Modeling credit spreads: The cross-section method revised

Aurélien Vermylen

Abstract:

The cross-section method as proposed in [NOM01] has a drawback when used in applications different than estimating a CVA VaR. Indeed, we show that by optimizing in least-squares on log-spreads, the method tends to approximate the geometric average of observed spreads instead of the arithmetic average, which by the AM–GM inequality will always be higher. We then propose an alternative methodology that optimizes on simple differences instead of log-differences.

1. Introduction

Estimating the creditworthiness of issuers that do not have a liquid CDS quote or liquid bond issues is a challenge for every institution that needs to determine the correct price or adequate level of provisioning for a particular deal with this issuer. One solution, as proposed by the EBA, is to categorize issuers by their defining characteristics and to take averages of their spreads in each of these categories (known as the "intersection" method). This offers a solution for issuers that have comparables, that is, other firms with the same characteristics. However, some issuers do not have this set of comparable firms from which it is possible to estimate a "proxy" spread.
The recently published cross-section method by Nomura [NOM01] describes an ingenious way to estimate spreads for these kind of issuers anyway. The method parametrizes each spread as the product of a few factors, which are simple real positive numbers. Each of these factors is representative of a particular characteristic of the spread's issuer. For instance, if we have a AA/financials/Belgian/senior issuance, we would estimate this spread by calculating the product of each factor representing this category: $Spread_{estimated} = Factor_{AA} \cdot Factor_{fin} \cdot Factor_{EUR} \cdot Factor_{sen}$. In the rest of this paper, we will call "AA", "financials", "Belgium" and "senior", categories, the combination of them (e.g. AA/financials/Belgium/senior), a bucket, and their group (e.g rating, sector, region, seniority), characteristics.

Notwithstanding the question of whether a spread is indeed fundamentally best parametrized as the product of factors, this method has a few advantages over intersection methods, and these are described in the paper. Obviously, the fact that once each factor is calibrated, every type of issuer's credit spread can be estimated, is the greatest advantage. Even if an issuer has no comparable firms with liquid quotes (no quotes in its bucket), one can just multiply the correct factors and obtain a spread. Another significant advantage is that we can take very granular categories in each characteristic of issuers. Indeed, when calibrating an intersection method, one must make sure that every bucket has sufficient amount of observable spreads so that the average is statistically significant. Indeed, if there are only three AA/financial/Belgian/senior liquid quotes, we will have to take a broader category definition of the region characteristic for example (e.g. Europe instead of Belgium). This requirement is not necessary for the cross-section method, since as long as the category Belgium has observable quotes, the method will estimate the general "level" of the Belgian factor from all these positions. And if there are no Belgian/financials/AAA observations, this is no problem, since the AAA factor will be estimated from all AAA positions and the same is true for the financials factor.

Other advantages over the intersection method can be read in the paper.

However, the proposed method has a big drawback also, that we will discuss in the next section.

2. The cross-section method of [NOM01]

We refer to the paper of Nomura for a detailed description of their methodology. We take the same notation as them in this section, except that we write factors as F instead of M.

The paper of Nomura states:

We want to find the optimal x [1]*that makes the proxy spreads* S^{proxy} *as close as possible to the market spreads* S^{mkt} *. Here we define "as close as possible" to mean "minimising total squared difference in log spreads", so finding the optimal* x *simply consists of performing a linear regression.*

And indeed, if we minimize log-spreads versus the log of their corresponding factor approximation, we get:

$$\min_F \sum_i \left(\log S_i - \log F_{glob} \cdot F_{sector_i} \cdot F_{region_i} \cdot F_{rating_i} \cdot F_{seniority_i}\right)^2$$

Which is equivalent to:

$$\min_F \sum_i \left(\log S_i - \log F_{glob} + \log F_{sector_i} + \log F_{region_i} + \log F_{rating_i} + \log F_{seniority_i}\right)^2$$

If we use an indicator vector **a** of zeros and ones as they define in the paper, we indeed get:

$$\min_{\boldsymbol{x}} \sum_i \left(\log S_i - \boldsymbol{a}^{\boldsymbol{T}} \cdot \boldsymbol{x}\right)^2$$

For which solving in least-squares is equivalent to solving the overdetermined system $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$ where **A** is the matrix of vectors $\mathbf{a}^T$ and **b** is the vector of log S_i.

So the authors have, in effect, linearized the problem, and this makes the method, of course, very attractive because any numerical linear algebra software solves this problem very easily.

We show in the next section, however, that this linearization by taking logarithms has a possibly unwanted effect: the method tries to come as close as possible to the geometric average of the spread-distribution of a particular category, and not the arithmetic average.

2.1 Optimizing log-spread differences

In order to prove that the method will tend to the geometric average of spreads of a category instead of the arithmetic average, we first show that a method with one degree of freedom per bucket (in effect, an intersection method) will tend towards these geometric means. Once this is shown, it is clear that the cross-section method, which simply optimizes the same objective, but with a parametrized representation with less degrees of freedom, will also tend to geometric averages of each bucket's spreads.

Again, we take a similar notation as above, but with one degree of freedom exactly per bucket:

1 Where x is the vector of log-factors: log F_i

$$\min_{\hat{S}} \sum_i (\log S_i - \log \hat{S}_{sector_i, region_i, rating_i, seniority_i})^2$$

Where the $\hat{S}$ represents this single spread per bucket.

We can now find the minimum for each $\hat{S}$ analytically, and even independently. Indeed, if we take the gradient of the above expression with respect to a specific $\hat{S}$, only the spreads of its particular bucket will have a non-zero gradient, since this particular $\hat{S}$ only occurs there. So we can rewrite the above optimization problem as multiple optimization problems:

$$\min_{\hat{S}_{bucket}} \sum_{S_i \in bucket} (\log S_i - \log \hat{S}_{bucket})^2 \, \forall \, bucket$$

Solving those problems is easy, we do it generically here:

$$\frac{\partial \sum_i (\log S_i - \log \hat{S})^2}{\partial \hat{S}} = \frac{\partial \sum_i (\log(\hat{S}/S_i))^2}{\partial \hat{S}} = \sum_i 2 \log(\hat{S}/S_i) \cdot \frac{1}{\hat{S}/S_i} \cdot \frac{1}{S_i} = \ldots$$

$$\sum_i 2 \log(\hat{S}/S_i)/\hat{S}$$

We will solve this last expression as equal to zero, in order to find all local optimums, which in this case is a global minimum, since the problem is clearly convex:

$$\sum_i 2 \log(\hat{S}/S_i)/\hat{S} = \sum_i 2 \frac{\log \hat{S} - \log S_i}{\hat{S}} = 0$$

Since it is clear that in zero, $\frac{\log \hat{S}}{\hat{S}}$ tends to minus infinity, we can solve this with the condition $\hat{S} > 0$.

We can thus write:

$$\sum_i^N \log \hat{S} = \sum_i^N \log S_i \Leftrightarrow N \cdot \log \hat{S} = \sum_i^N \log S_i$$

Finally:

$$\hat{S} = \exp\left(\frac{\sum_i \log S_i}{N}\right) = \sqrt[N]{\prod_i S_i}$$

Where we see that the right-hand side is simply the geometric average of all spreads in the bucket.

So as mentioned above, since the cross-section method as described in [NOM01] is simply a further parametrization of the problem described above, with less degrees of freedom, the method will also tend towards geometric averages of the spreads of each bucket, although not perfectly, since it must optimize with limited parameters on a great number of buckets.

2.2 Examples for a lognormal distribution of spreads

Although we will not discuss here the question of how credit spreads are distributed per bucket in reality, we will show the impact of using log-spread differences in the optimization if the distribution of observed spreads were lognormal.

Since the lognormal distribution is defined as the exponential of a normal distribution, the geometric average has a very simple form. We show this here:

$\sqrt[N]{\prod_i S_i} = \exp\left(\frac{\sum_i \log S_i}{N}\right)$ where $\log S_i$ is normally distributed.

So if $S \sim LogN(\mu, \sigma^2)$, then the geometric average is simply e^{μ} .

The arithmetic average, however, can be shown to be: $e^{\mu+\sigma^2}$

And so we have: $AM = GM \cdot e^{\sigma^2}$

This means that the greater the variance of the observed spreads, the greater the underestimation of the arithmetic average.

As an example, if we have a category of spreads with mean 50 and standard deviation 20, we have:

$$\begin{cases} 50 = e^{\mu+\sigma^2} \\ 20^2 = (e^{\sigma^2} - 1) \cdot e^{2\mu+\sigma^2} \end{cases}$$

Which, if we solve it, yields:

$$\begin{cases} \mu = \log \frac{50^2 - 20^2}{50} = 3.74 \\ \sigma^2 = \log 50 - \mu = 0.17 \end{cases}$$

And since $e^{0.17} = 1.19$, we see that the geometric average is somewhat 16% lower than the arithmetic average.

We observe, in practice, however, that spread distributions are not at all as concentrated as the example given above. Standard deviations can be much higher than 40% of the average (as in the example above) for low-rating buckets.

So this effect can cause quite extreme underestimations when the standard deviation of the observed spreads is large. We show the relative underestimation in the following chart for increasing standard deviations (as percentage of the average) for the lognormally distributed spreads:

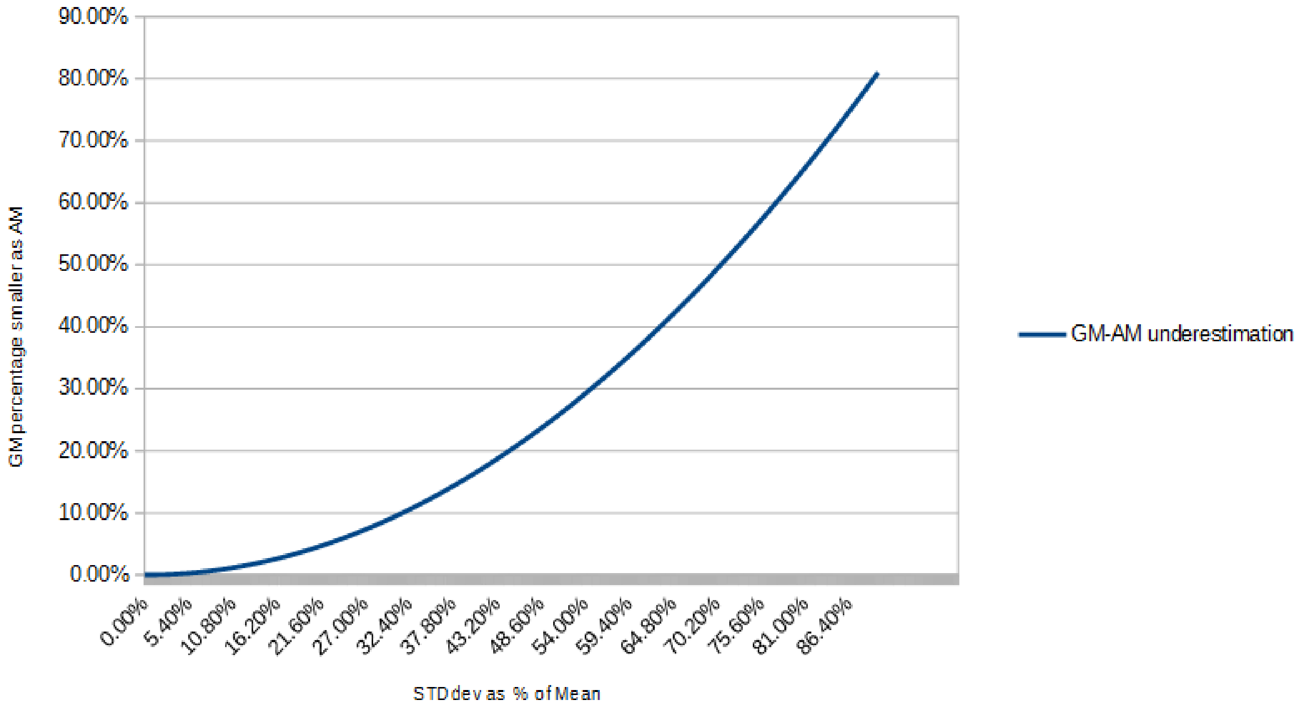

Illustration 1: GM-AM underestimation in function of STD

3. An alternative optimization technique for the cross-section method

The best way to avoid the effect shown above, is to optimize on a different objective, that does not use the logarithm to linearize the problem. Of course, this means the problem might get harder to solve numerically, but as we will show, the gradients of the new problem can be computed, in order to optimize numerically quite efficiently.

3.1 New objective formulation

We will now optimize on a simple least-squares objective in spread levels instead of log-spread levels. But first, we introduce our notation that is slightly more specific than in [NOM01], because we will need it later to compute the gradients.

Each observed spread is first categorized in its bucket:

$S_i \rightarrow (sector_i, region_i, rating_i, seniority_i)$, which we will write: $S_i \rightarrow (s_i, c_i, r_i, sn_i)$

Each factor is also assigned a single index, which ranges from one to the cumulative amount of categories in each characteristic:

$F_j \rightarrow j \in (1, \ldots, N_s + N_c + N_r + N_{sn})$ where N_s is the amount of categories in the sector characteristic, and so on.

We can then define the proxy spreads:

$\hat{S}_i = F_{s_i} \cdot F_{c_i} \cdot F_{r_i} \cdot F_{sn_i}$ where $s_i \in (1, \ldots, N_s), c_i \in (N_s + 1, \ldots, N_s + N_c)$, and so on.

Now, we can formulate the new objective in spread levels:

$$\min_{F_j} \sum_i^N (S_i - \hat{S}_i)^2$$

In the next sections, we show that this problem can be solved efficiently numerically, by computing the gradients, but first we show that this method will tend to the arithmetic averages per bucket, in a similar way as done above for log-spreads and the geometric average.

3.2 Convergence of new objective towards arithmetic average

We show this trivially, per bucket like above:

By the solution of a linear least-squares problem: $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$, we have:

$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}$, where in our case:

$\boldsymbol{X} = \boldsymbol{1}_{N \times 1}$ and $\boldsymbol{y} = \boldsymbol{S}_{N \times 1}$ (the spreads)

So we have:

$$\hat{S}=(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}=(\boldsymbol{1}^{\boldsymbol{T}}\boldsymbol{1})^{-1}\boldsymbol{1}^{\boldsymbol{T}}\boldsymbol{S}=N^{-1}\sum_i S_i$$

This is the arithmetic average of the spreads of the particular bucket.

3.3 Solution of the new optimization problem

In order to solve the new optimization problem numerically, we compute the gradient of the objective function:

$$\frac{\partial\sum_i^N (S_i-F_{s_i}\cdot F_{c_i}\cdot F_{r_i}\cdot F_{sn_i})^2}{\partial F_j}$$

This gradient can be seen as the sum of gradients by the linearity property of differentiation:

$$\frac{\sum_i^N \partial(S_i-F_{s_i}\cdot F_{c_i}\cdot F_{r_i}\cdot F_{sn_i})^2}{\partial F_j}$$

Which is zero if $j\notin\{s_i,c_i,r_i,sn_i\}$, and otherwise equal to:
$-2[(S_i F_{s_i}\cdot F_{c_i}\cdot F_{r_i}\cdot F_{sn_i}-F_{s_i}^2\cdot F_{c_i}^2\cdot F_{r_i}^2\cdot F_{sn_i}^2)]/F_j$

So that we can write the entire gradient as:

$$\frac{\partial\sum_i^N (S_i-F_{s_i}\cdot F_{c_i}\cdot F_{r_i}\cdot F_{sn_i})^2}{\partial F_j}=-2\sum_i^N \boldsymbol{1}_{j\in\{s_i,c_i,r_i,sn_i\}}[S_i F_{s_i}\cdot F_{c_i}\cdot F_{r_i}\cdot F_{sn_i}-F_{s_i}^2\cdot F_{c_i}^2\cdot F_{r_i}^2\cdot F_{sn_i}^2]/F_j$$

This expression can be computed numerically, and so it is possible to implement a steepest descent method for this problem, and to solve the cross-section method for this new objective.

We note, that the hessian of the objective can also be computed easily:

$$\begin{cases}\dfrac{\partial^2(S_i-F_{s_i}\cdot F_{c_i}\cdot F_{r_i}\cdot F_{sn_i})^2}{\partial F_j\partial F_k}=-2[(S_i F_{s_i}\cdot F_{c_i}\cdot F_{r_i}\cdot F_{sn_i}-2F_{s_i}^2\cdot F_{c_i}^2\cdot F_{r_i}^2\cdot F_{sn_i}^2)]/(F_j F_k)\,\forall\, j\neq k\\[2ex] \dfrac{\partial^2(S_i-F_{s_i}\cdot F_{c_i}\cdot F_{r_i}\cdot F_{sn_i})^2}{\partial F_j\partial F_k}=[F_{s_i}^2\cdot F_{c_i}^2\cdot F_{r_i}^2\cdot F_{sn_i}^2]/(F_j F_k)\,\forall\, j=k\end{cases}$$

This means that we can use a Newton method also to find the solution numerically.

Conclusion

We have shown that the cross-section method as proposed in [NOM01] will tend to optimize towards the geometric averages of spreads in each bucket. This is the result of optimizing on log-spreads. This effect might be undesired in certain applications, so we propose another methodology that converges towards the arithmetic averages instead of the geometric ones. The proposed methodology is a classical least squares problem, but which is not linear. We thus show how to compute the gradient and hessian of this problem to be able to solve it numerically.

Bibliography

NOM01: Kyriakos Chourdakis, Eduardo Epperlein, Marc Jeannin, James Mcewen, A cross-section across CVA, 2013