Credit risk capital allocation differences between developing and developed countries

Ashay Kadam*

August 1, 2016

Keywords : credit risk, economic capital, regulatory capital, Basel accord, developing countries, emerging markets, multiple logistic regression, CreditRisk+

preliminary! please consult the author before citation

*Faculty of Finance, NUS Business School, National University of Singapore. Please address all correspondence to Ashay.Kadam@nus.edu.sg This research is funded by grant R-315-000-104-646 (100%) titled 'Role of Government In Credit Markets'. The data used in this study was provided by the Credit Risk Initiative, Risk Management Institute, National University of Singapore.

Abstract

Basel accords do not differentiate between developing and developed nations. The uniformity in guidelines is despite a plethora of implementation difficulties that developing countries encounter. In this paper we provide evidence supporting differential treatment for developing vs. developed countries in the context of capital adequacy. We compare Asian countries treating the United States as the benchmark. We find that for the selected developed Asian countries the credit risk capital allocation for a typical portfolio of corporates is comparable to that in the US whereas for the selected developing Asian countries the capital allocation could be twice or thrice as much.

Extended Abstract

The Basel Committee on Banking Supervision was established by the central bank governors of the G10 countries in 1975. Earlier Basel accords were shaped mostly by concerns facing developed countries. Since 2009, the committee membership has expanded considerably, and now includes some influential developing countries as well. Several non-member developing countries are also keen on adopting the Basel accord. However, Basel accords do not differentiate between developing and developed nations. In this paper, we argue that for capital adequacy, developing and developed countries should be treated differently, at least in the Asian context.

We use a publicly available dataset with a few million observations for thousands of firms in four influential Asian countries, all of which are members of the Basel committee. For each country considered, we simulate a thousand "typical" well diversified portfolios of risky obligor firms in that country. We take a well-established reduced form model (multinomial logit) to compute obligor default risk and a well-known industry standard model (CreditRisk+) to compute portfolio loss distributions for these simulated portfolios.

From the loss distribution of each simulated portfolio we compute economic capital allocated to offset credit losses, as a proportion of total exposure in the simulated credit portfolio. A simple cross-country comparison of these samples of economic capital shows that in developed countries the capital allocated is comparable to that in the US whereas in developing countries it is not. The latter could be far more, even twice or thrice as much. Thus, we provide evidence in support of different regulatory capital frameworks for developing and developed and countries.

Our evidence is consistent with the fact that is already the case that regulatory capital requirements mandated by central banks of many developing countries are often found to be in excess of what the Basel committee stipulates. The higher capital allocation may be due to a combination of some or all of the following drivers viz. higher default risk, poorer measurement of default risk, or stronger default correlations - disentangling the drivers of differences in capital allocation is the focus of ongoing work.

1 Introduction

"The so-called Basel III rules will impose capital and liquidity requirements that were designed for US and Europe institutions but would be difficult to implement in emerging economies, according to a report set to be issued on Sunday by the B20 group of businesses, which advises the G20 group of nations." - Financial Times, 14th June 2012

The Basel Committee on Banking Supervision was established by the central bank governors of the G10 countries in 1975. Earlier Basel accords were shaped mostly by concerns facing developed countries. Since 2009, the committee membership has expanded considerably, and now includes some influential developing countries as well. Several non-member developing countries are also keen on adopting the Basel accord. However, Basel accords do not differentiate between developing and developed nations.[1] In this paper, we argue that for capital adequacy, developing and developed countries should be treated differently, at least in the Asian context.

Capital offers the last resort for keeping a bank solvent. Therefore, capital adequacy is perhaps the most crucial aspect of basel implementation. In theory, regulatory capital requirements should be designed taking economic capital allocation into consideration. Conceptually economic capital allocation can be quantified by taking the difference between some measure of maximum loss (such as VaR or expected shortfall) and the expected loss on the portfolio.

Taking the United States as the benchmark for comparison, and focusing exclusively on credit risk, we examine whether in the Asian context economic capital allocation is comparable for developing and developed countries.[2]

[1]This one size fits all approach is despite some concerns about inadequate data, inadequate technology and such other implementation aspects in developing countries. There have also been concerns about the potentially damaging after-effects of imposing unbearable regulation on weak financial institutions.

[2]We choose the US as a benchmark for two reasons. First is due to America's influence on world economy. Second is the wide financial data coverage for American firms.

We use a publicly available dataset with a few million observations for thousands of firms in four influential Asian countries, all of which are members of the Basel committee. For each country considered, we simulate a thousand “typical” well diversified portfolios of risky obligor firms in that country. We take a well-established reduced form model (multinomial logit) to compute obligor default risk and a well-known industry standard model (CreditRisk+) to compute portfolio loss distributions for these simulated portfolios.

We chose multinomial logit because it is one of the standard ways of mapping obligor specific data to probability of obligor defaults. Our model for obligor default is a special case of generalised linear models with the link function being logit.[3] We chose the CreditRisk+ framework to model portfolio credit risk because it is the simplest to implement among the three well known industry standards for this purpose. This is a simulation study, driven by real data, and reasonable modeling approaches. The focus here is not on any specific model of obligor default or any specific framework to compute portfolio loss. Our interest is in the differences that emerge between developing and developed countries, given the characteristics of their different datasets, having adopted the same model and methodology for all countries.

From the loss distribution of each simulated portfolio we compute economic capital allocated to offset credit losses, as a proportion of total exposure in the simulated credit portfolio. A simple cross-country comparison of these samples of economic capital shows that in developed countries the capital allocated is comparable to that in the US whereas in developing countries it is not. The latter could be far more, even twice or thrice as much. Thus, we provide evidence in support of different regulatory capital frameworks for developing and developed and countries.

It is not immediately obvious what drives the higher capital allocation in developing countries. Perhaps the firms in these countries have intrinsically higher default risk. Perhaps these countries suffer from a poorer measurement of default risk which itself could be due

[3]It turns out that modeling other exits such as mergers and acquisitions in addition to defaults significantly improves default probability estimates. Hence the use of multinomial (instead of binomial) logit.

to a myriad of factors like data availability, accounting standards, information and communication technology infrastructure etc. Perhaps it is the stronger default correlations between the firms that result in a propensity for higher losses. Perhaps it is some combination these drivers. Disentangling the drivers of higher economic capital allocation for developing countries is the focus of ongoing work.

2 Literature Review

Traditionally, credit risk models have fallen into one of the two categories viz. structural vs. reduced form models. Structural models attempt to model the obligor firm's fundamentals. See for instance Black and Scholes 1973, Merton 1974, Black and Cox 1976, Geske 1977, Leland 1994 and Longstaff and Schwartz 1995. While structural models were firmly grounded in economic theory, making transparent the relationship between a firm's capital structure and its credit risk, their (endogenous) default predictions were inconsistent with empirical observations. This shortcoming likely fueled the advance of credit risk literature in reduced-form models which treat the firm's (exogenous) default as a surprise event. See for instance Litterman and Iben 1991, R. Jarrow and Turnbull 1997, Madan and Unal 1998 and Duffie and Singleton 1999. See Lando 2004 or Bielecki and Rutkowski 2004 for a comprehensive overview of reduced form models for credit risk.

In reality, most reasonable approaches to modeling an obligor's credit risk are not on either of the two extremes, but take a somewhat hybrid approach. R. Jarrow 2004 have argued that the difference between the structural and reduced form model families is somewhat artificial. They argue that the difference between these two model families can be completely understood simply by focusing on the information assumed known by the modeler. Structural models form one extreme where the modeler is more like the firm's manager and can observe very detailed information at firm level, whereas reduced form models form the other extreme where the modeler is like the market investor who has no knowledge of the firm fundamentals.

The approach we take in this paper is also a hybrid approach but leans more towards structural models. Akin to most structural models, we use the concept of distance to default computed using an option theoretic characterization of the firm's capital structure. This is largely what drives the estimate of a firm's default probability in our analysis. However, we do supplement the information on the firm's distance to default by accounting data from the firm's financial statements as well as market data such as interest rates and stock indices. The latter improve the default probability estimation considerably.

Having modeled and estimated an individual obligor's probability of default, the next practical requirement for a financial institution exposed to credit risk is to compute the probability distribution of losses from a collection of default risky obligors (e.g. a basket of corporate bonds, a loans portfolio etc). This requires a careful treatment of the dependency between the individual obligor propensities to default (which may be summarised, for instance, by their default correlations). There are three widely used industry standards for this.

First, based on the credit migration approach, as proposed by JP Morgan with CreditMetrics, described in J. P. Morgan 1997. Second, the option pricing, or structural approach, as initiated by KMV, based on the Merton model described in Crosbie and Bohn 2003. Third, the actuarial approach as proposed by Credit Suisse Financial Products (CSFP) with CreditRisk+, described in CreditSuisse 1997, which focuses only on default and which assumes that default of individual obligors follows an exogenous Poisson process.

An excellent overview of all three, as well as a comparative analysis, can be seen in M. Crouhy and Mark 2000. In a similar vein, Wieczerkowski 2004 explores the relationship between CreditRisk+ and CreditMetrics by showing that there exists in general a consistent parametrization of that results in the same loss distribution.

In this paper we use the CreditRisk+ approach because we find it simpler than the other two alternatives mentioned above when it comes to implementation and repeated simulations.

Armed with the distribution of credit losses, a lender can estimate the capital to set aside so as to offset the (so called unexpected) credit losses. Kadam and Lenk 2008 show that rating migrations can vary significantly across countries, and that heterogeneity in obligor rating migrations affects economic capital allocation. This paper takes that train of thought forward by exploring in more detail, the heterogeneity in capital allocated across countries, especially for developing vs developed countries.

Our work is consistent with many others that describe challenges to Basel accord adoption by all developing countries. Griffith-Jones and Spratt 2001 and Gottschalk and Griffith-Jones 2006 have repeatedly raised concerns about both the reduction of lending to developing countries and the increased pro-cyclicality of bank lending resulting from Basel II. Tonveronachi 2009 elaborates on these and some other implementation related criticisms in more detail. Ward 2002 argues that both markets and supervisors are more likely to fail in developing countries making Basel accord unsuitable there.[4]

There have also been studies focusing on specific developing countries. For instance Barrell and Gottschalk 2006 find that a shock to capital adequacy ratios for Brazil and Mexico has an adverse effect on their GDP whereas S. Hai and Ahmed 2001 address Basel II implementation challenges in Pakistan.

While each of the above studies deals with a specific version of the Basel accord, their objections are generic enough to apply to all three versions to date.

Broadly speaking, the concerns raised so far fall into one of the two categories. First is the (predicted) ill-effects of Basel accord implementation on the developing country from a macro-economic perspective. Second is difficulty in implementation due to lack of technology, infrastructure, skills, historical data etc. What we argue in this paper does not neatly fall into either category.

We show that purely based on empirical results, applying the same rule for capital adequacy to developed and developing countries results in dramatically different amounts of capital allocated. Therefore uniformity in capital adequacy rules seems unreasonable. Our result has an overlap with the first set of concerns mentioned above in the sense

[4] Ward 2002 goes so far as to say that the Basel II "framework has not been designed with developing countries in mind, and it is especially likely to fail in developing countries."

that an unbearably higher capital requirement could damage a developing country's banking system. Our result has an overlap with the second set of concerns mentioned above in the sense that the drivers of higher capital allocation may very well be rooted in difficulties with implementation.

3 Model

We use two models. At obligor level, the probability of default (PD) is estimated using a multinomial logistic regression model. This estimation also yields standard errors of PD estimates. The PDs themselves and their standard errors then feed into a portfolio credit risk model which is CreditRisk+, an industry standard.

3.1 Model for obligor probability of default

At any given time t an obligor i is in state s where s takes one of the following values : survival, default, merged with or acquired by another firm, delisted for some other reason. The state s of the obligor i at time t gives us the dependent indicator variables $Y_{i,t,s}$.

In order to predict these dependent variables we use observable covariates $\{V_j\} = V_1, V_2, \ldots, V_n$ which are a mix of firm level observations (e.g. accounting ratios derived from its accounting statements) and economy level observations (e.g. interest rate, stock index). Exhibit 1 gives the full list of explanatory variables.[5]

A linear combination of these observations, for a known firm at a known time, is then mapped to the logarithm of the odds ratio for that state for that firm at that time. Thus our obligor level model is

$$\forall i, t, s \quad \log\left(\frac{P(Y_{i,t,s} = 1)}{1 - P(Y_{i,t,s} = 1)}\right) = \sum_{j=0}^{n} \beta_j V_{j,i,t,s} + \epsilon_{j,i,t,s} \qquad (1)$$

where the last term is an error term. In the model specification above $V_0 \equiv 1$ because we wish to have an intercept term. We assume that all error terms are i.i.d. with standard normal distribution.

[5] Note that these observations are made for each firm i at each time t but we dropped the subscripts i and t for convenience. Note further that prior to estimation all explanatory variables are standardised to have zero mean and unit variance.

At any time a firm must be in one of the states mentioned above i.e. the state space is exhaustive. Thus, at any time for a given firm, the probabilities corresponding to various states will sum to 1. Therefore, during model estimation we will need to choose a base case as reference for the firm's state and estimate parameters for all other states. Without loss of generality we choose reference state to be 'survival'.

3.2 Model for portfolio credit risk

CreditSuisse 1997 is the best resource for a full description of CreditRisk+, which is the portfolio credit risk model we use. In our implementation, the inputs to the model are a portfolio of default risky exposures (we assume constant loss given default), the industry sector that each exposure can be mapped to, and for each exposure an expected probability of default as well as the standard deviation of the probability of default. The output of our model implementation is a portfolio loss distribution, which naturally yields summary statistics such as expected loss and value at risk.

4 Data

We use data provided by the Credit Risk Initiative, Risk Management Institute, National University of Singapore. The size of the dataset is described in Table 1. The data has two components. The first component is the data on defaults as well as events such as mergers, acquisitions, delisting etc. This is firm level data with daily frequency. The second component is the data on explanatory variables we use to predict defaults. This data is firm level data at monthly frequency. The comprehensive list of explanatory variables used is given in Exhibit 1.

5 Methodology

For this study we focus on four Asian member nations viz. India, China, Japan, and South Korea. All four nations are members of the Basel committee but only the first two are considered developing countries. The methodology described in this section is applied to

Country	# Firms	# Observations	# Defaults	# Other Events
India	4794	497057	202	124
China	2487	315598	489	47
United States	13080	1520586	703	6834
South Korea	2429	320454	188	337
Japan	5134	843276	195	1385

Table 1: Data summary

This table summarises the data for each country in this simulation study. The last two columns summarise data on defaults as well as events such as mergers, acquisitions, delisting etc. This is firm level data with daily frequency. The first two columns summarise the data on explanatory variables we use to predict defaults. This data is firm level data at monthly frequency. The comprehensive list of explanatory variables used is given in Exhibit 1.

each of these four countries, as well as to the benchmark viz. United States. The analysis four each country is done in four steps.

The first step is to clean and format the data, thus preparing the inputs for the second step, which is to fit a multiple logistic regression model. The third step is to estimate default probabilities and their standard deviations. The fourth step is to compte the loss distributions for sample credit portfolios using the CreditRisk+ framework. A short summary of each step is given below. The detailed R code corresponding to each step is given in Exhibit 2.

5.1 Data cleaning and formatting

The data for explanatory variables (X data) and the data on defaults (Y data) are retrieved separately and reformatted for joining. Fine-grained event codes are collapsed into the following broad event categories: delisted, merged or acquired, defaulted (soft), defaulted (hard), survived.[6] Adjacent default events are treated as a single default event and defaults are recognised on their first occurrence. Information on soft and hard default events occuring close to each other is merged. X and Y data are joined treating the pair (firm id, month) as the key.

5.2 Multiple logistic regression model

The state space for the dependent variable is further collapsed by treating hard and soft defaults as identical. All explanatory variables

[6]Eventually, when running multiple logistic regression, hard and soft defaults will be further combined into one category viz. defaults.

are standardised i.e. adjusted for mean and variance so that the new means and variances for each explanatory variable are 0 and 1 respectively. Standardised data is then truncated to include only observations up until the end of the year 2013. This is the in-sample data. The rest of the data i.e. year 2014 onwards is treated as out-of-sample data and stored separately. The in-sample data is used to estimate the model coefficients in a multiple logistic regression model, whereas the out-of-sample data is where the coefficients are applied to make default predictions.

We feed the in-sample data and the model specification to the mlogit package in R. The model specification given to mlogit is exactly as described in Section 3.1 (with the list of variables as given in Exhibit 1) and an intercept term. Prior to this the data has to be reshaped as required by mlogit. The state of survival is treated as the base case for the dependent variable. See Croissant 2013 for more details on the mlogit package. The model estimation is with maximum likelihood estimation performed by the mlogit package (and transparent to us) using the package maxLik which is described in Henningsen and Toomet 2011.

5.3 PDs and their std deviations

Having fitted the multiple logistic regression model to in-sample data, the coefficients so obtained, and their standard errors are now used to make predictions for the out-of-sample data. From the coefficients and their standard errors, we get log odds ratios as well as their standard errors for each (firm, month) combination in the out-of-sample data. However, what we actually need as an input to the portfolio model is the firm-wise expected probabilities and their standard errors.

From the means and standard deviations of log-odds-ratios, the corresponding mean and variance for probabilities at firm level can be readily obtained by simulation. Log odds are assumed to be normally distributed.[7] Therefore, random samples can be created for log odds of each firm using the information on means and standard deviations of log odds. Each point in the sample is then mapped to a probability by inverting the relation between log odds and probability. The sample

[7]Recall that the last term in equation 1 has a standard normal distribution.

of default probabilities so obtained readily yields an expected default probability and the standard deviation of the default probability for each firm.

The probabilities we just obtained have a monthly horizon as we used monthly data. We multiply them by 12 so as to change the default time horizon from monthly to annual.[8] Thereafter, for each country, the bottom 50% (low risk) firms are discarded, and from the remainder, a random sample of 500 firms is chosen. All subsequent analysis therefore uses 500 randomly selected high risk firms from each country.[9]

5.4 Loss distributions using CreditRisk+

From the previous step we obtain a random out-of-sample set of 500 high risk firms for each country, for which the expected default probability and standard deviation of default probability are known. This set can be used to create the base credit risky portfolio of 500 exposures.

CreditRisk+ carefully accounts for the industry sector(s) to which each firm maps to. Our implementation is simplified by the restriction that each firm belongs to only one industry sector. Furthermore we select only the firms that belong to the those industry sectors which occur frequently in the data viz. Basic Materials sector, Communications sector, Consumer sector, Financial sector, Industrial sector and Technology sector.[10]

The loss distribution computed by CreditRisk+ is usually smoother for larger portfolios. Therefore, we create a large hypothetical portfolio with 10000 exposures by replicating each 500 firm base portfolio 20 times. The portfolio exposure amounts are random, but with tight

[8] This linear approximation can be shown to hold quite well if the monthly default probability is small and defaults occuring in distinct months are independent of each other.

[9] In other words we assume that a typical lending institution has a portfolio in which each obligor has a risk which is higher than the median risk among all firms in that country.

[10] Thus, for this simulation study, we do not include any firms in the energy, utility and government sectors due to their sparsity in the dataset.

Country	Expected Loss	99.9 % VaR	99.9 % Expected Shortfall
India	55.66	77.75	79.87
China	118.90	146.40	148.90
United States	10.32	18.82	19.65
South Korea	16.37	27.13	28.16
Japan	4.16	9.90	10.45

Table 2: Cross-country comparison of simulated credit portfolio loss

Table 2 summarises, for each country, the portfolio losses for a thousand simulated portfolios, holding the total portfolio exposure, as well as the number of exposures in the portfolio, to be approximately constant (at 1 million and 10000 respectively). The summary statistic used is the median computed from 1000 simulated portfolios for each country. Each entry is expressed in basis points per dollar of total portfolio exposure. Each column represents one important summary statistic from the portfolio loss distribution. The loss distributions for each portfolio are computed using the CreditRisk+ model.

variation around 100 units each. Thus the total portfolio exposure is approximately approximately 1 million units.

One thousand such random portfolios are simulated for each country, and CreditRisk+ framework is applied to each such portfolio. During the CreditRisk+ implementation, the following assumptions are made:

- portfolio losses are discretised to units of 10
- a constant recovery rate of 40% is assumed
- the confidence level is assumed to be 99.9%

6 Preliminary Findings

Table 2 summarises, for each country, the portfolio losses for a thousand simulated portfolios, holding the total portfolio exposure, as well as the number of exposures in the portfolio, to be approximately constant (at 1 million and 10000 respectively). The summary statistic used is the median computed from 1000 simulated portfolios for each country. Each entry is expressed in basis points per dollar of total portfolio exposure.

The loss statistics in developing countries, shown in the first two rows, are considerably larger than those in the case of the US or developed countries in Asia. This is true of expected loss, as well as unexpected losses (measured by value at risk or expected shortfall).

One could therefore surmise that economic capital allocation in developing countries would also be higher.

The difference between Value at Risk and expected loss yields the economic capital for each simulated portfolio. We express this as a proportion of total exposure for that portfolio, in basis points per dollar of total exposure. There are 1000 simulated portfolios for each country. We produce for each country one box plot summarising the thousand simulated economic capital amounts for that country.

Figure 1 shows the box plots for economic capital allocations. Each box plot corresponds to one country and summarises 1000 economic capital allocations for the 1000 portfolios simulated using firm level and country level data for that country. It is evident that the economic capital allocation for developing countries in Asia is much higher than that for developed countries in Asia. The latter are comparable to the economic capital allocation in the US but the former could be twice or thrice as much.

7 Further tests

While summaries in Table 2 and Figure 1 based on results for 4 member nations and recent data offer a strong indication of the difference between developing and developed countries, it is plausible that such a picture emerged purely by chance. A statistical (one sided) test is more appropriate to test the claim that capital allocation for developing countries is higher than that for developed countries.

7.1 Mann-Whitney-Wilcoxon Test

As there is no basis for economic capital numbers to be i.i.d, an ordinary means test under distributional assumptions cannot work. We do a non-parametric test on the two pooled samples (one pooled sample being for developing countries in Asia, the other for developed countries in Asia). We use the Mann-Whitney-Wilcoxon test with the null hypothesis that there is no difference between these pooled samples, against the alternative hypothesis that the average economic capital for developing countries is greater than that for developed countries.

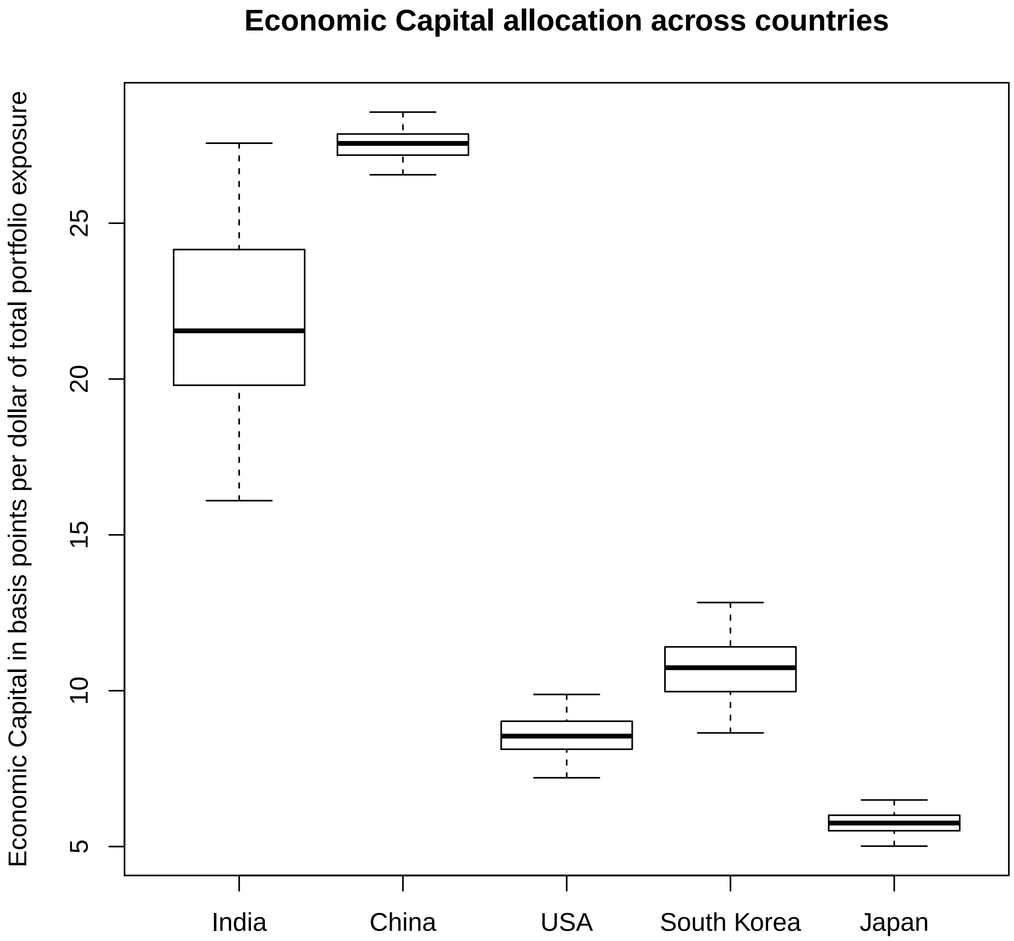

Figure 1: Economic capital allocation: developing vs. developed countries

This figure shows the box plots for economic capital allocations. Each box plot corresponds to one country and summarises 1000 economic capital allocations for the 1000 portfolios simulated using firm level and country level data for that country. It is evident that the economic capital allocation for developing countries in Asia is much higher than that for developed countries in Asia. The latter are comparable to the economic capital allocation in the US but the former could be twice or thrice as much.

7.2 Robustness

To further strengthen the claim, we add two more countries viz. Hong Kong and Malaysia for cross-sectional robustness, and test separately for three periods viz. before, around and after the global financial crisis (we took three separate five year periods 2000-2004, 2005-2009, and 2010-2014) for robustness across time.[11] Further, for the remaining analysis we drop the 'high-risk' filter mentioned in footnote 9 in order to remove potential biases arising from sampling from the closer to the tail of the distribution of risky firms.

7.3 Results

After making all the above changes, and repeating the Mann-Whitney-Wilcoxon test for each period, the test results were similar across the three five year periods, and similar to those obtained using the data from the entire range of fifteen years taken together. The null hypothesis of population similarity was rejected with a p-value of 0 for each test, supporting the alternative hypothesis that for developing countries the average economic capital is greater than that for developed countries. Changing the measure of tail risk from VaR to expected shortfall in computing the economic capital did not alter the result.

8 Conclusion

Basel capital rules do not differentiate between developing and developed countries. However, we find that for hypothetical "typical" simulated credit risky portfolios, credit risk capital allocation for developing countries can be significantly higher than that for developed countries, even twice or thrice as much. The finding is robust to cross-sectional variation in the countries used, and time variation as in before, around or after the recent global financial crisis. This dramatic difference in simulated economic capital amounts between developing and developed countries begs for a differential capital treatment.

[11] However, in adding countries and varying time, we had to change some aspects of the methodology due to accommodate for reduced data coverage. We constructed the sample of economic capital predictions for any given year based on model estimation for a five year rolling window prior to that year, and combined five such years of predictions to make up the data for one period. In adding more countries the data coverage being smaller for the newly added countries, the sizes of simulated portfolios were reduced from 500 to 100.

1 Exhibit

Independent variables in the obligor PD model

The 12 explanatory variables we use in estimating obligor default probabilities can be separated into two categories.

The first category captures market information. This has two elements. Stock index return **indexReturn** is the trailing one-year simple return on a major stock index of the country under consideration. Interest rate **r3month** is a representative 3-month short-term interest rate.

The second category captures firm specific information derived from the firm's financial statements as well as its stock price. **dtd** is the firm's distance to default. **cashTa** is the ratio of firms cash to total assets. **niTa** is the ratio of the firm's net income to total assets. **size** is number of shares outstanding times share price. **mktBook** is the ration of market to book value of the firm. **sigma** is the volatility of the firm. For some of these variables both the level and short term trend are computed and used as separate explanatory variables.

The full list of variables is as follows:

- indexReturn
- r3month
- dtdAvg, dtdDiff
- cashTaAvg, cashTaDiff
- niTaAvg, niTaDiff
- sizeAvg, sizeDiff
- mktBook
- sigma

2 Exhibit

2.1 R Code for the main program

```
home <<- "D:/Simulation Paper/"
set.seed(456)
fullModel <<- T
economies <- as.character(c(2,10,4,6,15))
names(economies) <- c("China", "South Korea", "India", "Japan", "United States")
cutoffYear <- 2013

economiesAll <<- economies
# the function rbindlist is redefined in package data.table, will produce errors
# to avoid confusion we define a new function rbindList and use this throughout
rbindList <<- function(l){Reduce(rbind,l)}

# source and run all the code modules
 setwd(paste0(home, "Code"))
 source("libraryDataPreparation.R")
 source("libraryLogisticRegression.R")
 source("librarySimulation.R")
 source("libraryCreditRisk+.R")
 setwd(home)

# read and pre-process X and Y datasets separately, then merge them
  #prepareData()

# fit logistic regression Y~X i.e regress log odds Y against X
 fitLogisticRegression()

# simulate PDs (on out-of-sample data)
# based on the coefficients of logistic regression (fitted on in-sample data)
 simulateProbabilities()

# create portfolio samples based on the simulated PDs for the out of sample universe
     of firms
 simulatePortfolios()

# apply CreditRisk+ model to the sample portfolios; get loss distributions and their
     summary statistics
 computeCreditLossDistributions()
```

2.2 R Code for data cleaning

```
######################################## X DATA PREPARATION
     ####################################################

readX <- function(economy){
# reads raw data from RMI in original format but split at firm level, combines into a
     single RData file at econ level
 setwd(paste0(home,"Data/X/Raw/",economy))
 dataX <- rbindList(lapply(list.files(),function(fileName){read.csv(fileName, header=F
      , stringsAsFactors = F)}))
 colnames(dataX) <- c("idRMI","year","month","indexReturn","r3month","dtdAvg","dtdDiff
      ","cashTaAvg","cashTaDiff","niTaAvg","niTaDiff","sizeAvg","sizeDiff","mktBook","
      sigma")
 return(dataX)
}
reformatX <- function(dataX){
# reformats company id and date information, sorts data frame by month
 # get rid of mapping numbers and allow multiple defaults for same firm
 dataX$idRMI <- as.factor(round(as.numeric(dataX$idRMI)/1000))
 # recompute month as number of months since 1900, factor month and year
 # note you need to do month reformat before year reformat
 # this is because for month recomputation you need year to be numeric not factor
 dataX$month <- (dataX$year - 1900)*12 + dataX$month
 # order chronologically so month and year can be stored as ordered factors
 print(dim(dataX))
```

```
 dataX <- dataX[order(dataX$month),]
 dataX$year <- ordered(dataX$year)
 dataX$month <- ordered(dataX$month)
 return(dataX)
}
outX <- function(dataX,economy){
 # Reorder and Output
 dataX <- dataX[order(dataX$idRMI, dataX$month),]
 path <- paste0(home, "Data/X/Cleaned/")
 write.csv(dataX, file = paste0(path,economy,"dataX.csv"), row.names = F)
 save(dataX, file = paste0(path,economy,"dataX.RData"))
}
dropRowsAllNA <- function(d){d[!is.na(rowSums(d)),]}
createX <- function(){
# reads, cleans, reformats and outputs X data in RData and csv formats
 setwd(paste0(home,"Data/X/"))
 for(economy in economies){
  unformatted <- dropRowsAllNA(readX(economy))
  dataX <- reformatX(unformatted); rm(unformatted); gc()
  outX(dataX, economy); gc()
 }
}

######################################### Y DATA PREPARATION FIRST PART
      ########################################

readCreditEventTable <- function(){
 # before using this function reformat the credit event table to remove data
       formatting errors
 # this can be done using the awk script in Data/Y folder thereafter use reformatted
       version
 eventsFilename <- paste0(home,"Data/Y/Raw/", "credit_event_table_reformatted.csv")
 events <- read.csv(eventsFilename,header = T, stringsAsFactors=F)
 colnames(events) <- c("idRMI", "eventDate", "eventCode", "eventType", "eventReason")
 return(events)
}
factorizeColumns <- function(events){
 # convert firm id to ordered factor type
 events <- events[order(as.numeric(as.character(events$idRMI))),]; events$idRMI <-
       ordered(events$idRMI)
 # convert event qualifiers to factor type
 events <- events[order(events$eventCode),]; events$eventCode <- ordered(events$
       eventCode)
 events$eventType <- factor(events$eventType)
 events$eventReason <- factor(events$eventReason)
 return(events)
}
manipulateEventCodes <- function(events){
 # Note : credit events with codes 29 (default resolution) and 74/75 (listing change)
 # are not default events.
 # Furthermore, they create data IO problems due to empty reason fields
 # Hence the step below to drop the rows corresponding to these event codes
  events <- events[!events$eventCode %in% c("29","74","75"),]
 # classify events as delisting and M&A
 acquisitionFlag <- events$eventReason %in% " Reason for delisting: Acquired/Merged"
 delistingFlag <- !(events$eventReason %in% " Reason for delisting: Acquired/Merged")
       & (events$eventType %in% "Delisting")
 events[delistingFlag, "classification"] <- "delisting"
 events[acquisitionFlag, "classification"] <- "acquisition"
 # classify events as hard or soft defaults
 # Note event code 203 is delisting for bankruptcy, was classified as delisting above;
        re-classify as hard default
 hardDefaultFlag <- events$eventCode %in% as.character(c(100:115,118:126,128,203,301))
 # Note that as per RMI classification event codes 116, 117, and 127 are soft defaults
        even though they map to bankruptcy
 softDefaultFlag <- events$eventCode %in% as.character(c(116,117,127, 300,302:334))
 events[hardDefaultFlag, "classification"] <- "hard default"
 events[softDefaultFlag, "classification"] <- "soft default"
 # several other event codes do exist, bucket them into one category called survival
 events[is.na(events$classification),"classification"] <- "survival"
 events$classification <- factor(events$classification)
 return(events)
}
manipulateEventDates <- function(events){
 # reformat as date object to extract month and year
 # both stored as ordered factors
 events$date <- as.POSIXlt(events$eventDate, format = "%Y-%m-%d", tz = "EST")
```

```
 events$year <- 1900 + events$date$year
 events$month <- 12*events$date$year + events$date$mon
 # drop both date related columns keeping only year and month
 events <- events[ , !names(events) %in% c("date","eventDate")]
 # sort the dataframe chronologically then convert month and year to ordered factors
 timeOrder <- order(events$month)
 events <- events[timeOrder,]
 events$month <- ordered(events$month)
 events$year <- ordered(events$year)
 return(events)
}
deleteDuplicates <- function(events){
 # delete duplicates due to change from daily to monthly, and multiple defaults in a
      month
 events <- unique(events)
 delFlag <- duplicated(events[,!names(events) %in% "eventCode"])
 events <- unique(events[!delFlag,])
}
createRawY <- function(){
 events <- manipulateEventCodes(factorizeColumns(readCreditEventTable()))
 dataY <- deleteDuplicates(manipulateEventDates(events)); rm(events); gc()
 write.csv(dataY, file = paste0(home, "Data/Y/Raw/dataY.csv"))
 save(dataY, file=paste0(home, "Data/Y/Raw/dataY.RData"))
 return(dataY)
}

################################################# Y DATA PREPARATION SECOND PART
     #################################

deleteOneDefault <- function(d){
 # d has same data structure as events but very few rows because it
 # has default data for only one firm; all d$idRMI values are identical
 d$classification <- as.character(d$classification)
 if(nrow(d) != 1){
  d$month <- as.numeric(as.character(d$month))
  d$gap <- c(0,diff(d$month))
  d$smallGap <- c(F, d$gap[-1] <= 6 )
  delRow <- match(T,d$smallGap)
  if(!is.na(delRow)) {
   if(d[delRow,"classification"] == "hard default") {
    d[delRow - 1,"classification"] <- "hard default"
   }
   d[delRow - 1,"eventCode"] <- d[delRow, "eventCode"]
   d <- d[-delRow,]
  }
  d$gap <- NULL
  d$smallGap <- NULL
 }
 return(d)
}
reduceSpecificDefaults <- function(defaults){
 defaults$idRMI <- ordered(defaults$idRMI)
 defaultList <- split(defaults,defaults$idRMI)
 totalOld <- Reduce(sum, Map(nrow,defaultList))
 defaultList <- lapply(defaultList, deleteOneDefault)
 totalNew <- Reduce(sum, Map(nrow,defaultList))
 print(totalOld - totalNew)
 defaults <- Reduce(rbind,defaultList)
 defaults$classification <- factor(defaults$classification)
 if(totalNew == totalOld) {return(defaults)}
 else {Recall(defaults)}
}
separateEvents <- function(events){
 softDefaultsFlag <- events$classification == "soft default"
 hardDefaultsFlag <- events$classification == "hard default"
 defaultsFlag <- softDefaultsFlag | hardDefaultsFlag
 defaultEvents <- events[defaultsFlag,]
 otherEvents <- events[!defaultsFlag,]
 separation <- list(defaultEvents, otherEvents)
 names(separation) <- c("defaultEvents", "otherEvents")
 return(separation)
}
reduceDefaults <- function(defaultEvents){
 softDefaults <- reduceSpecificDefaults(defaultEvents[defaultEvents$classification ==
      "soft default",])
 hardDefaults <- reduceSpecificDefaults(defaultEvents[defaultEvents$classification ==
      "hard default",])
```

```
 allDefaults <- reduceSpecificDefaults(rbind(softDefaults, hardDefaults))
}
createY <- function(events){
 # reorder by <firm,time> and output
 events <- events[order(events$idRMI, events$month), ]
 # rearrange columns - first 2 will form the key with X data
 events <- events[,c("idRMI", "month", "year", "classification", "eventCode")]
 # separate events into default and non-default events
 separation <- separateEvents(events)
 defaultEvents <- deleteDuplicates(reduceDefaults(separation$defaultEvents))
 otherEvents <- deleteDuplicates(separation$otherEvents)
 events <- rbind(defaultEvents, otherEvents)
 o <- order(events$idRMI, events$month)
 dataY <- events[o, ]; rm(events); gc()
 path <- paste0(home, "Data/Y/Cleaned/")
 write.csv(dataY, file = paste0(path, "dataY.csv"))
 save(dataY, file=paste0(path, "dataY.RData"))
}

#################################### MERGE X AND Y DATASETS
      ####################################################

replaceFactorNAs <- function(m,colName, levelName){
 levels(m[,colName]) <- c(levels(m[,colName]),levelName)
 m[is.na(m[,colName]),colName] <- levelName
 m[,colName] <- factor(m[,colName])
 return(m)
}
reformatXY <- function(m){
 m <- replaceFactorNAs(replaceFactorNAs(m, "classification","survival"), "eventCode","
      0")
 m$year <- factor(m$year, ordered=T)
 m$month <- factor(m$month, ordered=T)
 m <- m[order(m$idRMI, m$month), ]
}
mergeXY <- function(economy){
 load(paste0(home,"Data/Y/Cleaned/dataY.RData"))
 load(paste0(home,"Data/X/Cleaned/",economy,"dataX.RData"))
 print(economy)
 # MERGE X DATA with Y DATA, and output
 # use merge.data.frame instead of merge to avoid invoking merge() from data.table
      package
 dataXY <- reformatXY(merge.data.frame(dataX, dataY, all.x = T))
 write.csv(dataXY, file = paste0(home, "Data/XY/",economy,"dataXY.csv"))
 save(dataXY, file = paste0(home, "Data/XY/",economy,"dataXY.RData"))
 return(dataXY)
}
reformatSpecificEconomy <- function(economy){
 load(paste0(home,"Data/XY/",economy,"dataXY.RData"))
 dataXY$economy <- economy
 return(dataXY)
}
tabulateClassificationsByEconomy <- function(){
 load(paste0(home, "Data/XY/dataXY.RData"))
 dataXY <- dataXY[dataXY$economy %in% economies & dataXY$classification != "survival"
      ,]
 dataXY$economy <- factor(dataXY$economy)
 dataXY$classification <- factor(dataXY$classification)
 t <- table(dataXY$economy, dataXY$classification)
 row.names(t) <- names(economies[pmatch(levels(dataXY$economy), economies)])
 return(t)
}
createXY <- function(){
 for(economy in economies) mergeXY(economy); gc()
 dList <- lapply(economies, reformatSpecificEconomy)
 dataXY <- rbindList(dList); rm(dList); gc()
 dataXY$economy <- factor(dataXY$economy)
 dataXY$idRMI <- factor(dataXY$idRMI, ordered=T)
 dataXY$year <- factor(dataXY$year, ordered=T)
 dataXY$month <- factor(dataXY$month, ordered=T)
 dataXY <- dataXY[order(dataXY$economy, dataXY$idRMI, dataXY$month), ]
 write.csv(dataXY, file = paste0(home, "Data/XY/dataXY.csv"))
 save(dataXY, file = paste0(home, "Data/XY/dataXY.RData"))
 t <- tabulateClassificationsByEconomy()
 write.csv(t, file = paste0(home, "Data/XY/TableOfClassifications.csv"))
}
```

```
##################################### MAIN FUNCTION FOR DATA PREPARATION
     #####################################

prepareData <- function(){
 # create dataset for independent variables
    print("Creating X")
    createX(); gc()

 # create dataset for dependent variable
    print("Creating Y")
    createY(createRawY()); gc()

 # merge the above two datasets
    print("Creating XY")
    createXY()
}
```

2.3 R Code for multiple logistic regression

```
collapseStateSpace <- function(classification){
 levels(classification) <- sub("hard default", "default", levels(classification))
 levels(classification) <- sub("soft default", "default", levels(classification))
 levels(classification) <- sub("acquisition", "exit", levels(classification))
 levels(classification) <- sub("delisting", "exit", levels(classification))
 classification <- factor(classification)
}
createInputs <- function(economy, economyName){
 load(paste0(home, "Data/XY/", economy, "dataXY.RData"))
 print(c("creating logit inputs for economy ", economyName))
 m <- dataXY; rm(dataXY); gc()
 m$classification <- collapseStateSpace(m$classification)
 standardisedVars <- as.matrix(m[,names(m) %in% c("indexReturn","r3month","dtdAvg","
      cashTaAvg","niTaAvg","sizeAvg","mktBook","sigma","dtdDiff","cashTaDiff","
      niTaDiff","sizeDiff")])
 standardisedVarNames <- colnames(standardisedVars)
 colStdDevs <- apply(standardisedVars, 2, sd)
 standardisedVars <- standardisedVars %*% diag(1/colStdDevs)
 colnames(standardisedVars) <- standardisedVarNames
 otherVars <- m[,names(m) %in% c("idRMI","year","classification")]
 m <- cbind(otherVars, as.data.frame(standardisedVars))
 # separate data into two parts viz. up to cutoff year and post the cutoff year. Use
      the latter for out-of-sample tests
 flag <- as.numeric(as.character(m$year)) <= cutoffYear
 outOfSample <- m[!flag,]
 if(nrow(outOfSample) > 0) save(outOfSample, file=paste0(home,"Intermediates/
      Simulation/Inputs/OutOfSample/",economy,".RData"))
 rm(outOfSample); gc()
 m <- m[flag,]
 if(nrow(m) > 0) save(m , file=paste0(home,"Intermediates/LogisticRegression/Inputs/",
      economy,".RData"))
}
estimateLogitModel <- function(m){
 library(mlogit)
 d <- mlogit.data(m, shape = "wide", choice = "classification")
 rm(m); gc()
 # fullModel is a boolean constant defined in the main program
 if(fullModel) {
  f <- mFormula(classification ~ 1|indexReturn+r3month+dtdAvg+dtdDiff+cashTaAvg+
       cashTaDiff+niTaAvg+niTaDiff+sizeAvg+sizeDiff+mktBook+sigma)}
 else {
  f <- mFormula(classification ~ 1|indexReturn+r3month+dtdAvg+cashTaAvg+niTaAvg+
       sizeAvg+mktBook+sigma)}
 # note the formula does not control for year or firm id. 2 reasons for this are
 # first the estimation time and RAM being limited, second that prediction out of
      sample is difficult with new year and new firms
 fit <- mlogit(formula = f , data = d, reflevel = "survival")
}
outputLogitFit <- function(fit, economy, economyName){
 print(c("estimated logit model for economy ", economyName))
 sink(paste0(home,"Intermediates/LogisticRegression/Outputs/",economyName,".txt"))
 print(economyName)
 print(summary(fit))
```

```
 sink()
 save(fit, file=paste0(home,"/Intermediates/LogisticRegression/Outputs/",economy,".
      RData"))
}
fitLogisticRegression <- function(){
 for(i in 1:length(economies)){
  economy <- economies[i]
  economyName <- names(economies)[i]
   createInputs(economy, economyName); gc()
  load(paste0(home, "Intermediates/LogisticRegression/Inputs/",economy,".RData"))
  fit <- estimateLogitModel(m)
  rm(m); gc()
  outputLogitFit(fit, economy, economyName)
  rm(fit); gc()
 }
}
```

2.4 R Code for PD estimation

```
sampleSizeYtoP <<- 1000
sampleSizeFirms <<- 500
numSimulations <<- 1000
horizonMonths <<- 12
# if sampleSizeFirms chosen is larger than set of firms to sample from then
# will produce NAs from the rbinom command of function samplePortfolio() due to prob
     being > 1
dropLowPDs <<- T

extractFitInfo <- function(economy){
 library(mlogit)
 load(paste0(home,"Intermediates/LogisticRegression/Outputs/",economy,".RData"))
 if(fullModel){numCoeffs <- 13} else {numCoeffs <- 9}
 bHat <- fit$coefficients[1:numCoeffs *2 - 1]; sigma <- vcov(fit)[1:numCoeffs *2 - 1,
      1:numCoeffs *2 - 1]
 return(fitInfo = list(bHat = bHat, sigma = sigma))}

computeLogOddsMoments <- function(economy, bHat, sigma){
 load(paste0(home,"Intermediates/Simulation/Inputs/OutOfSample/",economy,".RData"))
 idRMI <- factor(outOfSample$idRMI)
 if(fullModel){xNew <- as.matrix(cbind(1,outOfSample[,-c(1:3)])) }
 else{ xNew <- as.matrix(cbind(1,outOfSample[,-c(1:3, seq(7,13,2))]))}
 yHat <- xNew %*% as.matrix(bHat)
 rm(outOfSample); gc()
 # computation below could be done in 1 step as SE <- sqrt(diag(xNew %*% sigma %*% t(
      xNew)))
 # however doing it that way requires too much RAM so we economise
 m1 <- xNew %*% sigma; n <- nrow(sigma); rm(sigma); gc()
 m2 <- cbind(m1, xNew); rm(m1, xNew); gc()
 SE <- apply(m2, 1, function(row){sqrt(sum(row[1:n]*row[(n+1):(2*n)]))}); rm(m2); gc()
 logOddsMoments <- data.frame(idRMI = idRMI, yHat = yHat, SE = SE); rm(idRMI, yHat, SE
      ); gc()
 save(logOddsMoments, file= paste0(home,"Intermediates/Simulation/Inputs/
      LogOddsMoments/", economy,".RData"))
 return(logOddsMoments)}

computeProbabilityMoments <- function(economy, logOddsMoments){
 ySample <- mapply(rnorm, n = sampleSizeYtoP, mean = logOddsMoments$yHat, sd =
      logOddsMoments$SE); gc()
 pSample <- 1/(1+exp(-ySample));pMeans <- colMeans(pSample);pStdDevs <- apply(pSample,
       2, sd)
 pMoments <- data.frame(idRMI = logOddsMoments$idRMI, pMeans, pStdDevs, ratiosMeanBySE
       = pMeans/pStdDevs)
 firmPDs <- aggregate(pMeans~idRMI, data=pMoments, median)
 firmRatiosMeanBySE <- aggregate(ratiosMeanBySE~idRMI, data=pMoments, median)
 probabilityMoments <- merge(firmPDs, firmRatiosMeanBySE)
 probabilityMoments$pSEs <- probabilityMoments$pMeans/probabilityMoments$
      ratiosMeanBySE
 probabilityMoments$ratiosMeanBySE <- NULL
 probabilityMoments[,-1] <- probabilityMoments[,-1] * horizonMonths
 save(probabilityMoments, file= paste0(home,"Intermediates/Simulation/Outputs/
      ProbabilityMoments/", economy,".RData"))}
```

```
generatePDSEs <- function(economy){
 print(economy)
 print("generating PDSE for the above economy")
 fitInfo <- extractFitInfo(economy)
 bHat <- fitInfo$bHat; sigma <- fitInfo$sigma; rm(fitInfo); gc()
 logOddsMoments <- computeLogOddsMoments(economy, bHat, sigma); gc()
 computeProbabilityMoments(economy, logOddsMoments)}

simulateProbabilities <- function(){
 for(economy in economies) {generatePDSEs(economy); gc()}}

mergeSectorInfo <- function(m){
 companyMaster <- read.csv(paste0(home, "Data/Company Master.csv"))[,c(2,10)]
 m <- merge(m,companyMaster);
 m <- m[! m$sector %in% c("NULL", "Diversified", "Funds", "Government", "Energy", "
      Utilities"),]
 m$sector <- factor(m$sector)
 return(m)}

createSamplePortfolio <- function(economy, simulation){
 load(paste0(home,"Intermediates/Simulation/Outputs/ProbabilityMoments/", economy,".
      RData"))
 portfolio <- mergeSectorInfo(probabilityMoments)
 medianPD <- median(portfolio$pMeans)
 if(dropLowPDs) portfolio <- portfolio[portfolio$pMeans > medianPD,]
 portfolio$ratiosMeanBySE <- NULL
 selectFlag <- rbinom(n = nrow(portfolio), size=1, prob=sampleSizeFirms/nrow(portfolio
      )) > 0
 selection <- portfolio[selectFlag,]
 selection$sector <- factor(selection$sector)
 portfolio <-  selection; rm(selection); gc()
 portfolioFileName <- paste0(home,"Intermediates/Simulation/Outputs/SamplePortfolios/"
      , economy, ".", simulation, ".RData")
 save(portfolio, file= portfolioFileName)}

simulatePortfolios <- function(){
 for(economy in economies) {
  for(simulation in 1:numSimulations){
   createSamplePortfolio(economy, simulation); gc()}}}
```

2.5 R Code for portfolio loss distributions

```
library(reshape2)
library(crp.CSFP)
# exposures at default for each counterparty are approximately 100 currency units
eadScale <<- 100
# when creating input for CreditRisk+ model, the original portfolio derived from real
     data (of size approx 100)
# is (conditionally) resampled with replacement. The new portfolio is
     portfolioSizeMultiplier times the oiriginal.
portfolioSizeMultiplier <- 20
# Note that the total portfolio value is approximately approximately 1 million
     currency units because
# approximately 500 firms, each allocated an exposure of approximately 100, and this
     set replicated 20 times
# unit for loss discretisation in the creditrisk+ framework
lossUnit <<- 10
# confidence level at which the VaR should be calculated
confidenceLevel <<- 0.999
# plot scale for losses (on the x axis)
plotScale <<- 100
# Note that as the total exposure is approx 1 million, the plot scale shows approx
     loss proportion in basis points
# for now we assume constant recovery rate of 40%, which translates to constant LGD =
     0.6
# However, the recovery rates can be easily made sector dependent later without
     introducing too many complications
constLGD <<- 0.6
# Set a threshold for minimum obligor risk in the portfolio chosen
# if this number is close to zero then all obligors equally likely
# if this number is high (max 80 approx) then all obligors are high risk
ratingThreshold <<- 0
```

```
createInputProbabilities <- function(simulationId){
 load(paste0(home, "Intermediates/Simulation/Outputs/SamplePortfolios/", simulationId,
      ".RData"))
 ratings <- rank(portfolio$pMeans, ties.method = "random")
 probabilities <- cbind(ratings, portfolio$pMeans, portfolio$pSEs)
 colnames(probabilities) <- c("RATING", "PD", "SD")
 probabilities <- probabilities[order(ratings),]
 folderName <- paste0(home, "Intermediates/CreditRisk+/Inputs/", simulationId,"/")
 if(!dir.exists(folderName)) dir.create(folderName)
 write.csv(probabilities, file=paste0(folderName,"rating_pd.csv"), row.names=F)
}

createInputPortfolioSource <- function(simulationId){
 load(paste0(home, "Intermediates/Simulation/Outputs/SamplePortfolios/",simulationId,"
      .RData"))
 numExposures <- nrow(portfolio)
 numSectors <- length(levels(portfolio$sector))
 sectorNames <- paste0("S",1:numSectors)
 portfolio$value <- T
 wide <- dcast(portfolio, idRMI + pMeans + pSEs ~ sector, fill=F)
 wide$ratings <- rank(wide$pMeans, ties.method = "random")
 rm(portfolio); gc()
 for(i in 1:numSectors) {wide[,3+i] <- as.numeric(wide[,3+i])}
 ead <- 1 + round((runif(numExposures,-eadScale*0.1,+eadScale*0.1) + (eadScale)))
 portfolio <- cbind(0,wide$idRMI, ead , constLGD, 1, wide$ratings, wide[,3 + 1:
      numSectors])
 rm(wide); gc()
 colnames(portfolio) <- c("CPnumber", "CPname", "exposure", "lgd", "maturity", "rating
      ", sectorNames)
 portfolio$CPnumber <- 1:nrow(portfolio)
 return(portfolio)
}

bootstrapPortfolio <- function(portfolio){
 # select only exposures with PD greater than a certain percentile
 selectedPortfolio <- portfolio[portfolio$rating > ratingThreshold,]
 if(nrow(selectedPortfolio) > 1){
  selectedPortfolio$CPnumber <- as.numeric(rownames(selectedPortfolio))
  newSize <- portfolioSizeMultiplier * nrow(selectedPortfolio)
  rowSample <- as.data.frame(sample(selectedPortfolio$CPnumber, size = newSize,
       replace=T))
  colnames(rowSample) <- "CPnumber"
  newPortfolio <- merge(rowSample, selectedPortfolio)
  rm(portfolio, selectedPortfolio); gc()
  newPortfolio$CPnumber <- 1:nrow(newPortfolio)
  newPortfolio$CPname <- paste0(newPortfolio$CPname,".",newPortfolio$CPnumber)
 }
 else{
 newPortfolio <- portfolio
 }
 return(newPortfolio)
}

createInputPortfolio <- function(simulationId){
 portfolioSource <- createInputPortfolioSource(simulationId)
 # create a template for the final portfolio (extra first row will be deleted later)
 portfolioFinal <- portfolioSource[1,]
 # standard input file format is columns CPnumber, CPname, exposure, lgd, maturity,
      rating, S1, S2, ...
 # so the number of sectors in this portfolio is got by counting total number of
      columns then subtracting 6
 numSectors <- ncol(portfolioSource) - 6
 for(i in 1:numSectors){
  portfolio <- portfolioSource[portfolioSource[,6+i] == 1,]
  portfolioBootstrapped <- bootstrapPortfolio(portfolio)
  portfolioFinal <- rbind(portfolioFinal, portfolioBootstrapped)
 }
 # delete the first (extra) row
 portfolioFinal <- portfolioFinal[-1,]
 folderName <- paste0(home, "Intermediates/CreditRisk+/Inputs/", simulationId,"/")
 if(!dir.exists(folderName)) dir.create(folderName)
 write.csv(portfolioFinal, file=paste0(folderName,"portfolio.csv"), row.names=F)
 return(portfolioFinal)
}

createInputSectorVariances <- function(simulationId){
```

```
 probabilities <- read.csv(paste0(home, "Intermediates/CreditRisk+/Inputs/",
      simulationId,"/rating_pd.csv"))
 portfolio <- read.csv(paste0(home, "Intermediates/CreditRisk+/Inputs/", simulationId,
      "/portfolio.csv"))
 m <- merge(probabilities,portfolio, by.x = "RATING", by.y = "rating")
 weights <- as.matrix(m[,-(1:8)])
 varCP <- as.matrix(m$SD ^ 2)
 muCP <- as.matrix(m$PD)
 varSector <- t(weights) %*% varCP
 muSector <- t(weights) %*% muCP
 varSector <- as.data.frame(varSector/muSector)
 colnames(varSector) <- "Var"
 folderName <- paste0(home, "Intermediates/CreditRisk+/Inputs/",simulationId,"/")
 if(!dir.exists(folderName)) dir.create(folderName)
 write.csv(varSector, file=paste0(folderName,"pd_sector_var.csv"), row.names=F)
 return(varSector)
}

prepareCRP <- function(simulationId){
 createInputProbabilities(simulationId); gc()
 createInputPortfolio(simulationId); gc()
 createInputSectorVariances(simulationId); gc()
}

runCRP <- function(economy, simulationId, statsList){
 print(paste0("Simulation ", simulationId))
 path.in <- paste0(home, "Intermediates/CreditRisk+/Inputs/", simulationId,"/")
 portfolio <- read.csv(paste0(path.in, "portfolio.csv"))
 totalExposure <- sum(portfolio$exposure)

 print("init"); cr <- init(path.in = path.in, loss.unit = lossUnit, alpha =
      confidenceLevel, PLOT.scale = plotScale, export.to.file = T, path.out = path.in)
 print("read"); cr <- read(cr)
 print("calc"); cr <- calc.portfolio.statistics(cr)
 print("loss"); cr <- loss.dist(cr)
 print("measure"); cr <- measure(cr)

 stats <- data.frame(EL = EL(cr), VaR = VaR(cr), ES = ES(cr), Total=totalExposure)
 stats$EL <- 10*round(stats$EL/10)
 #print(CDF(cr))
 statsList[[economy]] <- rbind(statsList[[economy]], stats)
 #print("plot"); plot(cr)
 print("save CDF"); cdf <- CDF(cr); save(cdf, file=paste0(home, "Intermediates/
      CreditRisk+/Outputs/CDFs/",simulationId,".RData"))
 return(statsList)
}

computeCreditLossDistributions <- function(){
 statsList <- list()
 stats <- data.frame(EL = 0, VaR = 0, ES = 0, Total = 0)
 for(economy in economies) {statsList[[economy]] <- stats}
 for(economy in economies){
  for(simulation in 1:numSimulations){
   simulationId <- paste0(economy,".", simulation)
   prepareCRP(simulationId); gc()
   statsList <- runCRP(economy, simulationId, statsList)
  }
 }
 for(economy in economies) {statsList[[economy]] <- statsList[[economy]][-1,]}
 save(statsList, file = paste0(home, "Intermediates/CreditRisk+/Outputs/statsList.
      RData"))
}
```

References

[1] R Barrell and S Gottschalk. "The Impacts of Capital Adequacy Requirements on Emerging Markets". Working paper, National Institute of Economic and Social Research. 2006.

[2] T. Bielecki and M. Rutkowski. *Credit Risk: Modeling, Valuation and Hedging*. Springer-Verlag, Berlin, 2004.

[3] F. Black and J. Cox. "Valuing corporate securities: some effects of bond indenture provisions". In: *Journal of Finance* 31 (1976), pp. 351–367.

[4] F. Black and M. Scholes. "The Pricing of Options and Corporate Liabilities". In: *Journal of Political Economy* 81 (1973), 637659.

[5] CreditSuisse. *CreditRisk+: A credit risk management framework*. 1997. URL: `http://www.csfb.com/creditrisk`.

[6] Yves Croissant. "Estimation of multinomial logit models in R : The mlogit Packages". In: (2013). URL: `https://cran.r-project.org/web/packages/mlogit/vignettes/mlogit.pdf`.

[7] P. Crosbie and J. Bohn. *Modeling Default Risk*. 2003. URL: `http://www.macs.hw.ac.uk/~mcneil/F79CR/Crosbie_Bohn.pdf`.

[8] D. Duffie and K. Singleton. "Modelling term structures of defaultable bonds". In: *Review of Financial Studies* 12 (1999), pp. 687–720.

[9] R. Geske. "The Valuation of Corporate Securities as Compound Options". In: *Journal of Financial and Quantitative Analysis* 12 (1977), 541552.

[10] R Gottschalk and S Griffith-Jones. "Review of Basel II Implementation in Low-Income Countries". Working paper, Institute of Development Studies, University of Sussex. 2006.

[11] S Griffith-Jones and S Spratt. “Will the proposed new Basel Capital Accord have a net negative effect on developing countries?” Working paper, Institute of Development Studies, University of Sussex. 2001.

[12] A. Henningsen and O. Toomet. “maxLik: A package for maximum likelihood estimation in R”. In: *Computational Statistics* 26 (3 2011), pp. 443–458.

[13] RiskMetrics (first published by J. P. Morgan. *CreditMetrics Technical document*. 1997. URL: `https://www.msci.com/resources/technical_documentation/CMTD1.pdf`.

[14] A. Kadam and P. Lenk. “Bayesian inference for issuer heterogeneity in credit ratings migration”. In: *Journal of Banking and Finance* 32.10 (2008), pp. 2267–2274.

[15] David Lando. *Credit Risk Modeling: Theory and Applications*. Princeton University Press, 2004.

[16] H. Leland. “Corporate Debt Value, Bond Covenants, and Optimal Capital Structure”. In: *Journal of Finance* 49 (1994), pp. 1213–1252.

[17] R. Litterman and T. Iben. “The Journal of Portfolio Management”. In: 17.3 (1991), pp. 52–64.

[18] F. Longstaff and E. Schwartz. “A simple approach to valuing risky fixed and floating rate debt”. In: *Journal of Finance* 50.3 (1995), pp. 789–819.

[19] D. Galai M. Crouhy and R. Mark. “A comparative analysis of current credit risk models”. In: *Journal of Banking and Finance* 24 (2000), pp. 59–117.

[20] D. Madan and H. Unal. “Pricing the risks of default”. In: *Review of Derivatives Research* 2 (2 1998), pp. 121–160.

[21] R. Merton. “On the Pricing of Corporate Debt: The Risk Structure of Interest Rates”. In: *Journal of Finance* 29 (1974), 449470.

[22] D. Lando R. Jarrow and S. Turnbull. “A Markov model for the term structure of credit risk spreads”. In: *Review of Financial Studies* 10 (1997), 481523.

[23] P. Protter R. Jarrow. "Structural vs. Reduced Form Models: A new information based perspective". In: *Journal of Investment Management* 2 (2 2004), 110.

[24] S. Minhaj S. Hai and R. Ahmed. "Implementation of Basel II: Issues, Challenges and Implications for Developing Countries". Working paper, University of Karachi. 2001.

[25] M. Tonveronachi. "Implications of Basel II for financial stability. Clouds are darker for developing countries". In: *PSL Quarterly Review* 62 (2009), pp. 121–146.

[26] J. Ward. "The new Basel accord and developing countries: problems and alternatives". CFAP Working Papers, No.4. University of Cambridge. 2002.

[27] C. Wieczerkowski. "Fourier Inversion Techniques for CreditRisk+". In: *CreditRisk+ in the banking Industry*. Ed. by M. Gundlach and F. Lehrbass. Springer-Verlag Berlin Heidelberg, 2004, pp. 45–68.