Parameter Uncertainty in Vasicek Model for Credit Risk

Martino Bernasconi
Laura Locatelli
Lorenzo Viola

Politecnico di Milano

2019

1 Introduction

The Vasicek model, refer to [BIS, 2005], is a Internal Rate Based (IRB) approach suggested by the BIS for capital requirement computations. This project will show that this model is extremely sensitive to parameter estimation and we will show the impact of this uncertainty in the capital requirement valuation.
After the 2008 Global Financial Crises, regulators put in place new frameworks to asses model governance within banking institutions, one important guidance, put in place by the FED, is the SR 11-7 which tells how financial institutions should govern and asses stability of internal models. This project is a toy example on a possible analysis of the Vasicek model, in particular we will focus of parameter estimation and briefly on sensitivity analysis.
In a financial institution every time a parameter is marked in a model, a special valuation group will take some parameter uncertainty reserve in order to account the fact that the marked parameters could be incorrectly marked, or estimated, by the model user.
Data about the subject is scarce and this cause problems in all techniques that tries to do inference about the parameter distribution. The data-sets used, cluster data about default events in year buckets, without specifying the function used in the weighing procedure.
This report is structured as following. First in chapter 2 we give a overview of the original problem we are dealing with, then in chapter 3 we define the alternative methodology chosen by [Baviera et al., 2019] and [Tarashev, 2010], an overview of the data is given in 4. In chapter 5 is discussed the distributional inference divided in classical and Bayesian in chapter 5.1 and 5.2 respectively. Results are discussed in chapter 7. In chapter 8 we briefly discuss the global sensitivity of the model, while in chapter 9 we describe the computational aspects of our work.

2 Problem formulation

The Vasicek model for credit risk is inferred from the hypothesis that a firm would defaults if it's firm value V_i goes below a threshold d at the final date T. Firm's values are then modeled as Gaussian random variable with a common risk factor Z, a idiosyncratic risk factor ϵ_i, both independent and Gaussian. Then the exogenous and idiosyncratic risk factor are correlated by a correlation parameter ρ.

$$V_i = \sqrt{\rho} Z + \sqrt{1-\rho}\epsilon_i$$

In this framework the loss function of the portfolio is

$$Loss = EAD \cdot LGD \cdot m \tag{1}$$

where EAD is the Exposure At Default, LGD is the Loss Given Default and m is the number of defaulted obligors. In our specification of the model the EAD is deterministic and fixed at $1/N$, LGD is a parameter marked by the model user and m is a random variable driven by the firm value process. Under this model assumptions the regulatory capital associated to this portfolio would be the following quantity:

$$\alpha RC := \alpha VaR - \mathbb{E}[L] \tag{2}$$

where L is the loss observed in the portfolio, αVaR is the α percentile of the loss distribution and $\mathbb{E}[L]$ the expected loss.

2.1 Large Homogeneous Portfolio

In the Large Portfolio assumption the obligors are considered homogeneous in the sense that the default thresholds and the recovery paid in case of defaults are the same for all obligors.
Thanks to [Gordy, 2002] we have a closed formula for RC in the case of perfect small granularity (LHP):

$$\alpha RC = (1 - RR)\mathcal{N}\left(\frac{\mathcal{N}^{-1}(DR) - \sqrt{\rho}\mathcal{N}^{-1}(1-\alpha)}{\sqrt{1-\rho}}\right) - DR\,(1 - RR) \tag{3}$$

where DR is the mean default rate, ρ the asset mean correlation and RR the average recovery rate of a single obligor.

2.2 Homogeneous Portfolio

In the case of a small portfolio we can compute explicitly $P(X \leq m)$ which is the probability of observing less then m defaults in our small portfolio.

$$P(X \leq m) = \int\limits_{-\infty}^{+\infty} \sum_{n=0}^{m} \binom{N_{ob}}{n} p(y)^n (1 - p(y))^{N_{ob}-n} \phi(y)\ dy \tag{4}$$

where $p(y)$ is the probability of default of a single obligor, $\phi(y)$ is the standard normal density function.
The from this distribution the VaR would be its α-quantile and the RC would easily follow from (2)

3 Alternative models

In the spirit of model risk the alternative model is a larger model that tries to estimate the impact of loosening some assumption. In this case the alternative model consider that the parameters could come from a distribution.

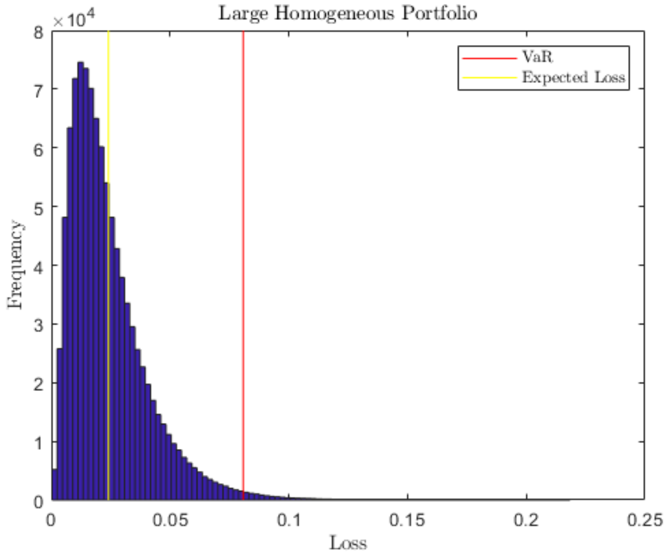

Figure 1: Common Loss distribution in LHP formulation

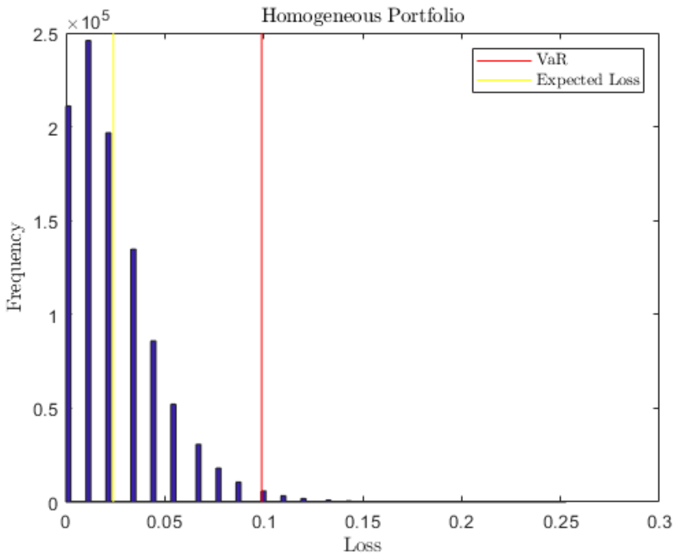

Figure 2: Common Loss distribution in HP formulation

In the case of a LHP (3) show that this a non-linear function of three parameters: DR, RR, ρ.
The first idea would be to infer a distribution from data and then computes with a MC the loss for each obligor and then take the percentile of the loss distribution. Note that in this case we are in reality dropping the assumption in Homogeneity rather then on parameter uncertainty since we

are considering N_{ob} obligors each with a different default probability, recovery and correlation with the overall market. Interesting enough this is both a model specification risk and a parameter uncertainty risk.
In the HP case we used a MC approach that simulates N_{sim} market situation (we have N_{sim} samples from the parameter distributions) and $N_{sim} \times N_{ob}$ firm values, computes N_{sim} losses and then takes the α-percentile of the loss distribution for the VaR.
The impact is then estimated in several ways, drawing one from the distribution and fixing the others or drawing two from a distribution and fixing the third and so on.

4 Data-set

The data-set used is the one reported in table 1. For each year from 1983 to 2017 we have one observation for the default rate of speculative grade issues, one for all issues and a recovery rate. The way this quantities are computed might impact the distribution of the data. Moreover we don't know what function is used to bucket data into a single figure per year.

The variables DRSpecultative and DRAll represents the default rate for speculative grate and total grade issuers in the corresponding year. Assuming a Gaussian Firm value we call DefaultBarrier the inverse of the normal cdf of the default rate.

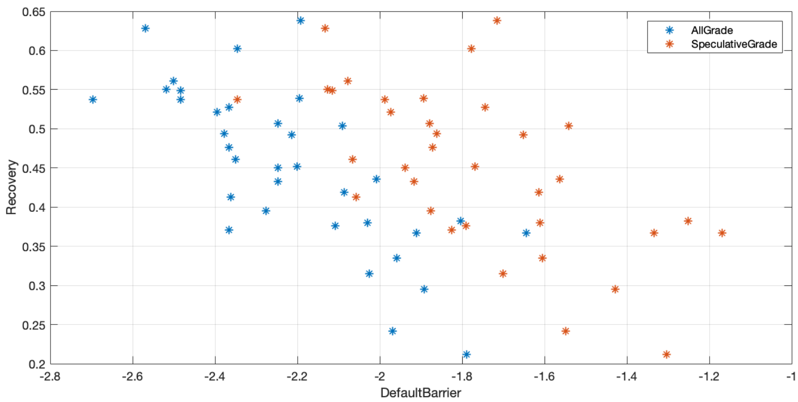

Figure 3: Scatter-plot for recoveries versus the default barriers

Oddly we are not give two different recovery rate for Speculative and All issue grade. Plotting the recoveries versus the default barriers we observe a clear linear relationship already observed by Altman in [Alt, 2019]. This is explained by the fact that in a distressed economy default rates are higher while goods are valued less and so we observe smaller recoveries. This fact is not used in the Alternative models described in [Baviera et al., 2019].

Year	DR_Speculative(%)	DR_All(%)	Recovery(%)
1983	4.06	0.9	52.7
1984	3.13	0.87	49.4
1985	3.77	0.95	60.2
1986	6.16	1.83	50.4
1987	4.31	1.42	63.8
1988	3.85	1.39	45.2
1989	5.91	2.23	43.6
1990	10.54	3.57	38.2
1991	9.1	2.8	36.7
1992	4.93	1.34	49.2
1993	3.4	0.9	37.1
1994	2.34	0.65	53.7
1995	3.06	0.9	47.6
1996	1.65	0.51	62.8
1997	1.89	0.62	56.1
1998	3.03	1.14	39.5
1999	5.36	2.12	38
2000	6.08	2.45	24.2
2001	9.63	3.68	21.2
2002	7.66	2.92	29.5
2003	5.33	1.85	41.9
2004	2.42	0.83	52.1
2005	1.72	0.65	54.9
2006	1.67	0.59	55
2007	0.95	0.35	53.7
2008	5.43	2.51	33.5
2009	12.11	5.01	36.7
2010	3.01	1.23	50.7
2011	1.98	0.91	41.3
2012	2.76	1.23	43.3
2013	2.63	1.23	45
2014	1.94	0.94	46.1
2015	3.67	1.75	37.6
2016	4.45	2.14	31.5
2017	2.91	1.41	53.9

Table 1: DataSet

As we can see we have no data regarding the correlation parameter ρ between firm values. This information is found in [Tarashev and Zhu, 2008] where it is inferred from publicly traded firms in the US from asset returns in the month of July 2006. The pairwise correlation seems to follow a beta distribution, hence in [Baviera et al., 2019] the distribution of the parameter ρ is assumed to follow the same distribution. This might be a far fetched assumption since using the pairwise correlation distribution for a single month is, in principle, very different then the firm value correlation we are interested in. Since the firm value is a not observable variable is natural to use the asset return as a proxy, the biggest drawback is the fact that we are using one month worth of data and this might bias the results, given that in principle the correlation distribution might not be stationary.

5 Distribution Inference

The two main approaches in order to estimate the parametric distributions are the classical statistical inference and Bayesian inference.
In classical statistical inference we assume stationary distributions for which we observe a sample, and then we have to estimate there parameters with a function of the sample, that are estimators.
Alternatively one can think as the the parameters are themselves random variable and having a conditional probability density function (pdf) of the observation given the parameter we can infer the posterior distribution of the parameters given the data

$$h(p|\hat{p}) \propto g(p)f(\hat{p}|p) \tag{5}$$

The pdf $g(p)$ is called *prior* and it is our assumption on the distribution of the parameter, while $f(\hat{p}|p)$ is the likelihood of observing $\hat{p}$ given the parameter ρ. This two are linked to the pdf of the true parameter given the observations $h(p|\hat{p})$ thanks to Bayes theorem.
Then from the posterior distribution we have to recover the class conditional density of the new data point $p(x|\hat{p}))$, with a straightforward application of the conditional expectation.

$$p(x|\hat{p}) = \int_{\mathbb{R}} f(x|p)h(p|\hat{p})dp \tag{6}$$

This step of the procedure (called Bayesian prediction) do not seems to be described in $[Bavieraet\ al., 2019]$.
In Bayesian inference is well know that if the sample size is large enough then the prior is irrelevant. In our case the sample size in somewhat small and hence changing the prior will impact the posterior. In [Baviera et al., 2019] and [Tarashev, 2010] the priors have been chosen as uninformative and hence the likelihood will dominate the posterior as explained in [Jackman, 2009]. For all this reasons using flat priors, we expect the roughly the same results from classical inference.

5.1 Frequentist Inference

Both for the Bayesian likelihood and for classical inference the parameters are assume to be distributed according to the following distributions:

- $d = \mathcal{N}^{-1}(DR) \sim \mathcal{N}(\hat{d}, \sigma_d)$
- $RR \sim \mathcal{N}(\overline{RR}, \sigma_{RR})$
- $\rho \sim Beta(\alpha_\rho, \beta_\rho)$

Using the Shapiro-Wilkinson test on the data of DefaultBarrier and Recovery we can accept the H_0 hypothesis of Gaussianity of the sample, with high p-value ($p_{RR} = 76.40\%$, $p_d = 75.50\%$). Having no data for the correlation we have to relay on the findings of [Tarashev and Zhu, 2008] and assume a Beta distribution with the parameter estimated by the authors. Using the NYSE (O) data-set of 36 firms from Jul 3 1962 to Dec 31 1984 (5651 days) we obtained $\binom{36}{2} = 630$ pairwise-correlations. Using this data-set we couldn't obtain the same distribution as in [Tarashev and Zhu, 2008].

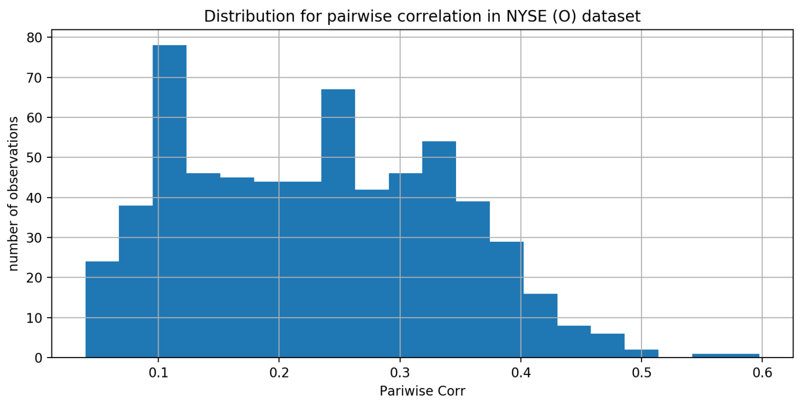

Figure 4: Pairwise correlation of NYSE(O) data-set

In particular the mean of this distribution is 23.14% well above the one found in [Tarashev and Zhu, 2008], which is 9.24%

Then the parameters of the distributions are inferred by simple estimators on the sample data. In particular we have two ways of choosing the mean of the distribution of the Default Barrier d. One is simply the sample mean of $\{\mathcal{N}^{-1}(DR_i)\}_i$ that gives as a result $\hat{d} = -1.7733$, another way suggested by [Baviera et al., 2019] is to take $\hat{d}$ such that

$$\mathbb{E}[\mathcal{N}(d)] = \frac{1}{N}\sum_{i=1}^{N} PD_i \tag{7}$$

which gives as a result $\hat{d} = -1.7731$ We do not observe any substantial difference in results in the two different approaches, and hence we will follow [Baviera et al., 2019].
Once the parameters of the distribution are fixed we will plug the independently drawn samples into the alternative model and then estimate the percentile increment with respect to the nominal model.
As in [Baviera et al., 2019] we will also use a deterministic correlation which is a function of the default probability as in [BIS, 2005]. The used function is a convex combination of 12% and 24% with exponential weights proportionally inverse the the default probabilities.

$$\rho(DR) = 0.12w(DR) + 0.24(1 - w(DR)) \tag{8}$$

with

$$w(DR) = \frac{1 - e^{-kDR}}{1 - e^{-k}}$$

k is a parameter that decides the decay of the exponential weights, in our case [BIS, 2005] set it to $k = 50$. This inverse relationship between default rates and correlation is explained in [BIS, 2005] by the fact that firm-specific information will impact asset prices more in near-default situations (low-correlation), rather then general economic information (high-correlation). We couldn't verify this claim with independent data.

5.2 Bayesian Inference

We said in chapter 5 that we can infer the parameter distribution given the observations by a Bayesian approach. Since we will use just uninformative priors and hence the posterior will be determined mostly by the likelihood. Moreover the likelihood, in most cases, will be evaluated just in one datapoint and so the overall effect will be just to invert the likelihood. If no information is coming from the prior and little information is coming from the sample we expect the Bayesian sampling pdf to be very similar to the inferential one.

5.2.1 Correlation

The prior distribution is taken as flat $g(\rho) = \mathbb{1}_{[0,1]}(\rho)$ while the likelihood $f(\hat{\rho}|\rho)$ is a beta distribution with parameters $\alpha(\rho)$ and $\beta(\rho)$. This give the following results for the parameters:

$$\begin{aligned} \alpha(\rho) &= \rho\frac{\rho(1-\rho) - \sigma^2}{\sigma^2} \\ \beta(\rho) &= (1-\rho)\frac{\rho(1-\rho) - \sigma^2}{\sigma^2} \end{aligned} \tag{9}$$

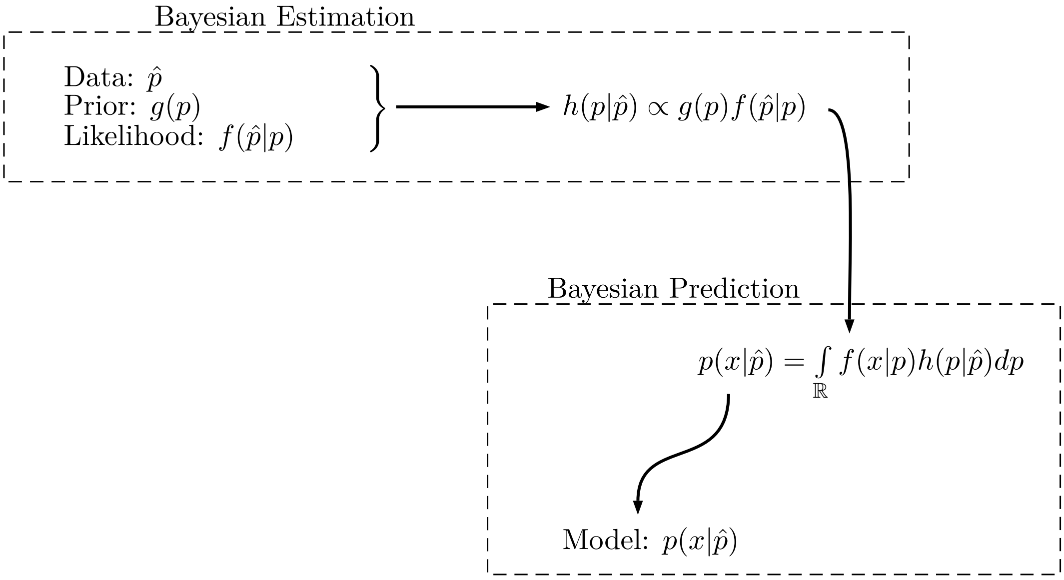

Figure 5: Bayesian prediction framework

where ρ and σ are the wanted mean and variance of the distribution as in [Tarashev and Zhu, 2008]. Hence the posterior can be written as

$$h(\hat{\rho}|\rho) \propto \mathbb{1}_{[0,1]}(\rho)Beta(\hat{\rho}|\alpha(\rho), \beta(\rho)) \tag{10}$$

and the only observation we have for the sample is the mean value in [Tarashev and Zhu, 2008] $\hat{\rho} = 9.24\%$. With just one observation and flat prior we are basically just using as posterior the likelihood as a function of the parameters. Moreover observe that the posterior is not a beta since is a function of the parameters.
In our opinion it is strange to use a flat prior (that indicates that all the information will come from the likelihood) with a sample of just one observation ($\hat{\rho} = 9.24\%$).
Then we recover the class conditional density as

$$p(x|\hat{\rho}) = \int_{\mathbb{R}} Beta(x|\rho, \sigma_\rho)h(\rho|\hat{\rho})d\rho \tag{11}$$

5.2.2 Default Barrier

In this case the the prior for the default barrier is $d \sim \mathcal{N}(0,1)$. This come from the uninformative flat prior for the default probability $DR \sim \mathcal{U}[0,1]$ and straight application of the Inverse Sampling Theorem. For the likelihood

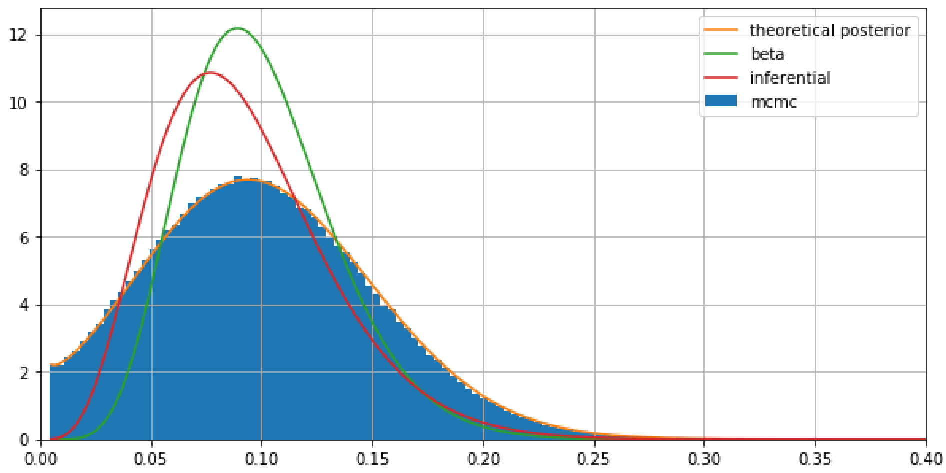

Figure 6: Distribution of the parameter ρ according to different approaches.

$f(\{d_i\}_i|d,\sigma_d)$ we used

$$f(\{d_i\}_i|d,\sigma_d) = \prod_{i=1}^{N} \phi(d_i|d,\sigma_d) \tag{12}$$

where $\phi(x|\mu,\sigma)$ is the normal pdf with mean μ and standard deviation σ and σ_d is the sample standard deviation of the sample $\{d_i\}_i = \{\mathcal{N}^{-1}(PD_i)\}_i$. Since the prior and the likelihood are conjugated we know that also the posterior $h(d|\{d_i\}_i)$ will be in the same parametric class of the prior and hence it will be normal with known mean and standard deviation given by:

$$\mu = \frac{\sum d_i}{N+\sigma_d^2} \qquad \sigma^2 = \frac{\sigma_d^2}{N+\sigma_d^2} \tag{13}$$

where N is the magnitude of our sample.
In [Baviera et al., 2019] we understood that the procedure used is somewhat different, in fact it seems that as observation was used just one observation which is $\hat{d}$ and the variance of the mean of the sample $\{d_i\}_i$ (even if $\hat{d}$ is not the mean of the sample) which is σ_d/N for a normal iid sample. The parameter of the posterior in this case are then

$$\mu = \frac{\hat{d}}{1+\sigma_d^2/N} \qquad \sigma^2 = \frac{\sigma_d^2}{N+\sigma_d^2} \tag{14}$$

And since $\hat{d} \approx \sum d_i/N$ the results obtained are very similar.
Formula di d hat?
Then the class conditional density is recovered as

$$p(x|\hat{d}) = \int_{d \in \mathbb{R}} \phi(x|d, \sigma_d) h(d|\hat{d}) \tag{15}$$

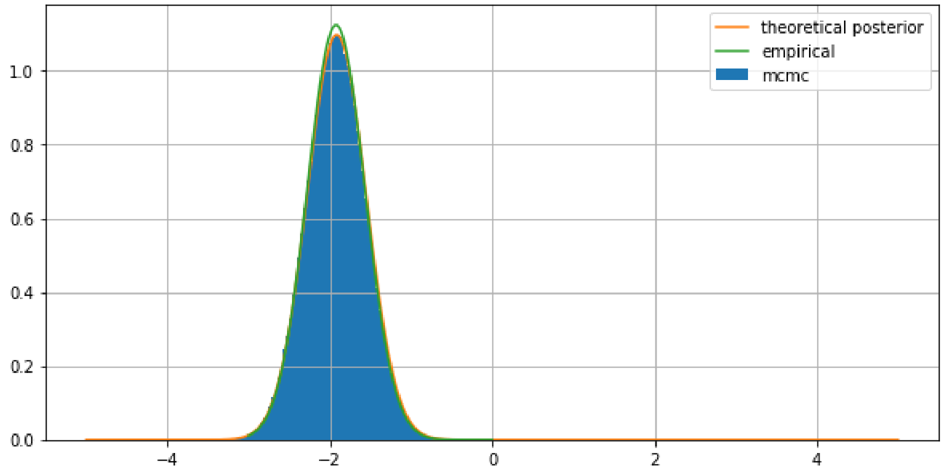

Figure 7: Distribution of the parameter d according to different approaches.

Observe that the class conditional density is close to the inferential. 7.

6 Cramer-Rao Lower Bound

In order to generate posterior distributions we always consider known variance. We might also want to use the Cramer-Rao lower bound for variance of unbiased estimators for the parameter, this would be a function of the unknown parameter p. It is defined as:

$$\sigma^2_{CR}(p) = -\frac{1}{T}\left(\mathbb{E}_y\left[\frac{\partial^2 LL(y|p)}{\partial p^2}\right]\right)^{-1} \tag{16}$$

where $LL(\cdot|p)$ is the log-likelihood function for the observed data and T the number of observation.

6.1 Correlation

As before we consider an uninformative prior $g(\rho) = \mathbb{1}_{[0,1]}(\rho)$ but this time the likelihood $f(\hat{\rho}|\rho)$ follows a beta distribution with parameters $\alpha(\rho)$ and $\beta(\rho)$ chosen such that the mean of the distribution $f(\cdot|\rho)$ is ρ and variance as the Cramer Rao lower bound. In the Homogeneous Portfolio case we have the closed formula in [Tarashev, 2010]

$$\sigma^2_{CR,\rho}(\rho) = \frac{2(1-\rho)^2(1+(N_{ob}-1)\rho)^2}{N_{ob}(N_{ob}-1)T} \tag{17}$$

where, since the data-set from which ρ is deducted covers one month time window we suppose to have 30 observation hence we plugged $T = 30$, while for the N_{ob} we choose 10891 since the sample analyzed by [Tarashev and Zhu, 2008] was composed by 10891 firms. To recover the class conditional density we computed it numerically as before

$$p(x|\hat{\rho}) = \int_{\mathbb{R}} Beta(x|\rho, \sigma_\rho) h(\rho|\hat{\rho}, \sigma_{CR,\rho}) d\rho \tag{18}$$

Note that the CR variance its used just in the posterior $h(\cdot, \hat{\rho}, \sigma_{CR,\rho})$ and in the model density we used the empirical variance of the sample found in [Tarashev and Zhu, 2008].

6.2 Default Rate

In this case the the prior for is $d \sim \mathcal{N}(0, 1)$. This come from the uninformative flat prior for the default probability $DR \sim \mathcal{U}[0, 1]$ and straight application of the Inverse Sampling Theorem.
As for the likelihood $f(\hat{d}|d, \rho) = \phi(\hat{d}|d, \sigma^2_{CR})$ where $\sigma^2_{CR}(d, \rho)$ is computed numerically as explained in chapter 9 since there are no closed formula. Then the class conditional density is computed as

$$p(x|\hat{d}) = \int_{d \in \mathbb{R}} \phi(x|d, \sigma_d) h(d|\hat{d}) \tag{19}$$

The resulting surface is shown in figure 8

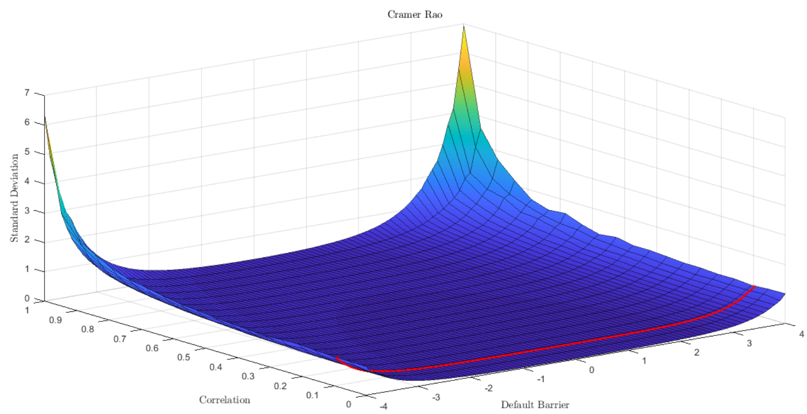

Figure 8: Surface for the standard deviation (Cramer Rao)

7 Results

Many of the results are obtained by some sort of Monte Carlo technique which in case we have to compute an expectation is known to converge as $\mathcal{O}(\sqrt{N})$. In our case we have to compute percentile of distributions and so we are much more susceptible to tail events and we claim that the rate of convergence will be much slower. Even though we haven't explored this path in full, we know that there are new techniques in Adaptive Sampling such as such as [Egloff and Leippold, 2010] to overcome this problem.
Another obvious thing to notice is that the Capital Requirement grows with the variance of the distribution we are sampling parameters from. This is due to the fact that from larger variance in the parameters follow a larger variance in the Loss.
Another striking observation is that the add-ons are in general very large in order of magnitude (even a 10% add-on is a very large figure). This is explained by the fact that we are looking at tail events and just few events would make the VaR to "jumps" to the next percentile.
In table 4 the correlation is the deterministic function of the Default Rate (8). We can see that the Capital Requirement is higher in this case since we are forcing the correlation to be in $\rho \in (12\%, 24\%)$ which is higher then the fixed correlation of $\rho = 9.24\%$ found in [Tarashev and Zhu, 2008].

7.1 Speculative grade

In table 5 we reported the Add-Ons considering the sensitivity with respect to different parameters. As we can see the Default Rate (linked to the default barrier d) is the most sensitive parameter in the model and provides the most variance to the output. This is somewhat alarming since this parameter would, in principle, be very hard to mark by the user. On the other hand, the Recovery Rate is not very responsible to the total output variance of the model: this parameter is much easier to mark, since it can be inferred with a little bit of expertise from financial statements.
Another interesting observation is that the correlation between parameters ρ is the least responsible in the sensitivity of the Capital Requirement. This contrasts with our *a priori* beliefs of the parameter sensitivities, and since this parameter is not observable we deemed to be relieved by this finding.
In table 9 we collected results obtained by adopting Bayesian Inference, comparing this results with 6 we can observe that the results are very similar, with exception with regards to the add-on of ρ. This happens because, as shown in figure 6, the class conditional density of rho is sensibly larger then in the inferential approach.

[1]Computed using closed formula $\alpha VaR = (1 - \overline{RR})\mathcal{N}\left(\frac{\hat{d} - \sqrt{\overline{\rho} + \sigma_d^2}\mathcal{N}^{-1}(1-\alpha)}{\sqrt{1-\overline{\rho}}}\right)$

RC	**99%**	**99,90%**
LHP	0,0564	0,0911
HP	0,0749	0,1189

Table 2: RC for LHP and HP

RC	**99%**	**99,90%**
LHP	0,0568	0,0919
HP	0,0749	0,1189

Table 3: RC for LHP and HP using MC simulations

RC	**99%**	**99,90%**
LHP	0,0738	0,1225
HP	0,0859	0,1409

Table 4: RC for LHP and HP using ρ Basel II correction

RC	**99%**	**99,90%**
d	0,0831	0,1405
$d^{[1]}$	0,0831	0,1395
RR	0,0612	0,1021
ρ	0,0588	0,1062
d, ρ	0,0841	0,1481
All	0,0888	0,1595

Add-on	**99%**	**99,90%**
d	47,33%	54,28%
$d^{[1]}$	47,25%	53,20%
RR	8,41%	12,04%
ρ	4,16%	16,60%
d, ρ	49,11%	62,54%
All	57,30%	75,08%

Table 5: RC and Add-on for LHP with frequentist inference

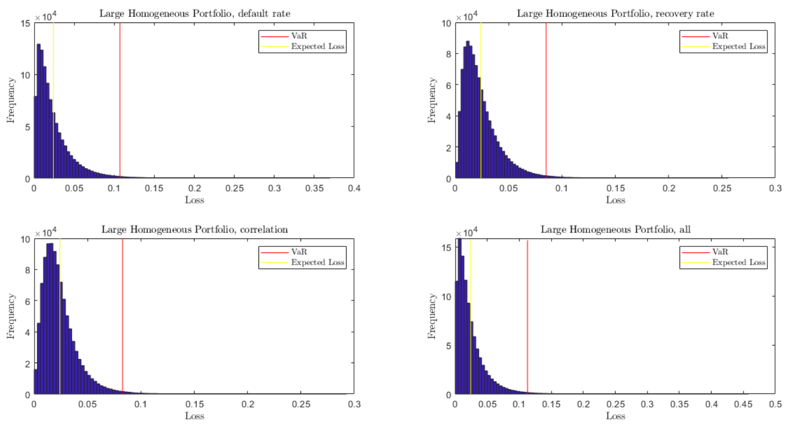

Figure 9: Typical Loss distribution in LHP Frequentist Framework

7.2 All grades

If we use the default rate computed considering all grading, we obtain the similar sensitivities that we obtain when we considered just specula-

RC	**99%**	**99,90%**
d	0,0969	0,1629
RR	0,0809	0,1315
ρ	0,0749	0,1299
d, ρ	0,0969	0,1739
All	0,1044	0,1841

Add-on	**99%**	**99,90%**
d	29,33%	36,97%
RR	7,93%	10,58%
ρ	0,01%	9,23%
d, ρ	29,34%	46,22%
All	39,30%	58,81%

Table 6: RC and Add-on for HP with frequentist inference

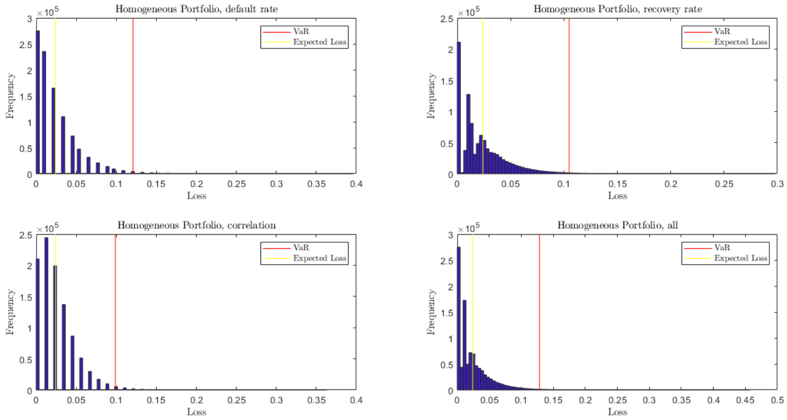

Figure 10: Typical Loss distribution in HP Frequentist Framework

RC	**99%**	**99,90%**
d	0,096	0,1621
RR	0,0786	0,1351
All	0,101	0,1769

Add-on	**99%**	**99,90%**
d	30,18%	32,38%
RR	6,57%	10,36%
All	36,88%	44,50%

Table 7: RC and Add-on for LHP with frequentist using ρ Basel II correction

RC	**99%**	**99,90%**
d	0,1079	0,1849
RR	0,0958	0,1616
All	0,1158	0,1995

Add-on	**99%**	**99,90%**
d	25,58%	31,20%
RR	11,51%	14,70%
All	34,80%	41,60%

Table 8: RC and Add-on for HP with frequentest using ρ Basel II correction

RC	**99%**	**99,90%**
d	0,0844	0,1421
ρ	0,0647	0,1195
All	0,0899	0,1612

Add-on	**99%**	**99,90%**
d	49,52%	56,00%
ρ	14,65%	31,22%
All	59,34%	76,**99%**

Table 9: RC and Add-on for LHP with Bayesian inference

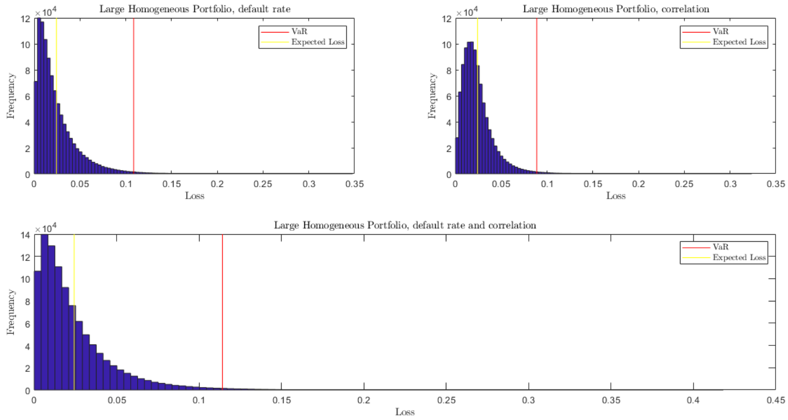

Figure 11: Loss distribution in LHP Bayesian Framework

RC	**99%**	**99,90%**
d	0,0966	0,1626
ρ	0,0859	0,1409
All	0,1077	0,1846

Add-on	**99%**	**99,90%**
d	28,97%	36,76%
ρ	14,66%	18,47%
All	43,66%	55,25%

Table 10: RC and Add-on for HP with Bayesian inference

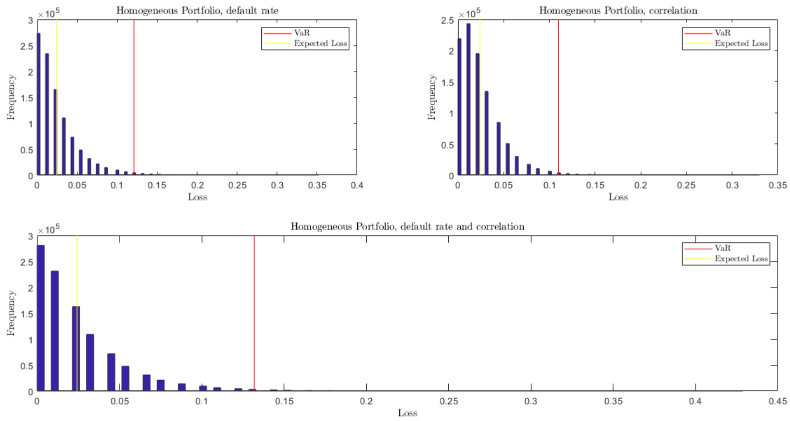

Figure 12: Loss distribution in HP Bayesian Framework

tive grades. Clearly the Capital Requirement is smaller in this situation as the default probabilities are smaller.

RC	**99%**	**99,90%**
d	0,0964	0,1624
ρ	0,0859	0,1408
All	0,1074	0,1843

Add-on	**99%**	**99,90%**
d	28,66%	36,52%
ρ	14,63%	18,47%
All	43,29%	54,96%

Table 11: RC and Add-on for HP with Bayesian inference using Cramer Rao lower bound as variance

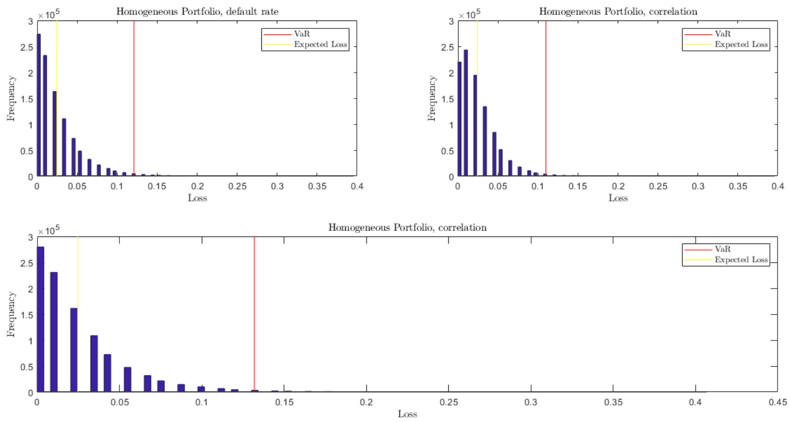

Figure 13: Loss distribution in HP Bayesian Framework with Cramer Rao Variance

RC	**99%**	**99,90%**
LHP	0,0272	0,0477
HP	0,0462	0,0681

Table 12: RC for LHP and HP

RC	**99%**	**99,90%**
LHP	0,0274	0,0478
HP	0,0462	0,0682

Table 13: RC for LHP and HP using MC simulations

RC	**99%**	**99,90%**
LHP	0,0451	0,0863
HP	0,0572	0,1012

Table 14: RC for LHP and HP using ρ Basel II correction

8 Extra

In principle we are estimating the sensitivity of a black box function with respect to some input parameter. To do this, first, we could do numerical

RC	**99%**	**99,90%**	**Add-on**	**99%**	**99,90%**
d	0,0385	0,0715	d	41,36%	50,02%
$d^{[1]}$	0,0384	0,0714	$d^{[1]}$	41,06%	49,81%
RR	0,0289	0,0517	RR	6,15%	8,44%
ρ	0,0281	0,0563	ρ	3,14%	18,12%
d, ρ	0,0388	0,077	d, ρ	42,44%	61,51%
All	0,0404	0,0816	All	48,42%	71,15%

Table 15: RC and Add-on for LHP with frequentist inference

RC	**99%**	**99,90%**	**Add-on**	**99%**	**99,90%**
d	0,0572	0,0902	d	23,82%	32,26%
RR	0,0457	0,0781	RR	-1,09%	14,55%
ρ	0,0462	0,0792	ρ	0,01%	16,13%
d, ρ	0,0572	0,0902	d, ρ	23,83%	30,25%
All	0,0544	0,1025	All	18,10%	50,37%

Table 16: RC and Add-on for HP with frequentist inference

RC	**99%**	**99,90%**	**Add-on**	**99%**	**99,90%**
d	0,0521	0,0979	d	15,52%	13,40%
RR	0,0466	0,0915	RR	3,44%	5,97%
All	0,0536	0,1044	All	18,82%	20,91%

Table 17: RC and Add-on for LHP with frequentist using ρ Basel II correction

RC	**99%**	**99,90%**	**Add-on**	**99%**	**99,90%**
d	0,0682	0,1122	d	19,22%	10,88%
RR	0,06	0,1123	RR	4,94%	11,02%
All	0,0664	0,1242	All	16,13%	22,78%

Table 18: RC and Add-on for HP with frequentist using ρ Basel II correction

RC	**99%**	**99,90%**	**Add-on**	**99%**	**99,90%**
d	0,039	0,0731	d	42,96%	53,34%
ρ	0,0312	0,0631	ρ	14,36%	32,25%
All	0,0416	0,0846	All	52,66%	77,41%

Table 19: RC and Add-on for LHP with Bayesian inference

partial derivatives with respect to the parameter of interest fixing the others. The problem with this approach is that the point in which we fix the other parameters can impact by a large magnitude the result of the partial derivative.
In the 1980, to overcome what is known as the local-limitation, one has

RC	**99%**	**99,90%**
d	0,0571	0,09
ρ	0,0462	0,0792
All	0,0571	0,101

Add-on	**99%**	**99,90%**
d	23,55%	32,06%
ρ	0,01%	16,11%
All	23,56%	48,19%

Table 20: RC and Add-on for HP with Bayesian inference

RC	**99%**	**99,90%**
d	0,0568	0,1008
ρ	0,0462	0,0792
All	0,0568	0,1008

Add-on	**99%**	**99,90%**
d	22,94%	47,79%
ρ	0,00%	16,10%
All	22,91%	47,81%

Table 21: RC and Add-on for HP with Bayesian inference using Cramer Rao lower bound as variance

been developed the Global Sensitivity Analysis (GSA) which aims to identify what inputs are the most influential in the model. One of the techniques of GSA tries to identify the most influential inputs by looking at the variance of the outputs of a function $f : \mathbb{R}^d \to \mathbb{R}$: this leads to the first order Sobol Indices. Here we summarize the main ideas behind it, for a more complete look see [Sal, 2004] and [Iooss and Lemaître, 2015].
Let $Y = f(X_1, .., X_d)$ be our model function and $X_1, \ldots, X_n$ be input random variables, possibly correlated, with support on $[0, 1]^d$.
The rationale is that an input X_i will be more influential as smaller the conditional variance of the output Y fixing the true unknown value X_i, if the conditional variance of Y is reduced just by condition on the value $X_i = x^*$ then this means that the output Y is very sensitive to the the input X_i and most of the variance of Y will be due to X_i. Formally we will look at

$$Var(Y|X_i = x^*)$$

but since we have no way of knowing the value x^* we will mute this dependency by integrating over the values x^*. Hence we will look at the value:

$$\mathbb{E}[Var(Y|X_i)] \tag{20}$$

Writing the variance decomposition of the output Y we can obtain the following relationship

$$Var(Y) = Var[\mathbb{E}(Y|X_i)] + \mathbb{E}[Var(Y|X_i)] \tag{21}$$

from which we can understand that highly relevant inputs X_i will have a greater first term $Var[\mathbb{E}(Y|X_i)]$ which can indeed be interpreted as the expected reduction in variance if we could fix the parameter X_i.

Hence we call first order Sobol indices the quantities

$$S_i = \frac{Var[\mathbb{E}(Y|X_i)]}{Var(Y)} \tag{22}$$

In order to compute analytically these values a common approach is to use MC simulations, hence we simulate a matrix of parameters for the model, drawn from a given distribution.
Fixing once at the time the parameter in a discrete grid for x^* in computationally inefficient and in order to efficiently compute those indices there are many tricks and in [Saltelli et al., 2010] we can find a complete survey and more interesting interpretations of the Sobol indices that involves the Hoeffding ANOVA functional decomposition.

We applied this methodology to our problem of capital requirement in the LHP case. Hence the functions considered are 3 function of DR, RR, ρ. To simulate the parameters we used a multivariate normal distribution for the joint distribution of the default barrier $d = \mathcal{N}^{-1}(DR)$ and the recovery rate RR, in order to exploit the linear relationship we mentioned in chapter 4, and an independent beta distribution for the correlation ρ.
The results are reported in the following where we also reported the normalized Sobol indices (so that they sum to 1). Normalization is needed since Sobol's indices $S1$ are just the first order sensitives and the little leftover of the variance is explained by second order sensitivities.

	DR	RR	ρ
S_1	36.84	7.99	23.16
NS_1	53.76	12.25	33.99

Table 22: Sobol indices for LHP (%)

Dropping the correlation between the default barrier and the recovery we obtain the following results.

	DR	RR	ρ
S_1	51.24	12.08	30.26
NS_1	54.75	12.91	32.33

Table 23: Sobol indices for LHP (%). Independent parameters

In either case we can see that the DR parameter is by far the most sensitive in the capital requirement function. A drawback a this analysis is that we do not have a dollar value for the risk associated to a parameter and hence it is unsuited to take reserves for parameter uncertainty. On the other hand this analysis might be interesting since we are not comparing two model and computing a percent error between the two, but we are looking directly at the amount of uncertainty coming from a parameter, hence we should, in theory, allocate proportional project budget to the estimation of more sensitive parameters.

9 Code

9.1 Nominal Model

To compute the Capital Required for LHP portfolio we implement the formula (3) in `CapitalRequirementNominalLHP`.
In `CapitalRequirementNominalHP` we just implemented the closed formula (4).

9.2 Alternative Model

In `CapitalRequirementAlternativeLHP` we had to use MC simulations of the systematic risk Z and the simulations of the variables in which we are interested (assuming they share the same size), sampled from a posterior or frequentist distribution. Then we compute the loss distribution and its percentile.
The same approach is followed with `CapitalRequirementAlternativeHP` where we need to give in input both the systematic risk factor and the individual risk factor.

9.3 Bayesian inference

9.3.1 Posterior distribution of correlation fixing variance equal to empiric one

To compute the posterior distribution of ρ, we select an equispaced vector of ρ, points on which we want to evaluate it, and a parameter σ_ρ^2, the empirical variance. Then in the function `betaParameter` we computes α and β parameters of a beta distribution following the equation 9.
In particular first we check the inputs in order to guarantee that α and β will be both strictly positive. Then with `posteriorDistributionRho` we compute the posterior distribution of ρ according to the formula (10). Finally we simulate our random variable via the function `samplingFromPosterior`, which uses the Inverse Sampling Theorem and interpolation in order to sample from the given pdf.

9.3.2 Posterior distribution of default barrier fixing variance equal to empiric one

Since the conjugated property entails that in this case the posterior distribution of d is again a normal distribution, hence to simulate the value of d we can use built-ins.

9.3.3 Posterior distribution of the correlation using Cramer Rao as variance

To compute σ_{CR} we implement the function `CramerRao_rho` that follow the formula (17) selecting as input the number of firms $N_{ob} = 10891$ as in [Tarashev and Zhu, 2008], and $T = 30$ as in discussed in chapter 6.1.

9.3.4 Posterior distribution of default barrier using Cramer Rao as variance

As there is not a closed formula, in order to compute the Cramer Rao lower bound:

$$\sigma^2_{CR,d}(d,\rho) = -\frac{1}{T}\left(\mathbb{E}_m\left[\frac{\partial^2 LL(m|d,\rho)}{\partial d^2}\right]\right)^{-1} \tag{23}$$

we have to calculate the expected value second derivative of the log-likelihood numerically. In particular the likelihood is:

$$P(X=m) = \int_{-\infty}^{+\infty} \binom{N_{ob}}{m} p(y)^m (1-p(y))^{N_{ob}-m} \phi(y)\ dy \tag{24}$$

where m is the number of defaults and $p(y)=\mathcal{N}\left(\frac{d-\sqrt{\rho}y}{\sqrt{1-\rho}}\right)$.

Thanks to the function `CramerRao_d` which, given the values of d and ρ (they can be vectors), computes the log-likelihood as defined above (with $T = 35$, the length of time series) and then numerically its second derivative with respect to d in the following way:

$$f''(x) = \frac{f(x+\epsilon) - 2f(x) + f(x-\epsilon)}{\epsilon^2} \tag{25}$$

In particular, we define a function handle which depends on the auxiliary variable s, that can assume values $-1, 0, 1$: it help us to compute the three addends of (25). So we construct a three dimensional matrix for each value of s where the $P_{i,j,k}$ element corresponds to the log-likelihood function evaluated in the correlation, default barrier and number of default respectively. Finally we compute the expected value of the second derivative of the log-likelihood on the variable m. In the end the function returns a matrix of the Cramer Rao limit function of the grid of the input ρ and d. To compute the posterior distribution of d and the sample we follow the analogous procedure described in chapter 9.3.1, but, since the Cramer Rao variance is the variance of the estimator $\hat{d}$ we use the function `posteriorDistributionD` which consider as observed parameters only $\hat{d}$.

This workaround help to improve drastically the computational time of procedure. For example in python we could reproduce this technique since python do not allows to compute integral of array valued functions and as a consequence it's painfully slow.

9.3.5 Posterior distribution using nested Monte Carlo

There is a dependence between the default threshold d distribution and the correlation value ρ, namely the Cramer Rao variance of the estimator $\hat{d}$ is function of ρ as seen in (23)

To take into account this link, we decide to perform nested Monte Carlo simulation. First of all, thank to the function `samplingFromPosterior` we simulate $N_{sim,\rho}$ ρ given its posterior distribution. Then, fixing a ρ, we estimate the corresponding Cramer Rao variance using `CramerRao_d`, determine the posterior law of d as described in chapter 5.2 and again `samplingFromPosterior` is used to simulate $N_{sim,d}$ value of d, which are associated to the fixed ρ. This procedure is repeated for every ρ, then we globally obtain $N_{sim,\rho} \cdot N_{sim,d}$ simulations.

9.3.6 Markov Chain Monte Carlo

In order to benchmark the sampling from posterior methodology of the Bayesian inference implemented with `saplingFromPosterior` we implemented the Metropolis Algorithm that allow us to sample from a distribution just having a function f proportional to the pdf. The algorithm samples a proposed x^* from a symmetrical distribution centered in x_i, and accept the proposal sample with a probability proportional to f. If the x^* is accepted then a new sample x_{i+1} is drawn and the procedure is repeated. In the limit the acceptance ratio will converge to the pdf we would like to draw from.

9.3.7 Class Conditional Density

Once we obtained the posterior distribution of the parameter by applying the Bayes Rule, we have to compute the pdf of the data (class conditional density). This is a weight of all parametric models and the weighting is done by the posterior.
This is implemented in `BayesianPredition` which take in input the vector on which we want to defined the output pdf, the vector on which the posterior is defined and a handle function the represent all parametric models.

References

[Sal, 2004] (2004). Sensitivity analysis in practice: a guide to assessing scientific models.

[BIS, 2005] (2005). An explanatory note on the basel II IRB risk weight functions - july 2005. page 19.

[Alt, 2019] (2019). The link between default and recovery rates: Theory, empirical evidence, and implications. page 26.

[Baviera et al., 2019] Baviera, R., Bianchi, G., and Canevisio, A. (2019). Parameters uncertainty impact on capital requirements. page 20.

[Egloff and Leippold, 2010] Egloff, D. and Leippold, M. (2010). Quantile estimation with adaptive importance sampling. *The Annals of Statistics*, 38(2):1244–1278. arXiv: 1002.4946.

[Gordy, 2002] Gordy, M. B. (2002). A risk-factor model foundation for ratings-based bank capital rules. page 35.

[Iooss and Lemaître, 2015] Iooss, B. and Lemaître, P. (2015). A Review on Global Sensitivity Analysis Methods. In Dellino, G. and Meloni, C., editors, *Uncertainty Management in Simulation-Optimization of Complex Systems*, volume 59, pages 101–122. Springer US, Boston, MA.

[Jackman, 2009] Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. Wiley Series in Probability and Statistics. Wiley.

[Saltelli et al., 2010] Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270.

[Tarashev, 2010] Tarashev, N. (2010). Measuring portfolio credit risk correctly: Why parameter uncertainty matters. 34(9):2065–2076.

[Tarashev and Zhu, 2008] Tarashev, N. and Zhu, H. (2008). Specification and calibration errors in measures of portfolio credit risk: The case of the ASRF model. 4(2):45.