score	2198.7891	2	<.0001
Wald	2183.5710	2	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
FICO_orig_time	1	−0.00442	0.0001114	1571.3114	<.0001	0.996
LTV_orig_time	1	0.01554	0.0008019	375.6040	<.0001	1.016

Exhibit 11.32 CPH Model

The computation of default and payoff probabilities is beyond the scope of this book. However, note that the cumulative incidence for the subcategories can be modeled by a methodology proposed by Fine and Gray (1999) with a slight variation of the MODEL statement.

Extensions

Other ways to control for adverse selection include bivariate probit models, estimation of the inverse Mills ratio in a first-stage probit model for the payoff event and control for the inverse

Mills ratio in a second-stage probit model for the default event, and the continued observation of paid-off borrowers (although this might be an impossible or very costly solution). For the last option, we refer to our discussion on reject inference in the credit scoring chapter.

PRACTICE QUESTIONS

1. Identify two credit exposures that have uncertain exposures at default. Identify the most suitable conversion measure for each exposure class, and justify your choice. It has been argued that regression models for conversion measures dominate models for EAD. Can you explain the reasoning for this argument?
2. Compute the conversion measures CCF, CEQ, LCF, and UACF based on the drawn loan amount one period prior to the observation of default events. Estimate a linear regression model for default events, and compute the EAD for all observations of the data set. Choose the dependent variable and covariates so that the R-squared is maximized. Use data set mortgage.
3. Compute the conversion measure LCF and estimate a Tobit model. Choose LTV_time and time to maturity (which you would have to compute using a DATA step) as covariates. Use data set mortgage.
4. Estimate a multinomial logit model for the default probability, and payoff probability, and interpret the parameters with regard to both probabilities. Include the following risk factors: FICO_orig_time, LTV_time, RE-type_CO_orig_time, REtype_PU_orig_time, and REtype_SF_orig_time. Stratify the data into a training sample and a validation sample with approximately an equal amount of observations. Compute the realized default and payoff rates and average estimated default and payoff probabilities for the validation sample by time and plot using PROC GPLOT. Use data set mortgage.

References

Agarwal, S., B. W. Ambrose, and C. Liu. 2006. “Credit Lines and Credit Utilization.” *Journal of Money, Credit and Banking*, 1–22.

Barakova, I., and H. Parthasarathy. 2012. “How Committed Are Bank Corporate Line Commitments?” Manuscript, Washington, DC: Office of the Comptroller of the Currency.

Basel Committee on Banking Supervision. 2006. “International Convergence of Capital Measurement and Capital Standards: A Revised Framework, Comprehensive Version.”

Basel Committee on Banking Supervision. 2014. “The Standardised Approach for Measuring Counterparty Credit Risk Exposures.”

Deng, Y., J. M. Quigley, and R. Van Order. 2000. “Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options.” *Econometrica* 68 (2): 275–307.

Fine, J. P., and R. J. Gray. 1999. “Proportional Hazards Model for the Subdistribution of a Competing Risk.” *Journal of the American Statistical Association* 94 (446): 496–509.

Jacobs, M. 2011. “An Empirical Study of Exposure at Default.” *Journal of Advanced Studies in Finance* 1: 31–59.

Moral, G. 2011. “EAD Estimates for Facilities with Explicit Limits.” In *The Basel II Risk Parameters*, edited by Bernd Engelmann and Robert Rauhmeier, 201–246. *Springer Science+Business Media*.

Qi, M. 2009. *Exposure at Default of Unsecured Credit Cards*. Washington, DC: Office of the Comptroller of the Currency.

Taplin, R., H. To, and J. Hee. 2007. “Modeling Exposure at Default, Credit Conversion Factors and the Basel II Accord.” *Journal of Credit Risk* 3: 75–84.

Tong, E. N., C. Mues, I. Brown, and L. C. Thomas. 2016. “Exposure at Default Models with and without the Credit Conversion Factor.” *European Journal of Operational Research* 252 (3): 910–920.

Valvonis, V. 2008. “Estimating EAD for Retail Exposures for Basel II Purposes.” *Journal of Credit Risk* 4 (1): 79–110.

Yang, B. H., and M. Tkachenko. 2012. “Modeling Exposure at Default and Loss Given Default: Empirical Approaches and Technical Implementation,” *Journal of Credit Risk 8* (2): 81.