ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Witzany, Jiří

Article

Exposure at Default Modeling with Default Intensities

European Financial and Accounting Journal

Provided in Cooperation with:
Faculty of Finance and Accounting, Prague University of Economics and Business

Suggested Citation: Witzany, Jiří (2011) : Exposure at Default Modeling with Default Intensities, European Financial and Accounting Journal, ISSN 1805-4846, University of Economics, Faculty of Finance and Accounting, Prague, Vol. 6, Iss. 4, pp. 20-48, https://doi.org/10.18267/j.efaj.18

This Version is available at:
https://hdl.handle.net/10419/109878

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Exposure at Default Modeling with Default Intensities[#]

Jiří WITZANY[*]

1 Introduction – The Concept of Exposure at Default

Basel II regulatory capital requires banks, in the advanced internal rating based approach (*IRBA*), to estimate for each credit exposure three key parameters: probability of default (*PD*), loss given default (*LGD*), and exposure at default (*EAD*). The regulatory capital formula for retail products can be expressed as $C = (UDR(PD) - PD) \cdot LGD \cdot EAD$. It is clear from the formula that the capital is as sensitive to the quality of the LGD estimates as well as to EAD estimates: 10% relative error in EAD (or LGD) leads to 10% error in the final regulatory capital, in the positive or negative direction. While PD estimation techniques, that are necessary for correct loan pricing, have been well developed many years before Basel II came into effect, banks still strive to develop sophisticated techniques to estimate the LGD and EAD parameters. There is quite limited literature on the subject (Araten – Jacobs, 2001; Moral, 2006, and Jacobs, 2008). The purpose of this study is to propose a new advanced methodology for EAD estimations incorporating not only various regression techniques but also the intensity of default modeling.

The most general EAD definitions and requirements are given in BCBS (2006). The concept is further specified and implemented in the European legislation EC (2006). Useful guidelines and interpretations can be found in CEBS (2006). According to BCBS (2006) the Exposure at Default (EAD) for an on-balance or off-balance sheet exposure is defined as the expected gross exposure of the facility upon default of the obligor. The EAD estimates are important in particular for off-balance sheet

[#] The research has been supported by the Czech Science Foundation grant no. 402/06/0890 *Credit Risk and Credit Derivatives* and by the grant no. 402/09/0732 *Market Risk and Financial Derivatives*.

[*] Doc. RNDr. Jiří Witzany, Ph.D. – assistant professor; Department of Banking and Insurance, Faculty of Finance and Accounting, University of Economics Prague, W. Churchill sq. 4, 130 67 Prague, Czech Republic; <jiri.witzany@vse.cz>.

exposures and for exposures that are composed of an on-balance sheet exposure (drawn amount) and off-balance sheet exposure (undrawn amount) like in the case of revolving credit, credit cards, and different lines of credit. EAD estimates may be theoretically challenging even for products with fixed principal and no undrawn amount due to the possibility of unpaid interest and late fees increasing the exposure or, on the other hand, repayments reducing the exposure at default compared to the current exposure. However, it follows from EC (2006) and CEBS (2006) that for those exposures it is sufficient to set EAD equal to the current gross exposure. This study will consequently focus on exposures generally composed of drawn and undrawn off-balance sheet amounts. The regulation requires EAD being estimated as the on-balance sheet exposure plus an amount reflecting the possibility of additional drawings. While BCBS (2006) does not stipulate any particular method for estimation of the expected additional drawings, EC (2006) does require banks to obtain so called Conversion Factors (*CF*) estimating the utilization of the undrawn amount upon default and calculate the Exposure at Default as

$$EAD = Current\ Exposure + CF \cdot Undrawn\ Limit\ , \tag{1}$$

The conversion factor (on a non-defaulted facility) is required to be always nonnegative. The estimation also strongly depends on the time horizon. Since PD and LGD are considered in one-year horizon, EAD should be also estimated conditional upon default in the same one-year horizon. There are several approaches, as noted by CEBS (2006), how to treat the time to default that is unknown for non-defaulted facilities, for example, the cohort approach, fixed time approach, or variable time approach.

As mentioned in CEBS (2006), some banks have traditionally expressed conversion factors out of the total credit limits not only out of the undrawn limit. We will call this factor *Credit Conversion Factor* (CCF). This method with $EAD = CCF \cdot \text{Limit}$, also called *the momentum approach*, in its simplest form does not fulfill the *Capital Adequacy Requirements* (*CAD*). However, according to CEBS (2006) the approach may be acceptable, if CCF just serves as a mean for the final CF estimation (for example, given a CCF estimation, calculate EAD and then solve the equation (1) for CF making sure that the conversion factor is nonnegative). We will also consider a generalized approach, where

EAD is estimated as a function of the current exposure, total limit, and other risk drivers, with CF recalculated according to (1).

Section 2 of this study outlines basic definitions, concepts, and data set requirements. We shall start with a full probabilistic definition of EAD that will be estimated using different methods depending also on quality and size of observed data available in the reference data set. Pool level methods will be described in Section 3 while account level estimation methods will be proposed in Section 4.

2 Definitions, Concepts, and Data

2.1 Ex-Post Exposure at Default and Conversion Factors

The ex-post EAD on a defaulted facility is defined simply as the gross exposure[1] $Ex(t_d)$ at the time of default t_d where $Ex(a,t) = Ex(t)$ denotes the on-balance sheet exposure of facility a at time t. We omit the argument a whenever it is clear from the context.

It is not so straightforward to define the ex-post conversion factor on a defaulted facility since it requires a retrospective observation point called the reference date t_r, where we observe the undrawn amount $L(t_r) - Ex(t_r)$, with $L(t)$ denoting the total credit limit at time t. Since a conversion factor measures the utilization out of the undrawn amount we need to have $L(t_r) - Ex(t_r) > 0$. Then, it makes sense to define the ex-post CF as

$$CF = CF(a,t_r) = \frac{Ex(t_d) - Ex(t_r)}{L(t_r) - Ex(t_r)}, \tag{2}$$

Note that an observed (ex-post) conversion factor may, in practice, be negative if the drawn exposure between the reference date and the default date declines, as well as larger than 1 if the exposure at default exceeds the limit effective at the reference date. This may happen if there is an increase of limit or a breach of limit, for example, caused by interest and

[1] Alternatively, the gross exposure could be split to the fees and interest and the principal amount. The principal amount drawing depends purely on the debtor's decision while fees and interest are in a sense automatic. Thus, the two components might be treated in different ways. However, in order to keep the framework simple, we are going to model only the total gross exposure development.

late fees. We will admit such observed values but the estimated ex-ante conversion factor still has to be nonnegative (regulatory requirement) and will be expected to be usually lower or equal than 1 (estimated CF larger than 1 exceptionally acceptable). Notice also that the expression (2) is very sensitive to the drawn amount if the undrawn amount is small. In case of simple average CF estimates a materiality threshold on $L(t_r) - Ex(t_r)$ should be set in order to eliminate unnecessary outliners. The materiality threshold is not needed in case of exposure-weighted or regression CF estimates described in Section 3.2.

2.2 Ex-Ante Exposure at Default and Conversion Factors

Regarding ex-ante EAD and CF we will start with a full probabilistic definition and analysis. Let τ denotes the time of default of a non-defaulted facility a at time t. Since we do not know the time of default, τ is a random variable and $\tau < \infty$ as we may assume that any debtor eventually defaults in the infinite time horizon. EAD is defined conditional upon default in the one-year horizon, hence the theoretical definition is

$$EAD = EAD(a,t) = E[Ex(\tau) \mid t < \tau \le t+1]. \tag{3}$$

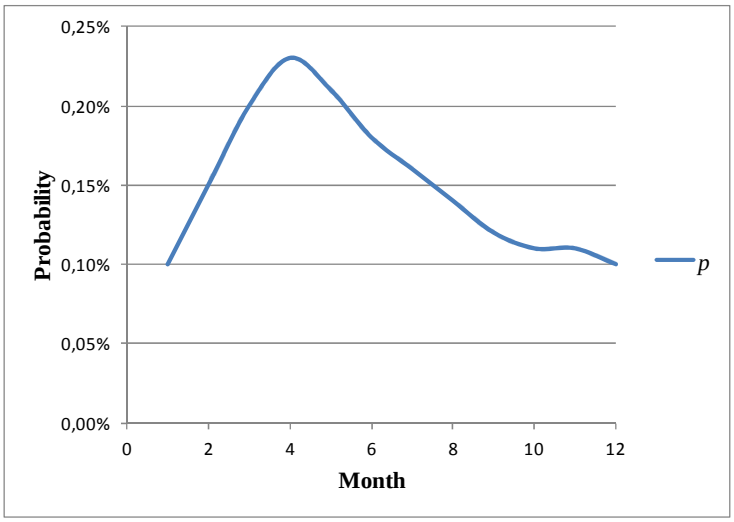
Note that $E[.]$ denotes the expected value, not the exposure where we rather use the notation $Ex(t)$. To decompose the unknown time of default and EAD conditional on the time of default we need to introduce the time to default density function $g(s)$. Here, $g(s)\Delta s$ is the probability that default happens during the time interval $[s, s+\Delta s]$. Consequently, EAD can be expressed as the $g(s)ds$ weighted average of the expected exposure upon default at $\tau = s$:

$$EAD = EAD(a,t) = \frac{\int_t^{t+1} E[Ex(\tau) \mid \tau = s] g(s) ds}{P[t < \tau \le t+1]}. \tag{4}$$

Thus, according to the analysis ex ante EAD does also depend on the probability distribution (density function) of the time to default. In particular, for short term retail loans, according to empirical experience, the time to default density function is large shortly after drawing and later significantly declines (see Fig. 1 for an illustration). This dependence

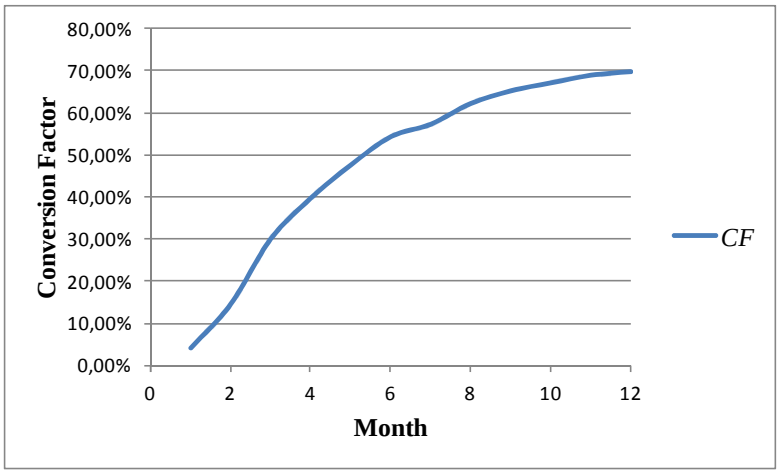
typically exists if we observe exposures homogenous in terms of time in bank, e.g. new credit cards or mortgages after a fixed number of years etc. If the portfolio is mixed with respect to the time in bank then the dependence usually disappears or becomes insignificant.

Fig. 1: Intensity of time to default from the first withdrawal – Credit Cards



The distribution of the time to default depends on a particular product, as well as on the time from the facility origination. Note that the same approach could be theoretically applied to LGD. However, the empirical experience is that there is no significant dependence of ex-ante LGD on the time to default in the one-year horizon, while there is a significant dependence of average observed CF on the time to default $t_d - t_r$ (see Fig. 2 for an illustration). This is confirmed for example by the study of Araten and Jacobs (2001). Consequently we will use the definition (4) which can be also called *PD-weighted approach.*

Fig. 2: Conversion factors and time to default dependence



In practice, we need to get an estimation $\widehat{EAD}$ of EAD defined according to (3) or (4). The hat notation will be sometimes omitted but we need to keep in mind that there are three different *EADs* or *CFs*: those calculated ex-post from the historical data, then the theoretical and unknown ex-ante values (parameters of a probability distribution without a hat), and their estimations which depend on the estimation method chosen (with a hat).

For example, the integral (4) can be approximated by a discrete summation: Let us split the one-year time interval into a sequence of subintervals $(t_0, t_1], \ldots, (t_{n-1}, t_n]$ where $0 = t_0 < t_1 < \cdots < t_n = 1$. Next we estimate $\widehat{EAD}_i$ conditional on time of default τ being in the interval $(t_{i-1}, t_i]$ and the probability $\hat{p}_i$ that default happens during this interval for $i = 1, \ldots, n$. Consequently, $\hat{p} = \sum_{i=1}^{n} \hat{p}_i$ estimates the probability of default within one year. Then, in line with (4), we get the approximation

$$\widehat{EAD} = \frac{1}{\hat{p}} \sum_{i=1}^{n} \hat{p}_i \cdot \widehat{EAD}_i \,. \tag{5}$$

In order to obtain the conditional $\widehat{EAD}_i$ estimations we must to split our observed data according to different distances between the reference date and the default date. The subintervals may have equal length of, for example, one or three months, or we can use an irregular splitting depending on the sensitivity of EAD on the time to default. This approach is clearly applicable if there is an approximation of the time to default density function. (The estimates $\hat{p}_i$ may be obtained e.g. observing a portfolio of non-defaulted accounts with certain characteristics at time T and counting the number of defaults in the interval $(T+t_{i-1}, T+t_i]$.)

Alternatively we could estimate the average time to default $\overline{\tau} = \frac{1}{\hat{p}} \sum_{i=1}^{n} \frac{t_{i-1}+t_i}{2} \hat{p}_i$ conditional on $\tau \le 1$ and set

$$\widehat{EAD} = \widehat{EAD}_{i_0} . \tag{6}$$

where i_0 is the first index i such that $\overline{\tau} \in [t_{i-1}, t_i]$. Such estimation should be better than, for example, one-year to default fixed time horizon estimation but its quality strongly depends on the distribution of the time to default and on the dependence of EAD on the time to default. Since distribution of the time to default varies across different products and facilities it is clear that (5) provides much more precise estimation compared to (6).

Similar approach can be applied to conversion factor estimation since

$$CF(a,t) = E\left[\frac{Ex(\tau) - Ex(t)}{L(t) - Ex(t)} \middle| t < \tau \le t+1 \right] = \frac{EAD - Ex(t)}{L(t) - Ex(t)} . \tag{7}$$

as $L(t)$ and $Ex(t)$, the current limit and the current exposure, are known at the reference time and can be taken out of the expectation operator. Combining (7) and (4) we obtain

$$CF = CF(a,t) = \frac{\int_t^{t+1} E[CF \mid \tau = s] g(s) ds}{P[t < \tau \le t+1]} . \tag{8}$$

Consequently, if $\widehat{CF}_i$ are estimates conditional on time to default being in the interval $(t_{i-1}, t_i]$ we may again use the PD-weighted average

$$\widehat{CF} = \frac{1}{\hat{p}} \sum_{i=1}^{n} \hat{p}_i \cdot \widehat{CF}_i \ . \tag{9}$$

2.3 Reference Data Set (RDS)

Reference data set is a set of historical observations used for ex-ante *EAD* estimations. Our notation follows Moral (2006). An observation $o = (a, t_r, t_d, \overrightarrow{RD})$ consists of defaulted facility identification, the reference date, the date of default, and a vector of risk drivers containing at least the information on exposures and limits at the reference and default dates ($Ex(t_r), L(t_r), Ex(t_d), L(t_d)$). Other risk drivers might capture the information on *qualitative risk drivers* as the facility type, customer type, rating class at reference date, or average rating during a period preceding the reference date, status of the facility (e.g. output of an Early Warning System), collateralization and third person guarantees, covenants (more appropriate corporate borrowers); and *quantitative risk drivers* like time in bank, time to maturity, expected LGD which could be a parameter aggregating a number of the other explanatory variables, etc. It is not necessary to record macroeconomic risk drivers on the account level as those depend only on the observation date and can be kept in a separate table.

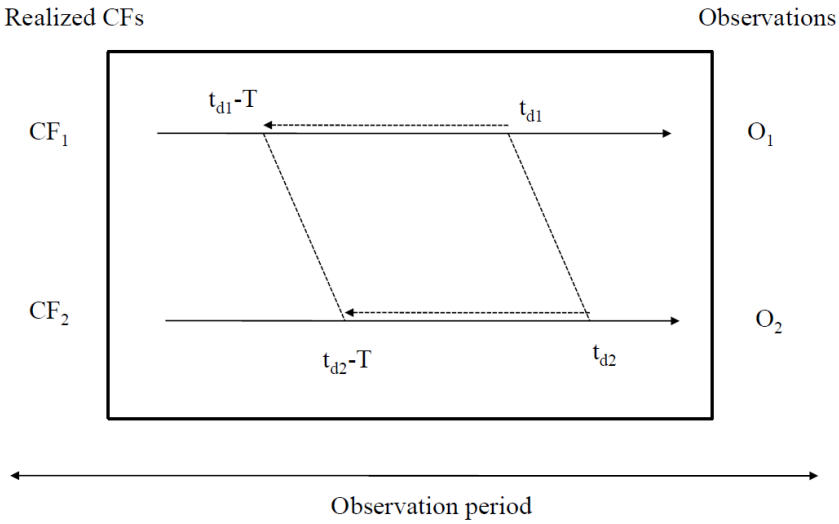
RDS should be created separately at least for different products, e.g. credit cards, overdrafts, lines of credit, etc. In case of lack of data the data sets could be possibly unified. Such an approach should be however exceptional due to possible different development of drawings before default for different products, for example, due to various controls and restrictions imposed by the bank. On the other hand, in the pooling approach with sufficient historical database product level RDS should be split to a number of subsets capturing certain risk drivers, e.g. facility status or macroeconomic situation. Similarly, to apply a time-series analysis approach the RDS needs to be further split into cohorts according to time of default or the reference date. The splitting of RDS is possible only as long as there are enough observations in the resulting pool level reference data sets. In order to calculate meaningful ex-post conversion factors we should require, depending on the estimation method employed,

that $L(t_r) - Ex(t_r)$ is larger than certain reasonable threshold. In other words the observations where the undrawn amounts fall below the threshold should be removed from the RDS. The threshold is to be applied only in case of default weighted average calculation (see Section 3.2) and is not necessary for the other estimation methods, where the observations are in principle weighted by the undrawn amount.

As explained in the definition of ex-post EAD and CF a single observation is not determined only by the facility that defaulted at time t_d but also by the reference date t_r at which we measure the retrospective drawn and undrawn amount. We do not exclude the possibility of more than one reference date for a given single defaulted facility in order to capture the dependence of EAD and CF on the time to default. The most common choice (and the most conservative in line with the analysis above) is the one year horizon corresponding to the unexpected credit losses estimation horizon, however, there are different alternatives (see also Moral, 2006): Fixed Time Horizon, Cohort Approach, or Variable Time Approach.

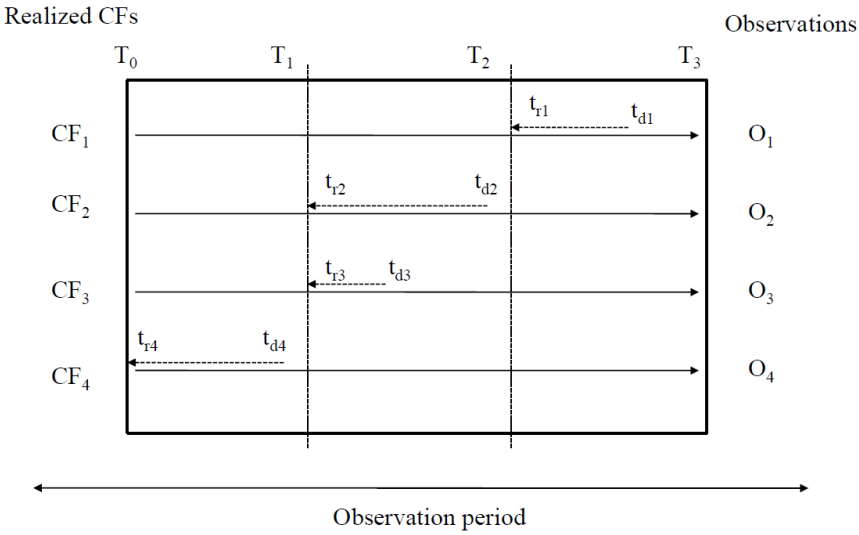
Fixed Time Horizon Approach sets $t_r = t_d - T$, where T is a fixed time horizon (see Fig. 3). RDS defined in this way in fact leads to an estimation of *EAD* and *CF* conditional on the time to default being equal exactly to T. Hence, a number of RDS with different fixed time horizons and based on the same set of defaulted facilities may be constructed in the *PD*-weighted approach. Nevertheless, banks often use $T = 1$ year as a standard choice. As explained above, the weighted time to default appears to be better, if just one fixed time horizon is to be used.

Fig. 3: Fixed Time Horizon approach



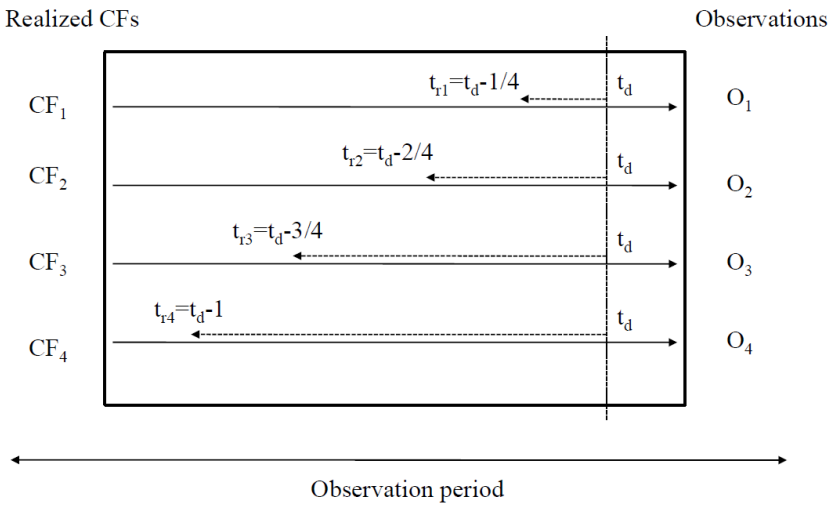
Cohort Method divides the observation period into intervals $(T_0, T_1], \ldots (T_{n-1}, T_n]$ of a fixed length, typically 1 year (see Fig. 4). Defaulted facilities are grouped into cohorts according to the default date. The reference date of an observation is defined as the starting point of the corresponding time interval. I.e. if $t_d \in (T_i, T_{i+1}]$ then we set $t_r = T_i$. In this case, the time to default probability distribution is implicitly captured in the data. However, the beginnings of intervals may cause a significant seasoning bias (for example T_i some time before Christmas will probably show higher drawing on credit cards or overdrafts than during the other months). Hence, it is advisable to set T_i at "normal" periods of the year with average drawings.

Fig. 4: Cohort approach



Variable Time Horizon Approach uses a range of fixed horizons $T_1,...,T_k$, e.g. one to twelve months, or 3, 6 , 9, and 12 months (see Fig. 5). For each observation we calculate realized conversion factors for the set of reference dates $t_r = t_d - T_i, i = 1,...,k$. The difference compared to the fixed horizon approach is that we put all the observations $(a, t_d - T_i, t_d, ...)$ into one reference data set (RDS). In the fixed horizon approach we admit different time horizons only in different reference data sets used for conditional *EAD* estimation. When all the observations are put into one RDS there might be a problem with homogeneity, for example, the facilities that have been already marked as risky with restrictions on further drawing should be treated separately. Moreover, there is an issue of high correlation of the different observations obtained from one defaulted account. The RDS on the other hand captures implicitly the possible dependence of *EAD* and *CF* on the time to default, but the distribution of the time of default (appearing flat in the *RDS*) is not realistically captured. This dataset is not suitable in the context of the *PD*-weighted approach by definition.

Fig. 5: Variable time horizon approach



The broadest RDS must contain all the observations of facilities for a given product type that have defaulted over the observation period. The length of the period must be in line with the regulatory requirement at least 5 years (or 2 years according to EC, 2006). If we interpret the requirement in the sense that RDS is based on all accounts defaulted during the last 5 years then, in fact, we need data starting 6 years ago since the reference dates are set up to one year before the default dates. The time period should optimally cover the full economic cycle according to EC (2006).

To summarize we recommend the fixed-time horizon approach for the *PD*-weighted approach (different time horizons for different RDS). Otherwise we prefer the cohort method unless the drawings show strong seasonality. In that case we recommend the variable-time horizon approach.

2.4 Empirical Example

We have randomly generated a number of defaulted accounts and calculated the corresponding observed conversion factors 1 to 12 months prior to default with dependence approximately corresponding to Fig. 2. Tab. 1 shows the average Conversion Factors depending on the time to

default. At the same time we assume that the density of default (given for months 1 to 12 in Tab. 1) has approximately the pattern given by Fig. 1.

Tab. 1: Conversion factors depending on time to default and the intensity of default

Month i	***CF***	***p***	$CF_i \cdot p_i$
1	4,14%	0,10%	0,23%
2	14,61%	0,15%	1,21%
3	30,10%	0,20%	3,33%
4	39,79%	0,23%	5,06%
5	47,71%	0,21%	5,54%
6	**54,41%**	0,18%	5,41%
7	57,40%	0,16%	5,07%
8	62,32%	0,14%	4,82%
9	65,32%	0,12%	4,33%
10	67,21%	0,11%	4,08%
11	69,02%	0,11%	4,19%
12	**69,90%**	0,10%	3,86%
Total		1,81%	47,13%

Equal weighted CF	**48,49%**
PD - weighted CF	**47,13%**

The 12 months fixed horizon approach gives the estimate $\widehat{CF} = CF_{12} = 69.9\%$. The variable time approach effectively yields the average $\widehat{CF} = 48.49\%$ of the 12 values and a similar result could be expected in the cohort approach depending on distribution of default in the cohort intervals. The PD-weighted approach according to (9), on the other hand, gives $\widehat{CF} = 47.14\%$. In the simplified approach we may firstly calculate the average time to default $\bar{\tau} = 5.54 \cong 6$ months and set $\widehat{CF} = CF_6 = 54.4\%$ according to (6).

The message of this exercise is that the CF estimate strongly depends on the method chosen. The 12 months fixed time horizon being clearly the most conservative while the PD-weighted and variable time estimates come out much lower and relatively close. The two values may, however, differ more significantly depending on the conditional CF and density of

default functions. The PD-weighed CF outlined in Section 2.2 presents, in our view, the best estimates from the theoretical point of view.

3 Pool Level Estimations

3.1 Definition of Pools and the Concept of Pool level Estimations

In the pool level approach defaulted and non-defaulted receivables are classified into a number of disjoint pools, that are homogenous with respect to selected risk drivers, and which contain at the same time sufficient amount of historical observations. Specifically, we determine certain defining properties ϕ_l, $l = 1,...,m$ and set $RDS(l) = \{o \in RDS \mid \varphi_l(o)\}$ where RDS is the broadest reference data set. By the pool l we understand not only $RDS(l)$ but also the set of all non-defaulted facilities satisfying ϕ_l. Consequently the defining properties may use only the information known at the reference date, in particular not the time to default $t_d - t_r$ which is known for defaulted but not for non-defaulted facilities (unless our estimation is conditional upon the time to default in the PD-weighted approach). Each $RDS(l)$ is used to obtain an estimation of the conversion factor $\widehat{CF}(l)$. Then, for a non-defaulted facility a we find the unique class (pool) l, so that a satisfies ϕ_l, and (in the basic approach) set

$$\widehat{EAD}(a,t) = Ex(a,t) + (L(a,t) - Ex(a,t)) \cdot \widehat{CF}(l). \tag{10}$$

Although $\widehat{CF}(l)$ is a pool level estimate (same for all non-defaulted facilities belonging to the pool l) the estimation $\widehat{EAD}(a,t)$ is, in fact, account specific as it uses the actual account exposure and limit. It could be also noted, that pool level *CCF* estimations (not allowed by EC (2006) in the simplest form) approach indeed generally lead to different account level *EAD* estimations, since $\widehat{EAD}(a,t) = \widehat{CCF} \cdot L(a,t)$ does not depend on the actual exposure $Ex(a,t)$.

The pool level estimations could be further improved using the PD-weighted approach applying account level distribution of the time to default: $RDS(l)$ needs to be split into a number of smaller sets $RDS_i(l)$

according to the time to default. For each of the reference data sets we obtain an estimation $\widehat{CF}_i(l)$ of the conversion factor and calculate $\widehat{CF}(a)$ according to (9). The probabilities of default $\hat{p}_i = \hat{p}_i(a)$ should depend on the obligor and facility rating, and on the time on books. This approach combining efficiently account specific information and pool level estimations shows, that there is no sharp borderline between so called *pool-level* and *account-level* estimations.

The definition of pools is based either on expert criteria, e.g. just according to facility (behavior) rating, or on an advanced technique using regression trees or EAD rating, e.g. based on the logistic regression of low and high drawings at default. EAD rating could be also a secondary product of account-level EAD estimates. The definition of pools must take into account the requirement that the pool level data sets $RDS(l)$ need to remain sufficiently large in terms of the number of observations. The same requirement applies, when we split $RDS(l)$ into the data sets $RDS_i(l)$ according to the time to default (though here we may produce more observations for each defaulted facility with different retrospective time horizons) as described above, or to cohort sets $RDS_v(l)$ according to the time period in which the observation appeared. The cohort estimation analysis will be described in Section 3.3 on margins of conservatism and time series analyses. It is clear that a very rich initial data set would be needed, if we wanted to combine the cohort time series analysis with the PD-weighted approach, effectively splitting the initial data sets in three dimensions into $RDS_{v,i}(l)$.

3.2 Pool Level Estimations of CF and EAD

Although EC (2006) requires banks to obtain primarily estimates of *CF*, it should be underscored that the final aim is to get estimations of the parameter *EAD* that enters the regulatory capital formula. Hence, quality of different estimation methods should be judged using a goodness of fit measure of the distance between the observed *EADs* (not *CFs*) and the corresponding ex-ante estimations. The standard measure defined as the sum of squared errors naturally leads to an estimation of *CF* being equal to a coefficient in a regression equation for *EAD*. This formula for *CF* can be interpreted as the mean value weighted by squared undrawn amounts. We list below some other formulas used by the banking industry.

Furthermore we propose a generalized *EAD* regression approach, where the coefficients are constant on a pool level, but *CFs* must be recalculated on the account level in line with our introductory remarks in Section 1.

In this subsection we consider a reference data set $RDS(l)$, which could be either the broadest product level data set, or the one resulting from subdivision according to certain pooling criteria $\phi_l, l = 1,...,m$, and/or from the cohort approach, and/or from the time-to-default conditional subdivision approach (we omit the possible sub-indices v and i).

Given a reference data set with calculated ex-post conversion factors $CF(o), o \in RDS$ the simplest approach is to calculate **the sample (default-weighted) mean:**

$$\widehat{CF}(l) = \frac{1}{|RDS(l)|} \sum_{o \in RDS(l)} CF(o). \tag{11}$$

The same weight is assigned to each observation disregarding the magnitude of undrawn amount or time of the observation. In particular, the observations with very low undrawn amounts might bring a significant random error into the estimation. This problem is in general solved by the **weighted mean** approach:

$$\widehat{CF}(l) = \frac{\sum w_o \cdot CF(o)}{\sum w_o}. \tag{12}$$

where w_o are appropriate positive weights (omitting the scope of summation $o \in RDS(l)$ for simplicity). The natural candidates for the weights are the undrawn limit amounts $w_o = L(o) - Ex(o)$. Then we get

$$\widehat{CF}(l) = \frac{\sum (EAD(o) - Ex(o))}{\sum (L(o) - Ex(o))}. \tag{13}$$

The weights could also reflect the time of observations assigning lower rates to older observations and higher rates to recent observations. Note that the standard approach according to BCBS (2006) is the default weighted one with no time dependence, however "*a credit institution*

need not give equal importance to historic data if it can demonstrate to its competent authority that more recent data is a better predictor of draw downs" according to EC (2006).

As outlined in the introduction, we prefer starting with the standard goodness of fit measure

$$GF = \sum_{o \in RDS(l)} (EAD(o) - \widehat{EAD}(o))^2 . \tag{14}$$

In other words we are looking for estimation methods producing ex-ante *EAD* estimates that minimize the sum of absolute squared differences between the realized *EADs* and the ex-ante predictions. If we restrict ourselves to estimations of the form

$$\widehat{EAD}(o) = Ex(o) + \widehat{CF}(o) \cdot (L(o) - Ex(o))$$

then we need to minimize

$$GF = \sum \left(EAD(a) - Ex(o) + \widehat{CF}(o) \cdot (L(o) - Ex(o)) \right)^2 . \tag{15}$$

which is equivalent to the **regression of the absolute increase without constant**:

$$EAD(o) - Ex(o) = \alpha + \beta (L(o) - Ex(o)) + \varepsilon \text{ with } \alpha = 0 \text{ and } \beta = CF . \tag{16}$$

Consequently

$$\widehat{CF}(l) = \frac{\sum (EAD(o) - Ex(o)) \cdot (L(o) - Ex(o))}{\sum (L(o) - Ex(o))^2} . \tag{17}$$

Note that this formula corresponds to weighted mean approach (12) with $w_o = (L(o) - Ex(o))^2$. We recommend using the formula (17) as the most consistent pool level *CF* estimation approach.

Alternatively we may apply regression of the **relative increase of exposure** $ead(o)-ex(o)$ in terms of $1-e(o)$, where $ead(o)=\frac{EAD(o)}{L(o)}$ and $ex(o)=\frac{Ex(o)}{L(o)}$. Hence, we scale the observations by the total credit limit and solve the regression equation $ead(o)-ex(o)=\alpha+\beta(1-ex(o))+\epsilon$ with the condition $\alpha=0$ and $\beta=CF$. Note that the goodness of fit in this case

$$GF=\sum(ead(o)-\widehat{ead}(o))^2=\sum\left(\frac{1}{L(o)}\right)^2(EAD(o)-\widehat{EAD}(o))^2$$

differs from (14) and so the result of the regression is

$$\widehat{CF}(l)=\frac{\sum(ead(o)-ex(o))(1-ex(o))}{\sum(1-ex(o))^2}. \tag{18}$$

The approach may be appropriate for reference data sets where we assign the same importance to observations with relatively low total limit as to observations with relatively high limit.

Since our goodness of fit measure (14) is focused on *EAD* rather than *CF* estimations the following generalized approach can be considered: express ex-ante $\widehat{EAD}(o)=\beta_1\cdot Ex(o)+\beta_2\cdot L(o)$ as a linear combination of the current exposure and the total limit and find the pool-level coefficients β_1 and β_2 minimizing the goodness of fit measure (14). In other words we regress

$$EAD=\beta_0+\beta_1\cdot Ex+\beta_2\cdot L+\epsilon \text{ with the condition } \beta_0=0. \tag{19}$$

It is clear that we generally get a better result in terms of goodness of fit since we have one additional explanatory variable compared to the regression approach based only on the undrawn amount. Note that this would be equivalent to the one parameter regression (16) if we assumed that $\beta_1+\beta_2=1$. In order to satisfy the regulatory requirement we may recalculate account-specific conversion factors

$$\widehat{CF}(o) = \max\left(0, \frac{\beta_1 \cdot Ex(o) + \beta_2 \cdot L(o) - Ex(o)}{L(o) - Ex(o)}\right). \tag{20}$$

We must use the maximum operator, since the conversion factor must be nonnegative due to the regulatory conditions. This may introduce a conservative bias into the final estimate $\widehat{EAD}(o) = Ex(o) + \widehat{CF}(o) \cdot (Ex(o) - L(o))$, but the recalculated goodness of fit measure (14) might still provide a better result than the pure *CF* approach according to (17).

3.3 Margin of Conservatism

The estimation techniques described so far provide, in line with the definition (3), the expected value of *EAD* or *CF*. The regulation (e.g. BCBS, 2006, Art. 475) in addition requires a margin of conservatism appropriate to the likely range of errors in the estimate, positive PD x EAD correlation, or downturn economic conditions.

The margin of conservatism may be based either on a time series analysis of cohort level CF estimates, or on an analysis of the CF distribution, in case there are not enough data to obtain cohort CF estimates.

Assume first that CF is obtained from (17) as a regression coefficient of (16), i.e. as the squared undrawn amount average of the ex post conversion factors. First of all a margin of conservatism set equal to the standard error of the regression coefficient (or its multiple) related to the estimation error should be added.

$$se(\widehat{CF}) = \left(\frac{\sum (EAD(o) - Ex(o) - \widehat{CF}(o) \cdot (L(o) - Ex(o)))^2}{(|RDS(l)| - 1) \cdot \sum (L(o) - Ex(o))^2}\right)^{1/2} \tag{21}$$

Note that the estimation error may be significant when the reference data set is small, while it diminishes when the RDS is large. Secondly, we want to add a margin of conservatism related to the systematic risk when CFs could be on a portfolio level larger than the long term average value. Let us calculate the average deviation of the observed values from the average with the squared undrawn amounts weights:

$$\hat{\sigma} = \sqrt{\frac{\sum_o \left(L(o) - Ex(o)\right)^2 \left(CF(o) - \widehat{CF}(o)\right)^2}{\sum_o (L(o) - Ex(o))^2}} \tag{22}$$

For other averaging techniques described in the previous sections the corresponding weights need to be applied. The average deviation might be used to obtain quantiles of the parameter *CF* accepting the (simplifying) assumption that it is normally distributed. For example, 95% percentile could be estimated as $\widehat{CF} + \hat{\sigma} \cdot N^{-1}(0.95)$ where N^{-1} is the inverse standardized cumulative normal distribution. However this is an account level stressed value, while the logic of the regulatory formula is to stress portfolio level average values. In other words $\widehat{CF} + \hat{\sigma} \cdot N^{-1}(0.95)$ is an estimate that (or worse) can be observed on a single defaulted account with 95% probability. But we rather need a 95%-probability stressed value that could be observed on average over a large portfolio of defaulted accounts. The transformation from the account level standard deviation to large portfolio level (asymptotic) standard deviation based on a uniform correlation ρ can be easily done using the normality assumption: if $X_i, i = 1, ..., N$ are normal random variables with mean μ, standard deviation σ, and with uniform mutual correlation ρ then it is easy to show that the standard deviation of the average $\frac{1}{N}\sum X_i$ tends to $\sigma\sqrt{\rho}$ when N is large. Hence if we estimate that the CF account level standard deviation is $\hat{\sigma}$ then a large portfolio average CF standard deviation is $\hat{\sigma}\sqrt{\rho}$ provided the mutual correlation is a positive constant ρ.

Finally our conservative estimation including the estimation error can be expressed as

$$\widehat{CF}_c = \max(\widehat{CF}, 0) + (se(\widehat{CF}) + \hat{\sigma} \cdot \sqrt{\rho}) \cdot N^{-1}(0.95)\,. \tag{23}$$

where we suggest to use regulatory correlation values used for unexpected PD modeling, e.g. $\rho = 0.04$ for revolving exposures, if an EAD specific estimate is not available. We use the standard 95% quantile corresponding to the worst year in every twenty years. If the observed

time period covers good and bad years, years with high and low *PD*, then the estimation (23) captures not only the estimation error, but also possible systematic variation due to economic downturn or high *PD* periods. If the observed period does not cover such years, then an additional conservative adjustment based on expert judgment or external data should be added.

If there are sufficient data to produce cohort level conversion factor $\widehat{CF}_v$ estimates we may apply a time series analysis. The approach could test the sensitivity of $\widehat{CF}_v$ with respect to macroeconomic variables or PD and separate the systematic factors influence from the estimation error. However, unless explicitly required by regulator, we propose to use the relatively simple and efficient formula (23).

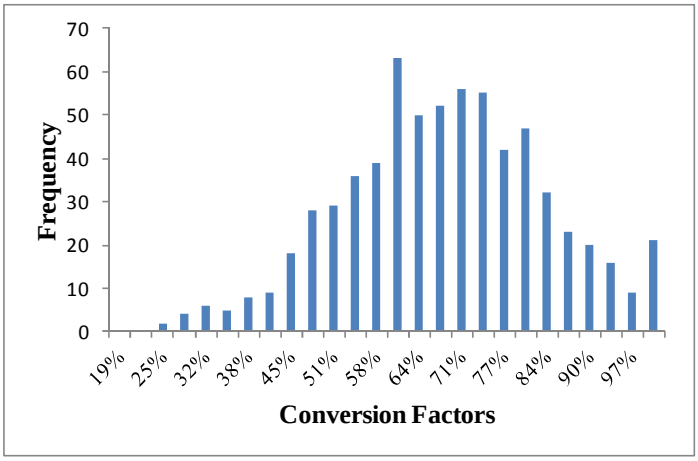
Alternatively, the regression (16) might be run with a different minimization function (Moral, 2006) that assigns a larger weight to positive estimation errors (underestimation of the real *EAD* by $\widehat{EAD}$ which is not desirable from the regulatory perspective), e.g.

$$\sum (a \cdot \max(EAD(o) - \widehat{EAD}(o), 0) + b \cdot \max(\widehat{EAD}(o) - EAD(o), 0)) \qquad (24)$$

where $\widehat{EAD}(o) = Ex(o) + CF \cdot (L(o) - Ex(o))$ and $a > b$ for example $a = 0.95$ and $b = 0.05$. The regression then yields the distribution $(b/(a+b))$ – quantile rather than the expected value estimate.

Example: We have randomly generated 620 defaults (of e.g. credit cards). The credit limits have been between 10 000 and 50 000 (e.g. CZK) and the drawn amount between 10% and 50% of the limit. The distribution of the ex post conversion factors in a fixed horizon (12 months) is shown on Fig. 6. We have used the data set to estimate and compare the ex ante conversion factor for the product applying the methods described above.

Fig. 6: Histogram of the ex post conversion factors



First we apply the simple average (11) and get $\widehat{CF_1} = 67.29\%$. Then we try the undrawn amount weighted mean (13) to obtain a slightly lower value $\widehat{CF_2} = 64.42\%$. Next, we employ the regression based technique (17) to get $\widehat{CF_3} = 62.27\%$. Note that this formula is equivalent to the squared undrawn amount weighted approach. Hence, the lower estimates indicate that the realized conversion factors are lower for higher undrawn exposures. Finally, when we apply the percentage increase of exposure regression based formula (18), we obtain $\widehat{CF_4} = 67.41\%$. The higher value may be explained by the fact that in this approach there is no difference between accounts with high and low limits.

Let us check that the sum of squared errors goodness of fit measure (14) comes out the best for the third (regression based) estimation. Instead of GF we may equivalently calculate the classical R^2 expressed as

$$R^2 = 1 - \frac{\sum_{o \in RDS(l)} (EAD(o) - \widehat{EAD(o)})^2}{\sum_{o \in RDS(l)} (EAD(o) - \overline{EAD})^2}.$$

The R^2 for $\widehat{CF}_1, \widehat{CF}_2, \widehat{CF}_3, \widehat{CF}_4$ came out as $77.33\%, 79.07\%, 79.46\%, 77.22\%$ respectively. Not surprisingly, R^2 comes out maximal for $\widehat{CF}_3$ as this estimate maximizes the measure by definition.

Finally, let us calculate, based on $\widehat{CF}_3$, the standard error according to (21), the average deviation according to (22), and the conservative CF estimation (23). We have obtained: $se(\widehat{CF}) = 0.62\%$, $\hat{\sigma} = 16.4\%$, and so $\widehat{CF}_c = 62.27\% + (0.62\% + 16.4\% \cdot 0.2) \cdot 1.65 \cong 68.7\%$, where the total margin of conservatism is 6.43%.

4 Advanced Methods – Conditional and Account Level Estimations

As pointed out in the previous section "the pool level techniques" described can be from certain perspective considered to be account-level: the parameter CF or β_1 and β_2 from (19) are estimated on a pool but the final EAD estimate is calculated using account specific information on the exposure and undrawn amount. If the PD-weighted approach is moreover applied, then we are also using account specific information to determine the probability distribution of the time to default. This section aims to describe regression techniques, where we estimate already the coefficients CF or β_1 and β_2 as functions of account specific explanatory variables with values known for non-defaulted facilities. We may also add the time to default as an additional explanatory variable (which is known ex-post but not ex-ante) and apply the PD-weighted approach.

4.1 Regression with CF in the form of the Logit Function

In this approach, we use again the regression equation (16), but with CF expressed in terms of the other explanatory variables (macroeconomic, facility, or obligor level risk drivers). Since all the relevant risk drivers become explanatory variables, we keep the broad product level reference dataset which is not split to smaller pool level datasets. Qualitative variables are categorized or represented by dummy variables using standard techniques (the regression could be equivalently performed in separate pools determined by the qualitative variables, but one regression is certainly more convenient). The parameter CF can be

modeled in different parametric forms. The simplest linear form would be $CF = \mathbf{b}'\mathbf{f}$ where $\mathbf{f}$ is a vector of relevant risk drivers and $\mathbf{b}$ is a vector of linear regression coefficients. Alternatively, we may use a link function e.g. the exponential $CF = e^{-\mathbf{b}'\mathbf{f}}$ where the outcome is always positive, but may be also larger than 1. If the historical data confirm that $CF \in [0,1]$ then the logit function would be more appropriate:

$$CF = \Lambda(\mathbf{b}'\mathbf{f}) = \frac{e^{\mathbf{b}'\mathbf{f}}}{1+e^{\mathbf{b}'\mathbf{f}}} . \tag{25}$$

The coefficients are obtained numerically minimizing either the sum of squared errors (15) or using the maximum likelihood approach (see Section 4.2).

If a is a non defaulted account with actual risk drivers $\mathbf{f}(a)$ then

$$\widehat{EAD}(a) = Ex(a) + \widehat{CF}(a)\cdot(L(a) - Ex(a))$$

where $\widehat{CF}(a) = \Lambda(\hat{\mathbf{b}}'\mathbf{f}(a))$ is our ex-ante account level estimate of expected exposure and conversion factor at default. If the risk drivers include time to default then we must use the PD-weighted average (9) to calculate $\widehat{CF}(a)$ where $\widehat{CF}_i(a)$ are the logistic-link regression estimated conversion factors with actual risk drivers of a conditioned on different times to default and $\hat{p}_i = \hat{p}_i(a)$ account specific estimates of probabilities of default for different time bands. In both cases, according to the regulatory requirement, we need to add a margin of conservatism. If the regression analysis confirms a significant dependence on macroeconomic variables (or experienced *PDs*) then those variables should be firstly stressed obtaining $\mathbf{f}_s(a)$ representing downturn economic condition, and then setting $\widehat{CF_s}(a) = \Lambda(\hat{\mathbf{b}}'\mathbf{f}_s(a))$. Alternatively, as in the pool level approach, we calculate the standard error $se(\widehat{CF})$ according to (21) and $\hat{\sigma}$ according to (22) but with $\widehat{CF} = \widehat{CF}(o)$ depending on the risk drivers. In the case of PD-weighted approach we take the PD-weighted average of the corresponding errors. The final conservative estimate then should be calculated according to the equation (23), i.e.

$$\widehat{CF}_c = \max(\widehat{CF}, 0) + (se(\widehat{CF}) + \hat{\sigma}\cdot\sqrt{\rho})\cdot N^{-1}(0.95) .$$

4.2 Beta regression

The proposed regression (25) will be statistically more consistent when we use an appropriate likelihood function. Let us assume that the relative account level $ead = \frac{EAD}{L}$ has a beta distribution with minimum 0 and maximum 1. See e.g. Smithson – Verkuilen (2005) for a detailed description of the beta distribution and the regression technique. Since ead is our targeted estimate, we recommend to use the log likelihood function expressed as follows

$$l(\mathbf{b},\phi) = \sum \ln Beta(ead(o), l(o) + \Lambda(\mathbf{b}'\mathbf{f}(o))(e(o) - l(o)), \phi) .$$

The beta distribution density function $Beta(y,\mu,\phi)$ is here parameterized by the mean μ and the precision parameter ϕ. While the mean is expressed as the logit transformation of a linear combination of the risk factors we propose ϕ to be regressed as a constant.

4.3 EAD regression

The regression above was based on the functional form $ead = e + \beta(1-e)$ with $\beta = CF = \Lambda(\mathbf{b}'\mathbf{f})$ or in a simpler parametric form. As noted in Section 2 we do not need to stick to this form in the account level approach as any account level *EAD* estimate can be mapped to a *CF* estimate and vice versa. For example the momentum (*CCF*) approach where we assume that EAD depends only on the limit would be given by the simple equation $ead = \alpha$ where $\alpha = \Lambda(\mathbf{a}'\mathbf{f})$ could be again regressed as the logit transformation of a linear combination of the risk drivers. In general, we could argue, as in the previous section, that EAD depends partially on the total limit and partially on the undrawn amount and regress $ead = \alpha + \beta(1-e)$ where $\alpha = \Lambda(\mathbf{a}'\mathbf{f})$ and $\beta = \Lambda(\mathbf{b}'\mathbf{f})$. To obtain the conversion factor estimate for a non-defaulted account a in line with the regulatory requirements, we firstly get $\widehat{EAD}(a) = \hat{\alpha}L(a) + \hat{\beta}(L(a) - E(a))$ and recalculate $\widehat{CF}(a)$ analogously to (20). The margin of conservatism can be obtained as above, stressing the macroeconomic risk drivers in $\mathbf{f}(a)$ and adding the margin of conservatism factor according to (23).

4.4 EAD (CF) rating – regression trees

Account level *EAD*, similarly to *LGD*, can be estimated in a one-step or two-step procedure. One-step estimation means direct regression estimation as described above. In a two-step procedure we firstly assign to a given account a rating class via an account-level estimate, and then obtain an *EAD* estimation (using the pool-level techniques) given by the rating. Hence the EAD rating approach is a combination of account-level and pool-level techniques.

The one-step account level estimation of *CF* may be used for the rating determination (e.g. according to CF intervals 0-10%,…, 90-100%). Conversion factors would be then re-estimated on the rating pools. Another approach would be to use the regression tree technique approach. If the realized conversion factors are distributed into low and high values logistic regression could alternatively be tested.

Conclusions

We have proposed a number of techniques to estimate the EAD parameter as required by the Basel II regulation. Applicability of the techniques depends on availability of data and in particular on availability of the intensity of default estimates. If those are not in hand then we propose to use the variable time RDS approach which implicitly captures the dependence of EAD on the time to default. The results of pool level and account level regression should be compared in terms of stability and estimation errors. If the intensity of default estimates is available then we recommend to use multiple RDS with different fixed time horizons to produce either pool level or regression EAD estimates conditional on the time to default. Finally a margin of conservatism capturing the estimation error and systematic factors related to potential downturn economic conditions must be added.

Our numerical examples have shown that the results may depend significantly on the method chosen. We have made a number of recommendations based rather on a qualitative analysis. However, additional empirical research comparing the different approaches and based on real banking data need to be done.

References

[1] Araten, M. – Jacobs, M. (2001): *Loan Equivalents for Revolving Credits and Advised Lines*. RMA Journal, vol. 83, no. 8, pp. 34-39.

[2] BCBS (2005): *Guidance on Paragraph 468 of the Framework Document*. [on-line], Basel, Basel Committee on Banking Supervision, c2005, [cit. 16. 6. 2011], <http://www.bis.org/publ/bcbs115.pdf>.

[3] BCBS (2006): *International Convergence of Capital Measurement and Capital Standards. A Revised Framework. Comprehensive Version*. [on-line], Basel, Basel Committee on Banking Supervision, c2006, [cit. 16. 6. 2011], <http://www.bis.org/publ/bcbs128.pdf>.

[4] BCBS (2004): *Studies on the Validation of Internal Rating Systems*. [on-line], Basel, Basel Committee on Banking Supervision, c2004, [cit. 16. 6. 2011], <http://www.bis.org/publ/bcbs_wp14.htm>.

[5] CEBS (2006): *Guidelines on the Implementation, Validation and Assessment of Advanced Measurement (AMA) and Internal Rating Based (IRB) Approaches, CP 10 Revised*. [on-line], London, Committee of European Banking Supervisors, c2006, [cit. 16. 6. 2011], <http://www.eba.europa.eu/getdoc/5b3ff026-4232-4644-b593-d652fa6ed1ec/GL10.aspx>.

[6] EC (2006). *Directive 2006/48/EC of the European Parliament and the Council of 14 June 2006 Relating to the Taking up and Pursuit of the Business of Credit Institutions (Recast).* [on-line], Brussels, European Commission, c2006, [cit. 16. 6. 2011], <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:177:0001:0001:EN:PDF >.

[7] Jacobs, M. (2008): *An Empirical Study of Exposure at Default*. [on-line], Washington, D. C., The Office of the Comptroller of the Currency Credit Risk Analysis Division, c2008, [cit. 16. 6. 2011], <http://michaeljacobsjr.com/Jacobs_AnEmpiricalStudyOfExposureAtDefault_11-08_V2.pdf >.

[8] Moral, G. (2006): *EAD Estimates for Facilities with Explicit Limits*. In: Engelman, B. – Rauhmeier, R. (eds.): *The Basel II Risk Parameters: Estimation, Validation, and Stress Testing.* Berlin, Springer, 2006, s. 197-242.

[9] Smithson, M. – Verkuilen, J. (2005): *Beta Regression: Practical Issues in Estimation*. [on-line], Canberra, Australian National University, c2005, [cit. 16. 6. 2011], <http://psychology3.anu.edu.au/people/smithson/details/betareg/Readme.pdf>.

Exposure at Default Modeling with Default Intensities

Jiří WITZANY

ABSTRACT

The paper provides an overview of the Exposure at Default (EAD) definition, requirements, and estimation methods as set by the Basel II regulation. A new methodology connected to the intensity of default modeling is proposed. The numerical examples show that various estimation techniques may lead to quite different results with intensity of default based model being recommended as the most faithful with respect to a precise probabilistic definition of the EAD parameter.

Key words: Credit risk; Exposure at default; Default intensity; Regulatory capital.

JEL classification: G21, G28, C14.