

# A Through-the-Cycle Model for Retail Lending Economic Capital

Joseph L Breeden  
Strategic Analytics Inc.  
Santa Fe, NM 87505, USA  
breeden@strategicanalytics.com  
+1 505-988-4747

Robert Parker  
Strategic Analytics Inc.  
Santa Fe, NM 87505, USA  
rparker@strategicanalytics.com  
+1 505-988-4732

Carsten Steinebach  
Strategic Analytics Inc.  
Santa Fe, NM 87505, USA  
csteinebach@strategicanalytics.com  
+1 505-988-4741

July 22, 2009

## **Abstract**

Problems in the US Mortgage industry have shown weaknesses in standard regulatory and economic capital approaches. Although significant discussion is occurring around how to segment the portfolios or predict key variables in order to better fit the existing formula, we believe that a re-examination of existing capital formulas is required.

In this paper we develop a formula specifically tuned to the dynamics of retail loan portfolios that could be employed for regulatory capital or economic capital. The key advantages of this approach are that it is based upon a much more accurate model of retail loan defaults, does not require any new data feeds, is based upon readily available modeling frameworks, and can adapt to the portfolio changes such as those observed in the US Mortgage Crisis.

Keywords: Age-Period-Cohort Model, Survival Model, Proportional Hazards Model, Dual-time Dynamics, Panel Data Methods, Retail Lending, Stress Testing

## 1 Introduction

Although the Basel II process has greatly increased the visibility of credit risk management within financial institutions, we know from experience that several weaknesses exist in the application of the Pillar 1 formula for regulatory capital. When Quantitative Impact Study 4.0 (QIS 4)[6] was conducted for the US in the fourth quarter of 2004, mortgage portfolios were assessed to need 60% less capital and home equity portfolios would need 70% less capital than under the Basel I accord.

This assessment of lower capital requirement was computed at a time of historically high originations volumes, meaning that the median age of a loan in these portfolios was very young. Significant changes also occurred in the quality of loans being originated, which were not captured by the traditional application scores.

By 2007, lenders had started to report extreme losses and many of the same institutions that computed lower capital needs under Basel II found themselves under-capitalized. Many forces were adding stress to these lenders, but here we would like to focus on the assessment of regulatory capital for credit risk in the context of its applicability to retail loan portfolios.

## 2 Basel II Regulatory Capital

The regulatory capital formula for credit risk in Basel II was derived from the Vasicek model of portfolio losses. [8] Vasicek used a mean-reverting stochastic model of asset values, and then assumed that when the asset value falls below a certain threshold the borrower would default on the loan.

Vasicek further considered the possibility of multiple loans defaulting simultaneously due to a common environmental factor  $Y_t$ , which he expressed as

$$V_{it} = \sqrt{\rho}Y_t + \sqrt{1 - \rho}Z_{it} \quad (1)$$

$Y_t$  is an unobserved common factor among the assets but is intuitively assumed to represent macroeconomic impacts.  $Z_{it}$  is an idiosyncratic factor for each asset. Both  $Y$  and  $Z$  will change over time, thus the subscript  $t$ .  $\rho$  represents the correlation among assets to  $Y$  and is assumed to be static over time.

A default is assumed to occur if  $V_{it}$  falls below a given threshold  $D_i$ , so using  $I$  as the default indicator

$$I_{it}|Y_t = \begin{cases} 1 & \text{if } V_{it}|Y_t < D_i \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The expected value of  $I_{it}$  is

$$E(I_{it}|Y_t) = \Phi\left(\frac{D_i - \sqrt{\rho}Y_t}{\sqrt{1 - \rho}}\right) \quad (3)$$

The threshold  $D_i$  is equated to  $\phi^{-1}(PD_i)$ , the inverse normal evaluated at the long run probability of default,  $PD$ .

For a portfolio of loans, the central limit theorem gives

$$DR_t = \Phi \left( \frac{\phi^{-1}(PD_i) - \sqrt{\rho}Y_t}{\sqrt{1-\rho}} \right) \quad (4)$$

where  $DR_t$  is the portfolio default rate.

From this relationship, the Basel II formula is obtained by integrating across possible realizations of the environment  $Y_t$  to create a cumulative loss distribution, from which a required solvency level may be evaluated.

$$F = \phi \left\{ \frac{\sqrt{1-\rho}\phi^{-1}(s) - \phi^{-1}(PD)}{\sqrt{\rho}} \right\} \quad (5)$$

This formula depends upon successfully estimating the correlation,  $\rho$ , and long-run  $PD$  as historical constants. This is where the formula breaks down for retail and much effort has gone into better ways of estimating these values. Instead, by considering what is not captured by Equation 5, we can solve the problem of estimating these parameters by altering the structure and assumptions that leads to their requirement.

### 3 Retail-specific Capital

For retail portfolios, the challenges come in what is not considered by Equation 4. When applied to corporate loans, the evolution of credit risk with the age of the loan occurs slowly and is usually ignored [1]. Consequently, it is assumed that ratings agencies have time to refresh the risk ratings for the loans within the pool, and the loans can be segmented by those risk ratings. Equation 4 is then applied independently to each segment.

For retail portfolios, much of the volatility in observed default rates is due to the boom-bust cycle of originations and how that impacts the portfolio via lifecycle effects. This lifecycle "volatility" has nothing to do with the economy and should not be viewed as volatility when computing capital, since lifecycle effects form the core of an expected loss forecast. In short, lifecycle effects should be part of product pricing and loss reserves.

We also know that adverse selection is a real and significant effect in retail portfolios. Credit quality being originated may diverge from that predicted by the credit score. Such adverse selection effects are inherently transitory as they can be explained by unrecorded changes in originations policies and changes in consumer appetite for loans due to macroeconomic conditions. We may, in fact, want to hold capital to protect against adverse selection, so we need to be able to explicitly capture this in the model.

The recent crisis in US mortgage illustrates both of these problems [4]. Just when Basel II was estimating that less capital would be required, the banks were rapidly accumulating risk that was not yet observed by the simple expression in Equation 4 coupled with behavior score segmentations. Said differently, Equation 4 is assumed to work for "well diversified" portfolios, but we now have

abundant evidence that even the largest institutions are not diversified in the timing or credit quality of originations. To be well-diversified, they must be diversified in origination volume, quality, and product. Since this requires an essentially steady-state portfolio, it is clear that few, if any, lenders qualify, and thus require a more robust approach to computing capital.

### 3.1 Derivation

A better model for retail lending regulatory capital can be derived using several different models, but Survival Models provide a simple conceptual framework for deriving a formula for capital.

A Survival Model or Proportional Hazards Model can be expressed as

$$p(t) = h_0(a)e^{B \cdot X_i} e^{C \cdot E(t)} \quad (6)$$

where  $p(t)$  is the probability that a loan will default at time  $t$ , conditional on not having defaulted in a prior period.  $h_0(a)$  represents the change in default risk as a function of age of the account (months-on-books).  $B \cdot X_i$  is essentially a credit score, represented as a linear combination of input factors  $X$  for account  $i$ .  $C \cdot E(t)$  is a linear combination of input macroeconomic factors driving default risk.

Instead of explicitly including macroeconomic factors  $E(t)$ , we represent the net impact of these factors as a non-parametric hidden variable  $z(t)$  analogous to what is done in the Vasicek formula. This is a more robust approach than relying upon specific macroeconomic factors, because it allows for the possibility of non-macroeconomic influences on portfolio performance.

$$p_i = h_0(a)e^{B \cdot X_i} e^{z(t)} \quad (7)$$

To compute capital over a one-year time span, we want to know the cumulative default probability over the next 12 months.

$$PD_i = \sum_{t=1}^{12} h_0(a)e^{B \cdot X_i} e^{z(t)} \quad (8)$$

The age of the account can be expressed as  $a = t - v$ . Also, the credit risk component is independent of time, so

$$PD_i = e^{B \cdot X_i} \sum_{t=1}^{12} h_0(t - v) e^{z(t)} \quad (9)$$

We know that macroeconomic impacts upon a retail portfolio are autocorrelated, but weakly so. The typical autocorrelation time scale for retail portfolios is 6 to 12 months. The autocorrelation length of overall performance time series appears longer because of the autocorrelation introduced by the lifecycle  $h_0(a)$ . When performance is decomposed as shown in 7, the induced autocorrelation for the lifecycle will not be confused with the autocorrelation from macroeconomic impacts.

If we want to compute annual capital, then we can ignore the intra-annual autocorrelation and write

$$PD_i = e^{B \cdot X_i} e^{Z(t)} \sum_{t=1}^{12} h_0(t-v) \quad (10)$$

where  $Z(t)$  is the average value of  $z(t)$  over the 12-month interval.  $Z(t)$  is assumed to be independent of  $Z(t-1)$ . This is an approximation for purposes of implementation simplicity, but we find that it is reasonable given experience with real data.

To compute regulatory capital at a given solvency level  $s$ , we need to know the distribution of  $Z(t)$ . We create a distribution of annual values of  $Z(t)$ . Over many economic cycles,  $Z(t)$  is stationary and should be computed as a mean-zero series, so assuming a normal distribution we obtain

$$CDF(s) = N(z, 0, \sigma) \quad (11)$$

from which

$$z = N^{-1}(s, 0, \sigma) \quad (12)$$

which gives us

$$PD_i = e^{B \cdot X_i} e^{N^{-1}(s, 0, \sigma)} \sum_{t=1}^{12} h_0(t-v) \quad (13)$$

Of course, we do not need to assume a normal distribution, and practical experience shows that a Normalized Inverse Gaussian (NIG) distribution is often useful, in that it allows for skew and kurtosis. In a regulatory context, Normal distributions are a sensible choice to keep the analysis simple for institutions with fewer modeling resources. In an economic capital context, testing other distributions such as NIG would be reasonable.

So far the derivation has maintained the credit score-like factor  $B \cdot X_i$ , but as we saw with recent mortgage history [4], adverse selection is a real concern. Our preference is to replace  $B \cdot X_i$  with a non-parametric function of vintage, since individual account dynamics are not important for capital calculations, but overall trends in credit quality are critical.  $f_Q(v)$  can be computed directly from the performance data during decomposition when the non-parametric maturation and exogenous functions are also computed. The previous equation can be rewritten as

$$PD_i = e^{f_Q(v)} e^{N^{-1}(s, 0, \sigma)} \sum_{t=1}^{12} h_0(t-v) \quad (14)$$

Since Equation 14 no longer has account-level dependency, we see that this expression can be viewed as the cumulative default rate  $CDR$  for a portfolio over a given time span.

$$CDR = e^{f_Q(v)} e^{N^{-1}(s, 0, \sigma)} \sum_{t=1}^{12} h_0(t-v) \quad (15)$$

The form of Equation 15 is actually compatible with more than just Survival Models. Many vendors sell lifecycle curves for various products, and Dual-time-Dynamics (DtD), Age-Period-Cohort (APC) models, and Panel Data methods can all be employed to estimate the components of this model, if they are structured to estimate the hidden variable  $Y_t$  non-parametrically. Traditionally, only DtD and APC have created such a function, but any of these methods could be modified to create appropriate estimates of  $Y_t$ .

Because of the range of well-tested methods that can be employed to estimate the components of this equation and vendors with readily available measures, Equation 15 should be straight-forward for financial institutions to implement.

The largest uncertainty in Equation 15 is  $\sigma$ , the width of the distribution of environmental impacts. Our experience has shown that this is very stable across many different portfolios. It scales with risk level, actually rising for more prime loans. One should expect the proportional impact from the economy to be greater for better loans, because those loans have a lower base loss rate. Subprime loans are proportionately less sensitive.  $\sigma$  should be viewed as a universal constant that will change with risk band and product type, but has been observed by the authors to be remarkably stable across institutions and around the world. It is dramatically more stable than  $\rho$  in the Vasicek formula in the context of retail lending, because lifecycle and credit quality effects have been explicitly incorporated in the formula.

The authors recommend that institutions estimate the lifecycle and credit qualities appropriate to their portfolio, but  $\sigma$  could be estimated and provided as part of the regulatory guidelines or purchase estimates from large data repositories.

To compute the capital requirement, one need simply combine the annual *CDR* of Equation 15 with stressed EAD and LGD estimates according to the existing Basel II guidelines.

## 4 Through-the-Lifecycle

Using the *CDR* in Equation 15 provides a Through-the-Cycle (TTC) calculation of capital, because the distribution used for  $Y_t$  is across the entire available history, not just next year's distribution. This is in contrast to the Point-in-Time (PIT) approach resulting from the Monte Carlo simulation described in Breeden & Ingram, 2008 [2].

However, Through-the-Cycle has begun to take on multiple meanings, so the one-year-forward calculation created above is more properly called a Through-the-Economic-Cycle (TTEC) approach. The other cycle to be considered is the lifecycle of the loan.

To avoid the severe liquidity problems that arose for financial institutions in 2008, one needs to compute not just next year's capital needs, but the needs throughout the lifetime of the loans. In 2004, a huge volume of new loans was being originated in the US mortgage industry, but because losses are predictably

low for the first one to two years of the loan, a standard one-year capital calculation left the industry unprepared for the dramatic rise in losses later.

A more appropriate procedure would be to compute the capital requirements for each of the future years for those loans. Typically a five-year horizon is sufficient. The institution can then plan ahead to increase capital as needed. This could be referred to as Through-the-Lifecycle-Capital (TTLC).

Numerically, computing TTLC capital is a trivial extension of Equation 15, where the range of the summation over the lifecycle is changed to capture the interval being studied. Alternatively, the calculation can be split into annual, quarterly, or even monthly numbers to reveal the timing of peak capital needs, allowing finance to prepare accordingly.

## 5 Segmentation

In order to compute reliable capital ratios, we need to choose an appropriate segmentation. With Equation 15, an appropriate segmentation scheme is one that enhances the stability of the lifecycle  $h_0(a)$  and environmental impact  $Y_t$  estimates. Experience has shown that product type, risk band, loan-to-value, and origination channel are commonly useful variables.

However, contrary to some standard practice, traditional behavior scores should not be used for segmentation. Proportional Hazards Behavior Scores incorporating macroeconomic factors may not exhibit these difficulties, but traditional logistic regression behavior scores without any adjustment for macroeconomic factors are strongly procyclical. Scores deteriorate as the economy deteriorates, but only as a trailing indicator, because they are based upon observed past performance. As a result, score distributions migrate according to lifecycle effects and changes in the macroeconomic environment. Since those factors are already incorporated into Equation 15, segmenting by a standard behavior score will introduce a double-counting and destabilize the capital calculations. Behavior scores of any type can be valuable for account management, but for portfolio modeling, we only want scores with a stable distribution across the economic cycle and lifecycle.

## 6 Estimating $\sigma$

If we accept the Normal distribution approximation of Equation 11, we need to address the issue of estimating  $\sigma$ . This is not unlike the problem of estimating  $\rho$  in the Basel II formula, Equation 5. Because of data limitations at individual institutions,  $\rho$  was viewed as something requiring regulatory guidance. It is also true that for rapidly changing retail loan portfolios, fitting  $\rho$  to the available data can be quite unstable. Estimating  $\sigma$  does not carry the same instabilities since it was designed with retail loan portfolios in mind, but it is still vulnerable to short time series challenges.

The best approach would again be to have regulatory guidance on the proper

value of  $\sigma$ , obtained from looking across many financial institutions around the world. Internally an institution can take steps to augment their estimates of  $\sigma$  for validation and use in economic capital. The typical five- to seven-year data set will not be robust for creating a distribution of annual changes as in Equation 11. The best approach appears to be create a regression model fitting  $Y_t$  to macroeconomic factors

$$\tilde{Y}_t = C \cdot E(t) + \eta_t \quad (16)$$

and then back-casting  $\tilde{Y}_t$  given historical values of  $E(t)$ . Using this approach, estimated time series spanning several decades can often be obtained. Rosenberg & Schuerman [7] used this approach to address issues in risk aggregation and Breeden & Ingram [2] used this approach to validate Point-in-Time economic capital models. This technique would work particularly well with industry-wide data sets, because management action residuals in  $\eta_t$  are reduced.

Experience has shown that the most important macroeconomic factors for retail loan portfolios are unemployment rates and changes in house prices [3]. Fortunately, most countries have long histories available for unemployment rates or levels of non-farm payroll. By looking at the distribution of the logit transform of the unemployment rate, or the log of the ratio of levels of non-farm payroll, year-over-year, we can obtain a good proxy for understanding long term variation in the environmental factor,  $Y_t$ .

## 7 Risk Aggregation

What we want from any capital model is an appropriate marginal distribution for each product type and a time series against which other products can be correlated. Given a correlation matrix, we can then use copulas to aggregate the risk. This is described as a top-down capital aggregation approach, in contrast to the loan-level, bottom-up risk aggregation that is often employed [5]. Bottom-up risk aggregation methods typically require that a single model be employed for all asset classes. With a top-down approach, we can utilize the retail lending specific marginal distributions described above, but allow the corporate loan, equities, derivatives, and commodities portfolios to use their best models to obtain their marginal distributions.

## 8 Conclusions

This paper describes a simple capital formula that derives naturally from the dynamics of retail loan portfolios and is easily implemented via a range of models, all of which are in use in retail lending today. The biggest obstacle to this approach is that shared by any method, limited portfolio performance history. However, the long histories of key macroeconomic factors provide a reasonable approach to estimating the width of the distributions TTC.

The US government bailouts of 2008 and 2009 demonstrate starkly that current capital calculations for retail portfolios are inadequate to the task. Although point-in-time (PIT) methods are available that appear to be effective, through-the-cycle calculations (TTC) such as are desired by regulators require a significant overhaul.

## References

- [1] Joseph L. Breeden. Validation of stress testing models. In George A. Christodoulakis and Stephen Satchell, editors, *The Analytics of Risk Model Validation*, pages 13–25. Academic Press, 2008.
- [2] Joseph L. Breeden and David Ingram. Monte carlo scenario generation for retail loan portfolios. *Journal of Operations Research*, xx:xx, xx 2009.
- [3] Joseph L. Breeden and Lyn C. Thomas. The relationship between default and economic cycle for retail portfolios across countries: identifying the drivers of economic downturn. *Journal of Risk Model Validation*, 2(3):11–44, 2008.
- [4] Joseph L. Breeden, Lyn C. Thomas, and John McDonald III. Stress testing retail loan portfolios with dual-time dynamics. *Journal of Risk Model Validation*, 2(2):43–62, 2008.
- [5] Christian Cech. Copula-based top-down approaches in financial risk aggregation. Technical Report 32, University of Applied Sciences of bfi Vienna, December 2006.
- [6] FDIC. Basel ii and the potential effect on insured institutions in the united states: Results of the fourth quantitative impact study (qis-4). *Supervisory Insights*, 2(2):27–32, 2005.
- [7] J. Rosenberg and T. Schuermann. A general approach to integrated risk management with skewed, fat-tailed risks. Technical report, Federal Reserve Bank of New York, 2004. staff report no. 185.
- [8] Oldrich Alfons Vasicek. Probability of loss on loan portfolio. Technical report, KMV Corporation, 1987.