

Choosing Markovian Credit Migration Matrices by Nonlinear Optimization

Maximilian Hughes, Ralf Werner

Angaben zur Veröffentlichung / Publication details:

Hughes, Maximilian, and Ralf Werner. 2016. "Choosing Markovian Credit Migration Matrices by Nonlinear Optimization." *Risks* 4 (3): 31. <https://doi.org/10.3390/risks4030031>.

Nutzungsbedingungen / Terms of use:

CC BY 4.0



Article

Choosing Markovian Credit Migration Matrices by Nonlinear Optimization

Maximilian Hughes ¹ and Ralf Werner ^{2,*}

¹ Fakultät für Informatik und Mathematik, Hochschule München, Lothstrasse 64, 80335 München, Germany; maximilian.hughes@hm.edu

² Institut für Mathematik, Universität Augsburg, Universitätsstrasse 2, 86159 Augsburg, Germany

* Correspondence: ralf.werner@math.uni-augsburg.de; Tel.: +49-821-598-5854

Academic Editor: Alexander Szimayer

Received: 4 July 2016; Accepted: 22 August 2016; Published: 30 August 2016

Abstract: Transition matrices, containing credit risk information in the form of ratings based on discrete observations, are published annually by rating agencies. A substantial issue arises, as for higher rating classes practically no defaults are observed yielding default probabilities of zero. This does not always reflect reality. To circumvent this shortcoming, estimation techniques in continuous-time can be applied. However, raw default data may not be available at all or not in the desired granularity, leaving the practitioner to rely on given one-year transition matrices. Then, it becomes necessary to transform the one-year transition matrix to a generator matrix. This is known as the embedding problem and can be formulated as a nonlinear optimization problem, minimizing the distance between the exponential of a potential generator matrix and the annual transition matrix. So far, in credit risk-related literature, solving this problem directly has been avoided, but approximations have been preferred instead. In this paper, we show that this problem can be solved numerically with sufficient accuracy, thus rendering approximations unnecessary. Our direct approach via nonlinear optimization allows one to consider further credit risk-relevant constraints. We demonstrate that it is thus possible to choose a proper generator matrix with additional structural properties.

Keywords: credit risk; embedding problem; transition matrix; generator matrix; homogenization; best approximation of the annual transition matrix; quasi-optimization of the generator; constraints; optimization; regularization

1. Introduction

Numerous methods exist for embedding a given discrete-time (in most cases, one-year) transition matrix into continuous-time, i.e., for finding the (infinitesimal) generator matrix. Simply computing the matrix logarithm does not necessarily result in a valid solution for the generator matrix. This seems especially to be the case for transition matrices in credit risk, where often, the matrix logarithm yields candidates for a generator, but with negative off-diagonal entries. Consequently, one needs to find procedures where, on the one hand, valid generators are produced and, on the other hand, the distance (for the moment using any matrix norm) of its exponential to the given transition matrix is minimal.

Our research mainly follows the work of [1], where the proposed method is also based on optimization. They suggest to avoid solving the true problem of interest, the best approximation of the annual transition matrix (BAM), but instead solving an approximation only, called quasi-optimization of the generator (QOG). For the solution of QOG, they propose the DMPG algorithm (somewhat confusingly called the distance minimization problem for the generator matrix). The DPMG

algorithm solves QOG by viewing each row of the transition matrix separately¹. Based on a surprising close connection between BAM and QOG, proven by [2], we suggest to solve BAM directly by methods of nonlinear programming based on a starting point obtained by the solution of QOG. To deal with the non-convex structure of BAM, we apply an iterative homogenization procedure to obtain more reliable solutions. This procedure yields quite robust solutions, as accuracy is gained step by step. Although our numerical experiments show no superiority compared to a direct solution of BAM for the considered problem instances, it gives the latter more credibility of being the best solution. Further, we provide a comparison to the diagonal (DA) and weighted adjustment (WA) techniques, which are quite popular techniques in current credit risk applications. In general, the optimization technique allows a more flexible interpretation and various additional applications, as opposed to non-optimization methods (e.g., DA or WA). For instance, constraints enforced by regulators, for example one-year default probabilities, need to be greater or equal to a given threshold or bank internal requirements may be incorporated. More precisely, one further main novelty of our approach lies with the adjustment of default and migration probabilities/rates during the optimization procedure to a specific desired outcome. These extensions are thoroughly investigated in our study.

In this context, it needs to be mentioned that several research papers also exist solely addressing computing the roots of transition matrices, e.g., [3–5]. Here, Lin, L. [6] proceeds in her dissertation similarly as in our work by solving the minimization problem as a whole for obtaining an optimal solution of BAM by utilizing MATLAB's² inbuilt optimization routines. There, analytic formulas for the gradient and the Hessian matrix of the objective function are derived. Further attention is paid to the choice of the nonlinear optimization method and the choice of the starting point. However, the focus there is on local methods only, with almost no discussion of the global aspects of the problem. More importantly, Lin does not exploit this formulation by adding further highly-relevant credit risk-related constraints.

The remainder of this paper is organized as follows: First, we briefly review existing results on whether the logarithm of a given transition matrix exists and whether the one-year matrix can be embedded into continuous time, which is given in Section 2. These checks are conducted on commonly-used one-year transition matrices in credit risk-relevant assessments, which are published on a regular basis by rating agencies. Section 3 introduces sufficient conditions and additional constraints, as well as the underlying optimization problems. An extensive application in Section 4 compares the underlying techniques and addresses the impact of the additional constraints imposed specifically on present and future default probabilities. Again, transition matrices estimated from real data are utilized for an extensive analysis. In Section 5, a brief summary is given.

2. Transition Matrix Analysis

The common approach, in the context of embedding a transition matrix³ in continuous time, is based on the notion of the generator of a continuous-time Markov process. A generator (or intensity matrix) of a Markov process $X(t)$ is a matrix G , for which:

$$\begin{aligned} \sum_{j=1}^K g_{ij} &= 0, \quad i = 1, \dots, K; \\ g_{ij} &\geq 0, \quad \forall i \neq j \text{ and } i, j = 1, \dots, K, \end{aligned} \quad (1)$$

¹ An application of the row-by-row approach to the progression of a chronic disease through various severity states can be found in [7] and is also compared in [8].

² Version 8.4.0 (R2014b), The MathWorks Inc., Natick, MA, USA.

³ We distinguish between general transition matrices in theory, denoted by P , and given transition matrices by rating agencies, denoted by $P(0, t)$, if not stated otherwise.

and for which the entries are rates at which the Markov process jumps from state i to j , with:

$$g_{ij} = \lim_{h \rightarrow 0} \frac{P(\tilde{X}(t+h) = j | \tilde{X}(t) = i)}{h}, \quad \forall i \neq j \text{ and } i, j = 1, \dots, K.$$

These necessary and sufficient properties of (1) for being a generator matrix can be derived in three different ways, as can be seen in [9]. A formal proof of (1) is given in [10]. In Reference [11], Bielecki, T. and Rutkowski, M. provide this compact definition of the embedding problem:

Definition 1 (Embedding problem). *Find a $K \times K$ matrix G satisfying (1) (non-negative off-diagonal entries and all rows summing to zero), such that $\exp(G) = P$.*

As $P(0,1) = \exp(G)$ from Definition 1 needs to hold, a feasible candidate may be the matrix logarithm of a given transition matrix. The work in [12] gives two necessary and sufficient conditions on the existence and uniqueness of the matrix logarithm. The existence is given for $P_{S\&P}(0,1)$, $P_{Moodys}(0,1)$ and $P_{Moodys\ XL}(0,1)$, resulting in a real-valued solution, but not in a valid generator according to (1), as negative off-diagonal values arise. Moreover, in general, the issue of multiple solutions may occur. For example, when computing the eigenvalues of $P_{Moodys\ XL}(0,1)$, a complex conjugate eigenvalue pair exists, so that one needs to search in every existing branch of the matrix logarithm for a valid generator. Applying the search algorithm proposed by [13,14], three real-valued solutions of the matrix logarithm exist, yet no valid generator can be found for this particular matrix due to the corresponding smallest off-diagonal entries -13.390 , -0.0013 and -19.061 of branches -1 , 0 and 1 . At this point, many publications addressing the matter resort to the adjustment techniques (DA and WA) for eliminating negative off-diagonal entries. Both methods are thoroughly described, for example, in [14].

The embedding problem itself relies on more restrictive conditions for obtaining valid generators which consists of the existence of an embeddable matrix in continuous time and its uniqueness (also referred to as identification), reaching back as far as [15]. Thus far, necessary and sufficient conditions have only been formulated for 2×2 (see [16], attributed to D. G. Kendall, and [17]) and 3×3 matrices (see [18–22]). For higher dimensions, the issue still remains vague. The work in [6] states a rather complete list of nine necessary conditions in her dissertation, where we only list the most important conditions according to [14]:

Condition 1 (Necessary conditions for the existence of a generator G with $\exp(G) = P$).

- (a) For every pair of states i and j , such that j is accessible from i , $p_{ij} > 0$. (cf. [23,24])
- (b) $\det(P) > 0$ (cf. [16])
- (c) $\det(P) \leq \prod_{i=1}^K p_{ii}$ (cf. [25])
- (d) The entries of P must satisfy:

$$p_{ik} \geq m^m r^r (m+r)^{-m-r} \sum_j (p_{ij} - b_m)(p_{jk} - b_r) \mathbf{1}_{p_{ij} > b_m, p_{jk} > b_r},$$

for any positive integers m and r . Here, $b_m = \sum_{l=m+1}^{\infty} e^{-\mu} \mu^l / l! = 1 - \sum_{l=0}^m e^{-\mu} \mu^l / l!$, which equals the probability that $N' > m$, where N' is a Poisson random variable with mean $\mu \equiv \max_i(-g_{ii})$, with $\max_i(-g_{ii}) \leq -\text{trace}(G) = -\log(\det(P))$. Furthermore, $\mathbf{1}_B$ is the indicator function of the Boolean event B . (cf. [14])

While Conditions (b) and (c) of Condition 1 are satisfied by all given matrices in Table A1, Condition (a) is not. For example, in $P_{S\&P}(0,1)$, it is possible to migrate from Class AAA to CCC-C via Class A ($p_{13}^{S\&P}(0,1) > 0$ and $p_{37}^{S\&P}(0,1) > 0$), but not directly, as $p_{17}^{S\&P}(0,1) = 0$. This violation is a frequently-observed phenomenon when examining credit risk transition matrices (due to

non-observable defaults for high ratings in historical data, as also pointed out in [14]). Therefore, [14] established a more quantitative version of (a), being the most recent condition added to the list, (d). However, the newly-computed lower bounds for the zero entries are not sufficiently large in order to obtain a valid generator, as still negative off-diagonals remain. This finding applies to all three example transition matrices.

It can be concluded that specifically credit risk transition matrices have problems fulfilling conditions on matrix entries, while eigenvalue conditions are usually satisfied. Thus far, we can conclude that typically, no valid generator for transition matrices exist.

Considerations concerning the uniqueness of the generator are, however, omitted at this stage, and the interested reader can resort to [13,14] where the five known sufficient conditions for the uniqueness of a generator G are thoroughly investigated and where examples are given.

3. Optimization Problems

To this end, numerically solving the embedding problem has been investigated in several research papers. A brief chronological listing of regularization-based methods is outlined, before formalizing the optimization problems proposed by [1]. First, regularization techniques of adjusting negative off-diagonal entries, arising when computing the matrix logarithm, can be found in [3,4], where DA and WA are described. A similar procedure is given in the seminal work of [26]. These are further investigated for example in [1,8,14]. The aim of [1] was to solve the optimization problem of the best approximation of the annual transition matrix. BAM is the actual problem to solve; however, it represents a high-dimensional non-convex optimization problem. Thus, [1] replaced the problem by an approximation, now minimizing the sum of squared deviations between $\log(P)$ and G , referred to as quasi-optimization of the generator (QOG). The approximation allowed [1] to solve the problem almost analytically, considering the rows of the input matrix separately (they refer to this algorithm as the distance minimization problem for the generator matrix (DMPG)). Until 2010, it remained open if there were a (weak/strong) relation between BAM and its approximation QOG. Then, [2] derived Condition 2, which gives the distance of any solution of a generator matrix G to the matrix logarithm $M = \log(P)$. In the context of optimization, it follows that should G be close to $\log(P(0, 1))$, then so is $\exp(G)$ to $P(0, 1)$.

Condition 2. Let P be a transition matrix such that all eigenvalues are strictly positive, and put $M = \log(P)$. If G lies within the set of Equation (1) and $\|M - G\|_\infty = \epsilon$, then:

$$\|P - \exp(G)\|_\infty \leq \min\{2, \exp(\epsilon) - 1\} \leq \min\{2, 2\epsilon\}, \quad (2)$$

where the used norm is $\|A\|_\infty = \max\{\|Av\|_\infty : \|v\|_\infty \leq 1\} = \max_{1 \leq i \leq n} \{\sum_{j=1}^n |A_{i,j}|\}$.

The first attempt of solving QOG and BAM (the actual problem) as a whole, hence nonlinearly, was conducted by [6] in her dissertation. This has one great advantage over all other methods so far, namely additional credit risk-relevant constraints can be incorporated, which are introduced in the following.

3.1. Constraints

For the optimization problems, we need to define the set of all valid generators; hence, Definition 2 is derived from: (1).

Definition 2 (Generator matrix constraints).

$$\mathcal{G} := \left\{ \mathbf{X} \in \mathbb{R}^{K \times K} : \sum_{j=1}^K x_{ij} = 0, \quad i = 1, \dots, K; \right. \quad (\text{G1})$$

$$\left. x_{ij} \geq 0, \quad \forall i \neq j \text{ and } i, j = 1, \dots, K \right\} \quad (\text{G2})$$

As the main novel idea, we suggest to additionally impose constraints⁴ specifically addressing credit risk-relevant characteristics. The first constraint (D1) reflects Basel II regulation requirements demanding that one-year default probabilities need to be larger or equal to 3 basis points (bp). Constraint (D2) is a logical continuation and a realistic internal banking requirement stating that default probabilities should be monotonic increasing with decreasing rating classes. Thus, abnormal behaviors as observed in Moody's transition matrix in Table A1 where $p_{38}^{\text{Moody's}}(0, 1) \leq p_{28}^{\text{Moody's}}(0, 1)$ can be corrected. While both constraints are optional, (D1) can be seen as necessary from a regulator's point of view, whereas (D2) can be considered as completely optional.

Definition 3 (Default probability constraints).

$$\mathcal{D} := \left\{ \mathbf{P} \in \mathbb{R}^{K \times K} : P_{i,K} \geq 0.0003, \quad i = 1, \dots, K-1; \right. \quad (\text{D1})$$

$$\left. P_{i,K} \leq P_{i+1,K}, \quad i = 1, \dots, K-2 \right\} \quad (\text{D2})$$

Continuing the idea of including additional constraints, we define a constraint specifically for the migration rates from one state to another. Similarly to (D2), we impose a monotonic behavior. For example, it is more probable to migrate from, for example, AAA to AA than from AAA to C, and vice versa. (M1) handles the upper triangular matrix and (M2) the lower triangular matrix, both excluding the main diagonal. Also not included is the default state 'D', as it is absorbing; thus, the last column and the last row are omitted from (M1) and (M2), respectively. Taking the rates into consideration has the advantage of being linear constraints, and "jumping too far" is less likely. The migration rate constraints of Definition 4 are again purely optional.

Definition 4 (Migration rate constraints).

$$\mathcal{M} := \left\{ \mathbf{X} \in \mathbb{R}^{K \times K} : x_{i,j-1} \geq x_{i,j}, \quad i = 1, 2, \dots, K-3 \text{ and } j > i+1 \right. \quad (\text{M1})$$

$$\left. x_{i,j} \geq x_{i,j+1}, \quad i = 3, 4, \dots, K-1 \text{ and } j < i-1 \right\} \quad (\text{M2})$$

Lastly, in our selection of additional constraints, we consider a general monotonic behavior of the rating concept in credit risk. The first state is associated with AAA, being the best rating class and having the least risk. With decreasing ratings, the risk increases. Following [26] (also found in [14]; the corresponding proof is given in [27]), this requirement can be ensured by one of the following two conditions of Condition 3.

Condition 3. Let \mathbf{G} be a valid generator for \mathbf{P} . Then, the following statements are equivalent:

1. $\sum_{j \geq k} p_{ij}$ is a nondecreasing function of i for every fixed k and
2. $\sum_{j \geq k} g_{ij} \leq \sum_{j \geq k} g_{i+1,j}$ for all i, k such that $k \neq i+1$.

⁴ Each single additional constraint can be switched on or off according to the user's preferences. Further extensions are conceivable.

An example where the first condition does not apply, we find that with $k = 5$ and $i = 6$, the condition is violated for $P_{\text{Moody's}}(0, 1)$, as $\sum_{j \geq 5} p_{6j}^{\text{Moody's}}(0, 1) = 0.9912$ and $\sum_{j \geq 5} p_{7j}^{\text{Moody's}}(0, 1) = 0.9829$. In [14], Israel, R. B. et al. point out that this violation is frequently observed with given transition matrices. With both conditions being equivalent, the less cumbersome choice for the upcoming optimizations is the second condition because of its linear nature.

Definition 5 (Rate constraints).

$$\mathcal{R} := \left\{ \mathbf{X} \in \mathbb{R}^{K \times K} : \sum_{j \geq k} x_{ij} \leq \sum_{j \geq k} x_{i+1,j} \text{ for all } i, k \text{ such that } k \neq i + 1. \right\} \quad (\text{R1})$$

Remark 1. Definition 5 can be regarded as a combination of Definition 4 and (D2), with the difference that (D2) is applied to each column and where the rates are considered instead of probabilities.

3.2. Best Approximation of the Annual Transition Matrix

Problem 1 (Best approximation of the annual transition matrix (BAM)). Find a generator matrix G that, when exponentiated (by the matrix exponential), most closely matches a given annual transition matrix $P(0, 1)$, with:

$$\begin{aligned} G &= \arg \min_{\mathbf{X} \in \mathcal{G}} \quad \|\exp(\mathbf{X}) - P(0, 1)\|_{\text{F}}^2, & (\text{BAM}) \\ \text{subject to} \quad & \exp(\mathbf{X}) \in \mathcal{D} \\ & \mathbf{X} \in \mathcal{M} \\ & \mathbf{X} \in \mathcal{R} \end{aligned}$$

where $\|\cdot\|_{\text{F}}$ denotes the Frobenius norm.

The above Problem 1 is a high-dimensional and nonlinear optimization problem, thus from a computational perspective, not an easy undertaking. Therefore, [1] suggest modifying the nonlinear problem to QOG.

3.3. Quasi-Optimization of the Generator

Problem 2 (Quasi-optimization of the generator (QOG)). Find a generator matrix G that most closely matches the unique matrix logarithm of the annual transition matrix $P(0, 1)$, with:

$$\begin{aligned} G &= \arg \min_{\mathbf{X} \in \mathcal{G}} \quad \|\mathbf{X} - \log(P(0, 1))\|_{\text{F}}^2, & (\text{QOG}) \\ \text{subject to} \quad & \exp(\mathbf{X}) \in \mathcal{D} \\ & \mathbf{X} \in \mathcal{M} \\ & \mathbf{X} \in \mathcal{R} \end{aligned}$$

where $\|\cdot\|_{\text{F}}$ denotes the Frobenius norm.

Problem 2 is an approximation to Problem 1, which is a convex quadratic problem with linear constraints, if $\exp(\mathbf{X}) \in \mathcal{D}$ is omitted. Further, it can be simplified to a row-by-row linear problem if $\mathbf{X} \in \mathcal{M}$ and $\mathbf{X} \in \mathcal{R}$ are also left out, leading to a simple algorithm, referred to as the distance minimization problem for the generator matrix (DMPG); see [1] for details.

4. Optimization Application

Here, the above problems in Section 3 are investigated for the given one-year transition matrices provided by Moody in Table A1, specifically $P_{\text{Moody}}(0,1)$, $P_{\text{Moody}}^{\text{XXL}}(0,1)$ and $P_{\text{Moody}}^{\text{XXL}}(0,5)$. $P_{\text{S\&P}}(0,1)$ is omitted at this stage, as it has a similar structure and the same properties (see Section 2) as $P_{\text{Moody}}(0,1)$, thus yielding similar results. Additionally, for reasons of reproducibility and validation of the results (any kind of credit risk transition matrix getting thrown at the above optimization techniques should yield stable and valid solutions), a five-year forecasted transition matrix⁵ is included, denoted by $P_{\text{Moody}}^{\text{XXL}}(0,5)$. For the computations, MATLAB's inbuilt nonlinear optimization routine `fmincon()` with the 'sqp' algorithm is utilized. The default optimization options defined by MATLAB are not changed, except for raising the maximum number of function evaluations and iterations, allowing the algorithm more computation time for the complex underlying task (64 variables need to be calibrated for the 8×8 and 441 variables for the 21×21 matrices). All computations were executed on a MacBook Pro with a 2.4-GHz Intel Core i5 processor and 8 GB 1333 MHz DDR3 memory.

4.1. Regularization Comparison

In this analysis, we solve the problems QOG and BAM w.r.t. $\|\cdot\|_F$ and compare with DA, WA and DMPG (see [1] for details), based on the (G1) and (G2) constraints only. An example of a valid generator matrix is given in Table A2, which is in this case obtained with BAM. Throughout Table 1, the optimization approaches (including DMPG) show overall better results than DA and WA w.r.t. to the averaged Frobenius norm ($\frac{1}{K^2} \|\exp(G) - P(0,1)\|_F$). This is in line with the results from [8] where a more elaborate comparison of the regularization methods can be found. The related methods DMPG and QOG obviously reveal the same outcome. The best results are achieved by BAM. It is also noteworthy that the differences between methods increase the larger the input matrix is. This is especially the case for $P_{\text{Moody}}^{\text{XXL}}(0,5)$.

Table 1. Comparison of the regularization methods based on averaged Frobenius norm (QOG and BAM optimizations based on Constraints (G1) and (G2)), for given transition matrices from Moodys in Table A1.

Constraints	Method	Start Matrix	$P_{\text{Moody}}(0,1)$	$P_{\text{Moody}}^{\text{XXL}}(0,1)$	$P_{\text{Moody}}^{\text{XXL}}(0,5)$
(G1), (G2)	DA		8.86×10^{-6}	7.72×10^{-6}	2.28×10^{-5}
	WA		6.74×10^{-6}	5.38×10^{-6}	1.75×10^{-5}
	DMPG		6.33×10^{-6}	4.82×10^{-6}	1.26×10^{-5}
	QOG		6.33×10^{-6}	4.82×10^{-6}	1.26×10^{-5}
		X_0^{Log}	(0.21 s)	(7.18 s)	(10.70 s)
		X_0^{WA}	(0.38 s)	(9.91 s)	(21.06 s)
		X_0^{DMPG}	(0.14 s)	(4.02 s)	(4.60 s)
	BAM		6.28×10^{-6}	4.67×10^{-6}	5.62×10^{-6}
		X_0^{Log}	(0.16 s)	(1.92 s)	(17.33 s)
		X_0^{WA}	(0.17 s)	(2.22 s)	(27.07 s)
		X_0^{DMPG}	(0.16 s)	(3.41 s)	(18.26 s)
		X_0^{QOG}	(0.20 s)	(4.24 s)	(23.72 s)

⁵ As can be seen on the right picture of Table A1, the majority of probability mass for lower ratings is shifted from the main diagonal to the default column, revealing a rather atypical transition matrix.

Additionally, different start matrices are inserted for **QOG** and **BAM**. One obvious choice is the principal logarithm⁶ ($X_0^{\text{Log}} = \log(P(0, t))$), which is computed by MATLAB's `logm()` function. An improved starting point would be to eliminate the negative off-diagonal entries a priori. Therefore, at no noteworthy additional computational cost, we compute WA (this is the preferred choice as the starting matrix of [6]) and DMPG. For **BAM**, additionally, the solution of **QOG** (X_0^{QOG}) is considered. Remarkable is the fact that the start matrix does not have an effect on the overall outcome. Eventually, regardless of the inserted start matrix, the optimization routine will find the same optimal solution. The only influence can be observed in the computation times. However, this seems to only apply to **QOG**, whereas **BAM** is relatively indifferent (with X_0^{WA} for $P_{\text{Moodys XXL}}(0, 5)$ being an exception) to the choice. For these reasons and assuring comparable and consistent results, all of the following outcomes are based on X_0^{DMPG} .

The results from Table 1 are further supported by taking a closer look at the default probabilities. Here, the absolute differences are compared with each other⁷. Therefore, in addition, we compute the future default probabilities of ten years. Visually, it can be inferred that a certain hierarchical order of the applied approaches can be observed in Figure 1, with **QOG** being superior to WA and **BAM** again being superior to **QOG** (and WA).

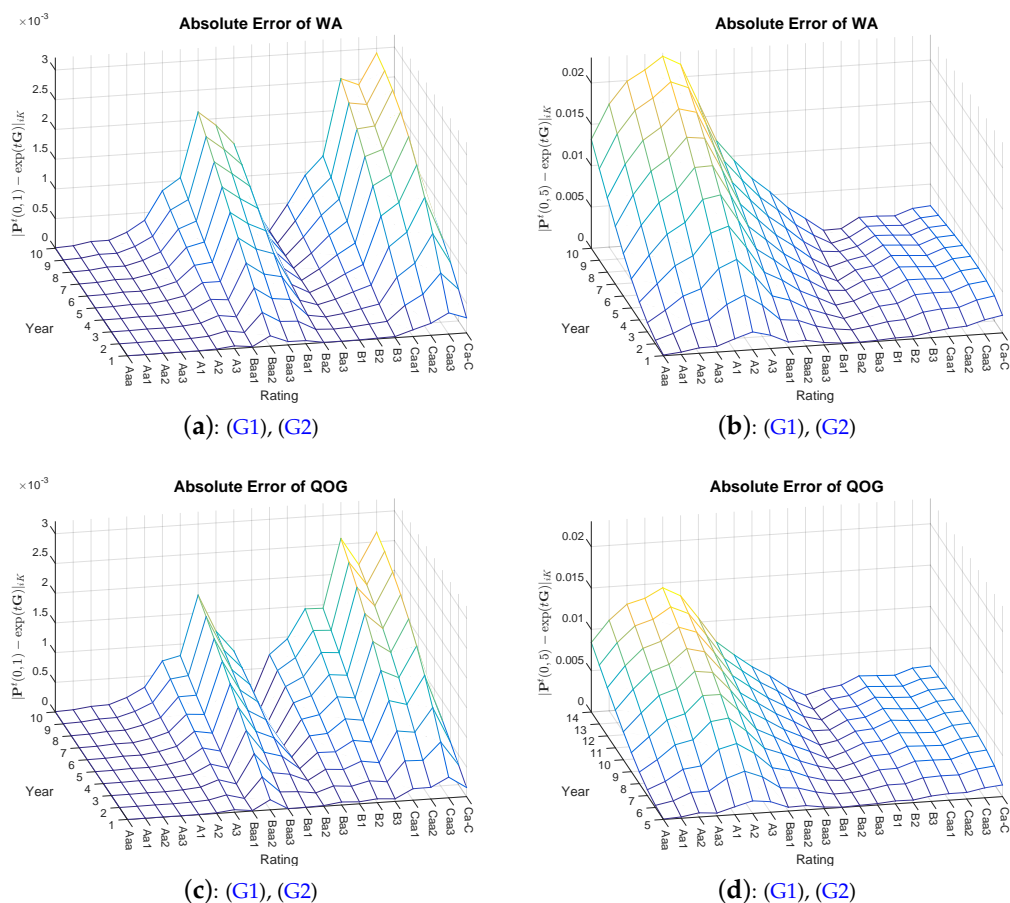


Figure 1. Cont.

⁶ The principal logarithm of $P_{\text{Moodys XXL}}(0, 1)$ has the nearest negative off-diagonal entry to zero. This also applies to $P_{\text{Moodys XXL}}(0, 5)$.

⁷ DA and DMPG are deliberately omitted at this point due to the results in Table 1. Furthermore, the visual depiction of $P_{\text{Moodys}}(0, 1)$ is spared due to too small differences.

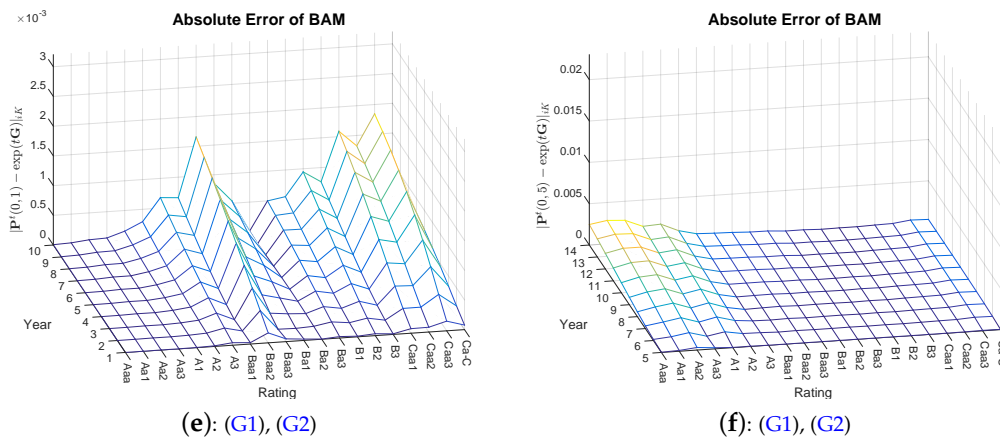


Figure 1. Absolute errors of the methods WA, QOG and BAM (QOG and BAM optimizations based on Constraints (G1) and (G2)) with starting point X_0^{DMPG} . Left side: (a, c, e) transition matrix $P_{\text{Moodys XXL}}(0, 1)$; right side: (b, d, f) transition matrix $P_{\text{Moodys XXL}}(0, 5)$.

4.2. Towards Global Optimization

Global optimization approaches are introduced for potentially further improving the results from Section 4.1 benchmarked by BAM. Therefore, the following methods are applied:

- A homogenization approach is solved iteratively, where $\alpha \in [0, 1]$ and the interval is discretized, for example, into steps of 0.01, thus moving from Problem 2 to Problem 1, where in each step, the start matrix is replaced by the optimal solution of the previous step. This way, a more robust procedure is obtained as the simpler (as convex problem) QOG is utilized for generating better start matrices, accordingly gaining accuracy in each iteration. More precisely, we find a generator matrix G that, when exponentiated (by the matrix exponential), most closely matches the annual transition matrix $P(0, 1)$, with:

$$G = \arg \min_{X \in \mathcal{G}} \alpha \| \exp(X) - P(0, 1) \|_F^2 + (1 - \alpha) \| X - \log(P(0, 1)) \|_F^2 \quad (3)$$

where $\| \cdot \|_F$ denotes the Frobenius norm. Equivalently, as in Problems 2 and 1, the sets \mathcal{M} (Definition 4) and \mathcal{R} (Definition 5) can easily be included. Set \mathcal{D} (Definition 3), on the other hand, also needs to be homogenized, as well.

- Global solutions in combination with MATLAB's local solver fmincon() with multiple start points.
 - 'MultiStart' (MS)
 - 'GlobalSearch' (GS)

Table 2 shows the results of the global optimization in comparison to BAM. No global technique is thereby superior to BAM; thus, the local solver is sufficient for obtaining a valid generator. From a computational perspective, it is also advisable to use BAM, as the durations are significantly shorter. Concluding, this outcome can also be interpreted as the confirmation of BAM delivering overall reliable results, although being a high-dimensional and nonlinear problem.

Table 2. Comparison of global solutions based on the averaged Frobenius norm and Constraints (G1), (G2) with starting point X_0^{DMPG} , for given transition matrices from Moodys in Table A1 (* the MS method uses a set of 100 different start points, which are parallelized).

Constraints	Method	$P_{\text{Moodys}}(0,1)$	$P_{\text{Moodys XXL}}(0,1)$	$P_{\text{Moodys XXL}}(0,5)$
(G1), (G2)	Homogen	6.28×10^{-6} (24.95 s)	4.67×10^{-6} (718.74 s)	5.62×10^{-6} (1938.16 s)
	MS ^(*)	6.28×10^{-6} (32.67 s)	4.67×10^{-6} (1353.93 s)	5.62×10^{-6} (12509.95 s)
	GS	6.28×10^{-6} (2.77 s)	4.67×10^{-6} (18.01 s)	5.62×10^{-6} (9447.82 s)

4.3. Constraints Analysis

We now extend the already conducted comparison of the default probabilities from Section 4.1 to the additional constraints. First of all, the averaged Frobenius norm of the complete resulting generator matrix is compared w.r.t. the optimization approaches with their corresponding constraints (Section 3). By adding the additional Constraints ((D1), (D2), (M1), (M2) and (R1)), the overall fit deteriorates as one deliberately moves away from the ‘true’ given transition matrix. Larger discrepancies are therefore to be expected when incorporating (M1) and (M2), as well as (R1), potentially effecting the whole matrix. Clearly, all additional constraints also depend on the state of the given transition matrix itself. Monotonically sorting w.r.t. migration probabilities, default probabilities and ratings and default probabilities larger or equal to 3 bp will have the same optimization result when only using (G1) and (G2). Overall better results can again be obtained with BAM as opposed to QOG. Similarly to Section 4.1, the improvement is smaller for the one-year transition matrices, while the difference in the distance decrease becomes more significant for the five-year forecast.

By increasing the constraints imposed to the optimization, the computation durations also increase. When looking at the two approaches more closely, the results show that for $P_{\text{Moodys}}(0,1)$, BAM and QOG are approximately equally fast. For $P_{\text{Moodys XXL}}(0,1)$, BAM is consistently faster than QOG, while for $P_{\text{Moodys XXL}}(0,5)$, the opposite seems to be the case for some constraints. Albeit, the computations are overall in acceptable ranges from a modeling perspective, even for larger transition matrices. All results of the comparison based on the Frobenius norm and corresponding computation times can be viewed in Table 3.

Table 3. Comparison of the additional constraints of QOG and BAM based on the averaged Frobenius norm with starting point X_0^{DMPG} , for given transition matrices from Moodys in Table A1.

Constraints	Method	$P_{\text{Moodys}}(0,1)$	$P_{\text{Moodys XXL}}(0,1)$	$P_{\text{Moodys XXL}}(0,5)$
(G1), (G2), (D1)	QOG	9.51×10^{-6} (0.62 s)	5.30×10^{-6} (11.13 s)	1.27×10^{-5} (22.70 s)
	BAM	9.43×10^{-6} (0.31 s)	5.16×10^{-6} (3.31 s)	5.75×10^{-6} (26.25 s)
(G1), (G2), (D2)	QOG	6.74×10^{-6} (0.66 s)	5.11×10^{-6} (18.21 s)	1.26×10^{-5} (4.65 s)
	BAM	6.70×10^{-6} (0.31 s)	4.96×10^{-6} (6.32 s)	5.62×10^{-6} (26.36 s)
(G1), (G2), (D1), (D2)	QOG	9.49×10^{-6} (0.49 s)	5.53×10^{-6} (16.07 s)	1.27×10^{-5} (22.44 s)
	BAM	9.43×10^{-6} (0.31 s)	5.39×10^{-6} (3.93 s)	5.75×10^{-6} (27.44 s)

Table 3. Cont.

Constraints	Method	$P_{\text{Moody's}}(0,1)$	$P_{\text{Moody's XXL}}(0,1)$	$P_{\text{Moody's XXL}}(0,5)$
(G1), (G2), (M1), (M2)	QOG	1.46×10^{-5} (0.22 s)	4.55×10^{-5} (13.97 s)	4.25×10^{-5} (22.26 s)
	BAM	1.45×10^{-5} (0.28 s)	4.49×10^{-5} (5.46 s)	2.83×10^{-5} (114.72 s)
(G1), (G2), (R1)	QOG	9.98×10^{-5} (0.42 s)	2.12×10^{-5} (14.91 s)	1.40×10^{-4} (78.80 s)
	BAM	9.78×10^{-5} (0.33 s)	2.01×10^{-5} (4.30 s)	5.50×10^{-5} (32.88 s)
All	QOG	1.00×10^{-4} (0.42 s)	4.76×10^{-5} (15.95 s)	1.09×10^{-4} (38.06 s)
	BAM	9.89×10^{-5} (0.51 s)	4.70×10^{-5} (6.41 s)	5.83×10^{-5} (51.18 s)

Next, the impact of the additional constraints on the default probabilities is analyzed. Table 4 displays the post optimization default probabilities of the top four investment grade Classes Aaa, Aa, A and Baa of the transition matrix $P_{\text{Moody's}}(0,1)$. Evidently, (D1) corrects $p_{1K}^{\text{Moody's}}(0,1)$ and $p_{3K}^{\text{Moody's}}(0,1)$ to 3 bp, and $p_{2K}^{\text{Moody's}}(0,1)$ is affected by (D2), where the hierarchical order of ratings is restored. Subsequently, combining (D1) and (D2) will yield probabilities equal or larger to 3 bp. (M1) and (M2) do not show a major impact on the default probabilities, at least not on the higher ratings. However, a slight change can be registered compared to the (G1) and (G2) case. This may be attributed to the minor correction of $p_{1,4}^{\text{Moody's}}(0,1)$ being smaller than its right neighbor and, thus, also affecting the default probabilities to some extent. This will become more obvious when analyzing $P_{\text{Moody's XXL}}(0,1)$ below. (R1), in turn, has a similar effect as (D2) with $k = 8$ and $i = 2$. No major changes could be detected for Baa, all within a deviance of ± 1 bp to the original probabilities at $t = 1$. Overall, the probabilities of BAM compared to QOG do resemble the original values of $P_{\text{Moody's}}(0,1)$ more accurately.

A visual depiction⁸ is given specifically of $P_{\text{Moody's XXL}}(0,1)$, as one can easily trace which constraint is responsible for its corresponding impact on the default probabilities. First, Constraint (D1) is added where in (a) and (b) of Figure 2, the ratings Aaa to A3 are adjusted to 3 bp. Not surprisingly, the errors grow accordingly, while all other rating classes remain unaffected (compare Figure 1c,e). Continuing with (c) and (d) of Figure 2, larger errors seem to occur for the rating Class Baa2. Indeed when looking at the matrix entry $p_{9,21}^{\text{Moody's XXL}}(0,1) = 0$ and its previous neighbor $p_{8,21}^{\text{Moody's XXL}}(0,1) = 0.001$, it violates Constraint (D2). Thus, the post-optimization entry is then adjusted to $p_{9,21}^{\text{Moody's XXL}}(0,1) \geq p_{8,21}^{\text{Moody's XXL}}(0,1)$ causing higher errors. The effect of combining (D1) and (D2) can then be seen in (e) and (f), where the errors accumulate accordingly. A different picture than one may expect is presented in (g) and (h) of Figure 3, as (M1) and (M2) do not influence the default probabilities directly. Yet, we see a rise in discrepancy in rating Class Caa2.

For the following computations, we pretend that $\log(P_{\text{Moody's XXL}}(0,1))$ is a potential generator of the original given transition matrix for the sake of comparability of the pre- and post-optimization values. Indeed, $g_{18,19}^{\text{Moody's XXL}}(0,1) = 0.0440$, whereas $g_{18,20}^{\text{Moody's XXL}}(0,1) = 0.0838$, almost doubling the amount of its left neighbor, so that a larger correction is required for $g_{18,19}^{\text{Moody's XXL}}(0,1)$. This can

⁸ Again, $P_{\text{Moody's}}(0,1)$ is omitted, yielding too close results where detecting differences visually is not feasible. Furthermore, $P_{\text{Moody's XXL}}(0,5)$ is not included, as some of the probability mass has already migrated to the default state; thus, visually, there is not much insight to gain.

be seen in Table 5, where for both optimizations QOG and BAM, the values are adjusted, so that $g_{18,19}^{\text{Moody's XXL}}(0,1) \geq g_{18,20}^{\text{Moody's XXL}}(0,1)$ is enforced. However, one can see that QOG already fails to maintain the original default probability of 0.1735. In the long run, it is obvious that manipulated pre-default stages (here, Caa2 and Caa3) will then have an impact on the future default probabilities, which explains the offset in (g) and (h). Between Rows 17 and 18, several violations of (R1) can be detected. Exemplary for all, the largest discrepancy is picked with $i = 17$ and $k = 14$. Table 6 shows that the original $\sum_{j \geq 14} g_{18,j}^{\text{Moody's XXL}}(0,1) < \sum_{j \geq 14} g_{17,j}^{\text{Moody's XXL}}(0,1)$ (again, simply $\log(\mathbf{P}^{\text{Moody's XXL}}(0,1))$ is computed of the original transition matrix), but is adjusted by QOG and BAM as required. Simultaneously, BAM can maintain the cumulative sum $\sum_{j \geq 14} g_{18,j}^{\text{Moody's XXL}}(0,1)$ more closely to the original. The rectification of the rating Caa1 comes at a cost of losing accuracy in the default probability. However, it can be observed (similarly to the (M1) and the (M2) case) that BAM is still closest to the original PDs. A summary of all effects can then be found in (k) and (l). Noteworthy is the positive effect when combining (M1) and (M2) with (R1), resulting in lower absolute errors in Caa1 compared to (R1) by itself.

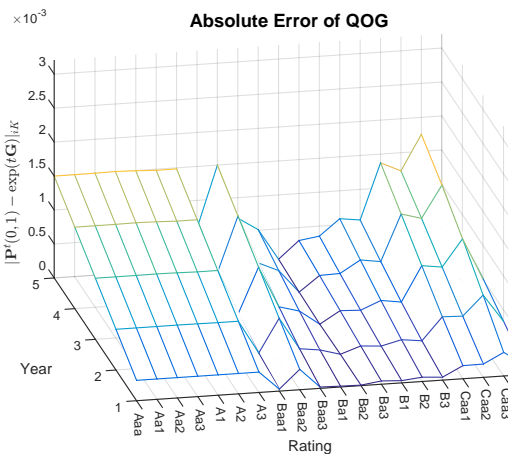
Table 4. Comparison of one- and five-year default probabilities in the basis points (bp) of QOG and BAM, for given transition matrix $\mathbf{P}^{\text{Moody's}}(0,1)$ in Table A1.

$(\mathbf{P}^{\text{Moody's}}(0,1))_{i,K}$		One Year ($t = 1$)				Five Year ($t = 5$)			
Constraints	Method	Aaa	Aa	A	Baa	Aaa	Aa	A	Baa
	Original	0.00	3.11	1.04	15.87	5.01	28.85	47.69	231.67
(G1), (G2)	QOG	0.17	3.12	1.42	15.89	6.35	29.69	49.17	231.92
	BAM	0.17	3.05	1.40	15.86	6.20	29.32	48.94	231.82
(G1), (G2), (D1)	QOG	3.00	3.49	3.00	16.01	18.04	32.56	55.81	233.29
	BAM	3.00	3.12	3.00	15.87	17.65	30.99	55.19	232.73
(G1), (G2), (D2)	QOG	0.12	2.08	2.08	15.95	5.39	26.11	51.73	232.46
	BAM	0.12	2.06	2.06	15.87	5.29	25.96	51.35	232.17
(G1), (G2), (D1), (D2)	QOG	3.00	3.00	3.00	16.01	17.65	30.66	55.70	233.26
	BAM	3.00	3.00	3.00	15.87	17.55	30.52	55.16	232.73
(G1), (G2), (M1), (M2)	QOG	0.16	2.93	1.41	15.89	5.36	25.88	48.97	231.88
	BAM	0.17	3.08	1.41	15.86	5.52	26.51	49.06	231.82
(G1), (G2), (R1)	QOG	0.07	1.28	1.99	15.09	4.36	21.77	49.44	222.44
	BAM	0.08	1.47	2.19	15.57	4.59	22.98	51.09	227.43
All	QOG	3.00	3.41	4.08	15.17	17.41	29.36	56.96	223.86
	BAM	3.00	3.42	4.12	15.50	17.46	29.70	58.22	228.29

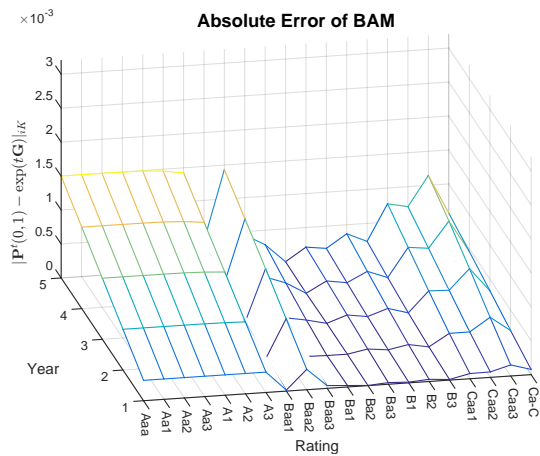
Table 5. Corrections of the additional Constraints (M1) and (M2) and their impact on the default probabilities of QOG and BAM, for transition matrix $\mathbf{P}^{\text{Moody's XXL}}(0,1)$ in Table A1.

Constraints	Method	$g_{18,19}^{\text{Moody's}} \text{ XXL}(0,1)$	$g_{18,20}^{\text{Moody's}} \text{ XXL}(0,1)$	$p_{18,21}^{\text{Moody's}} \text{ XXL}(0,1)$
	Original	0.0440	0.0838	0.1735
(G1), (G2), (M1), (M2)	QOG	0.0638	0.0638	0.1721
	BAM	0.0641	0.0641	0.1734

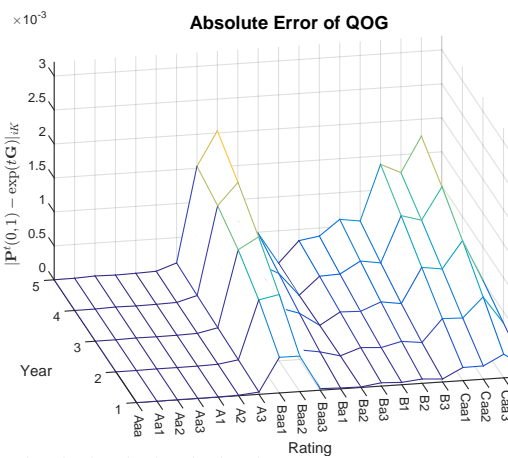
(a): (G1), (G2), (D1)



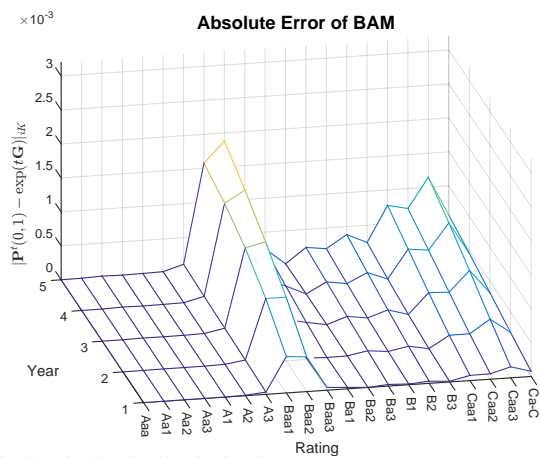
(b): (G1), (G2), (D1)



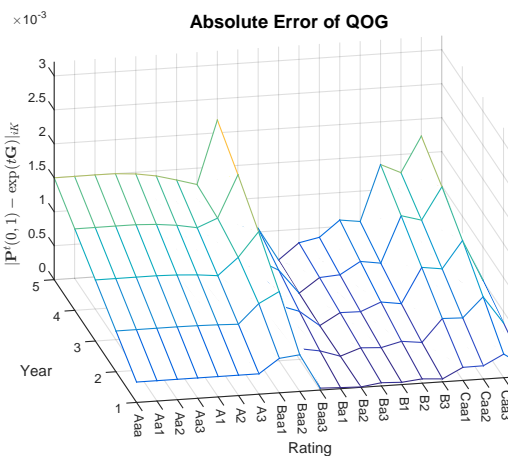
(c): (G1), (G2), (D2)



(d): (G1), (G2), (D2)



(e): (G1), (G2), (D1), (D2)



(f): (G1), (G2), (D1), (D2)

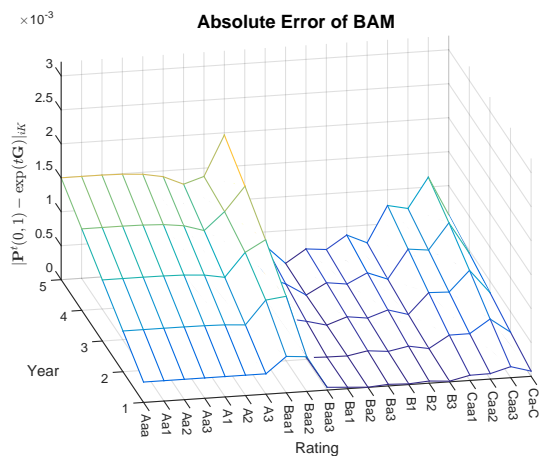
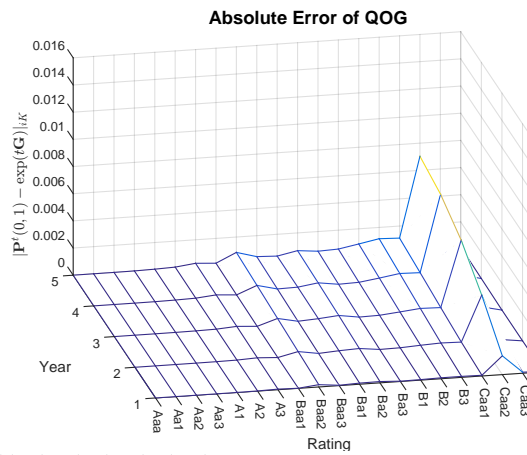
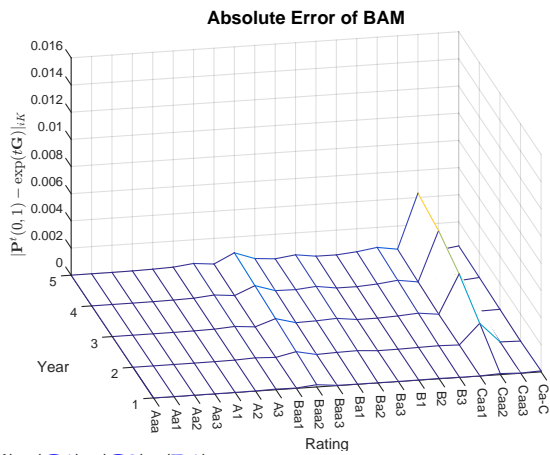


Figure 2. Absolute errors of Problems **QOG** and **BAM** based on various additional constraints with starting point X_0^{DMPG} , for transition matrix $P^{\text{MoodyS XXL}}(0, 1)$ in Table A1. Left side: (a,c,e) **QOG**; right side: (b,d,f) **BAM**.

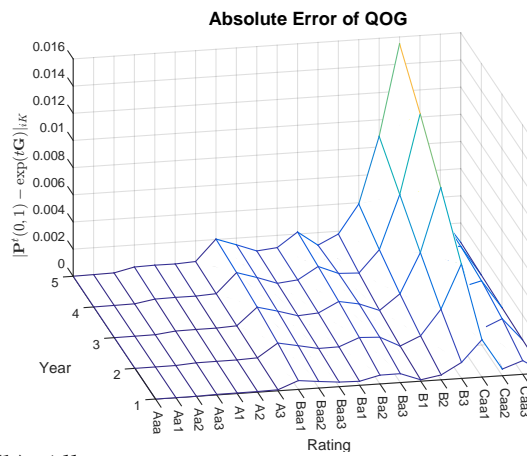
(g): (G1), (G2), (M1), (M2)



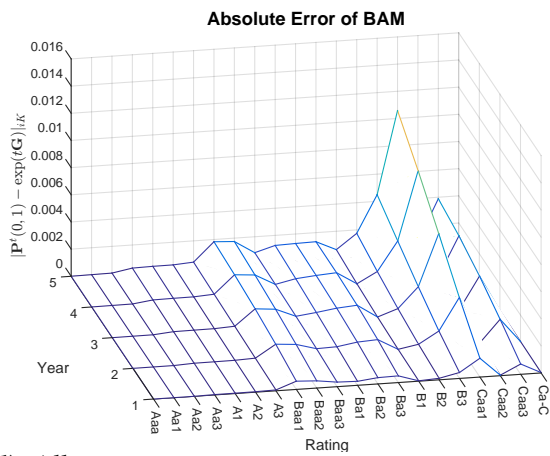
(h): (G1), (G2), (M1), (M2)



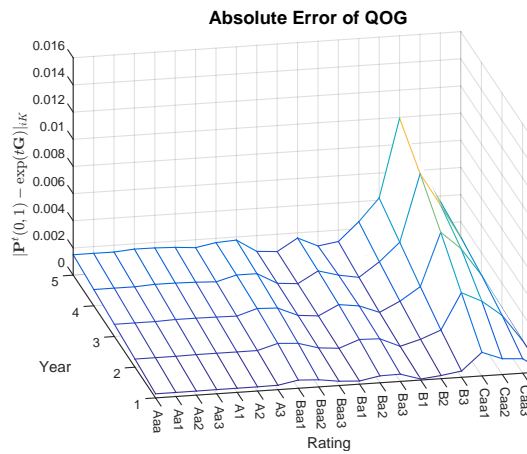
(i): (G1), (G2), (R1),



(j): (G1), (G2), (R1),



(k): All



(l): All

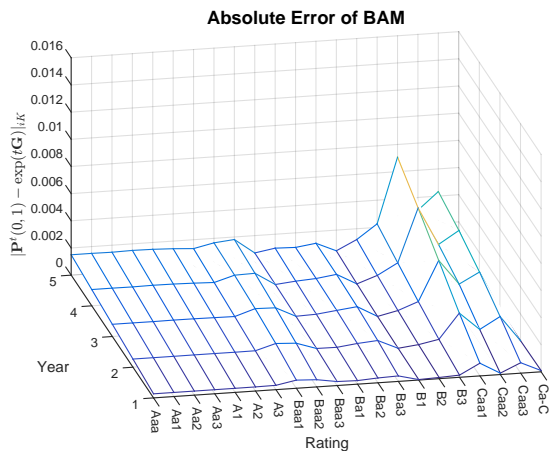


Figure 3. Absolute errors of Problems QOG and BAM based on various additional constraints with starting point X_0^{DMPG} , for transition matrix $P^{\text{Moodys XXL}}(0,1)$ in Table A1. Left side: (g,i,k) QOG; right side: (h,j,l) BAM.

Table 6. Corrections of the additional Constraint (R1) and its impact on the default probabilities of QOG and BAM, for transition matrix $P^{\text{Moody's XXL}}(0,1)$ in Table A1.

Constraints	Method	$\sum_{j \geq 14} g_{17,j}^{\text{Moody's XXL}}(0,1)$	$\sum_{j \geq 14} g_{18,j}^{\text{Moody's XXL}}(0,1)$	$p_{17,21}^{\text{Moody's XXL}}(0,1)$
(G1), (G2), (R1)	Original	−0.0059	−0.0173	0.0931
	QOG	−0.0189	−0.0189	0.0904
	BAM	−0.0163	−0.0163	0.0917

5. Conclusions

Embedding a discrete transition matrix in continuous time is not an easy undertaking, as pre-conditions need to be met beforehand. The analysis of given annual transition matrices has revealed that it is often the case that theoretically, a valid generator does not exist, due to non-observed defaults in historical data for high rating classes. However, optimization problems can deal with this shortcoming and provide the closest solutions.

Three related optimization techniques have been considered in the literature. While QOG can be solved via a linear row-by-row approach, it has been shown that it can be solved similarly easily with additional linear constraints by linear programming. Analogous observations apply to BAM, which is the actual optimization problem of interest. The results are interesting for several reasons. Firstly, the non-convex and non-linear problem BAM is not only solvable, but, as expected, also yields (slightly) better results than QOG. Secondly, BAM has additionally gained more credibility of being the best optimal solution by introducing a robust homogenization approach.

The most noteworthy additional value however is the fact that credit risk-relevant constraints can be incorporated (and easily extended) into the optimization problems. It is up to the practitioner to decide which constraints are necessary or even additionally needed. In doing so, crude pre-manipulation of the transition matrix can be avoided, but instead be directly incorporated into the optimization, yielding a more elegant approach.

Acknowledgments: This research paper is part of the project regarding the modeling of the Pfandbriefe (German covered bonds), financially supported by the Bayerische Staatsministerium für Wissenschaft, Forschung und Kunst and backed by the corporate partners Allianz Deutschland AG, Deutsche Pfandbriefbank AG and DEVnet GmbH. Special thanks to Dirk Banholzer and Georg Schlüchtermann and participants of the annual OR2015 conference for their valuable input.

Author Contributions: The authors, Maximilian Hughes and Ralf Werner, contributed equally to this research paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BAM	best approximation of the transition matrix
DA	diagonal adjustment
DMPG	distance minimization problem for the generator matrix
GS	global search
Homogen	homogenization
MCMC	Markov chain Monte Carlo
MS	multi-start
QOG	quasi-optimization of the generator
WA	weighted adjustment

References

1. Kreinin, A.; Sidelnikova, M. Regularization Algorithms for Transition Matrices. *Algo Res. Q.* **2001**, *4*, 23–40.
2. Davies, E. Embeddable Markov Matrices. *Electron. J. Probab.* **2010**, *15*, 1474–1486.
3. Stromquist, W. *Roots of Transition Matrices*; Practical Paper; Daniel H. Wagner Associates: Exton, PA, USA, 1996.
4. Araten, M.; Angbazo, L. *Roots of Transition Matrices: Application to Settlement Risk*; Practical Paper; Chase Manhattan Bank (JPMorgan Chase & Co.): New York, NY, USA, 1997.
5. Higham, N.J.; Lin, L. On p th Roots of Stochastic Matrices. *Linear Algebra Appl.* **2011**, *435*, 448–463.
6. Lin, L. Roots of Stochastic Matrices and Fractional Matrix Powers. Ph.D. Thesis, University of Manchester, Manchester, UK, 2011.
7. Charitos, T.; de Waal, P.R.; van der Gaag, L.C. Computing Short-Interval Transition Matrices of a Discrete-Time Markov Chain from Partially Observed Data. *Stat. Med.* **2008**, *27*, 905–921.
8. Inamura, Y. *Estimating Continuous Time Transition Matrices from Discretely Observed Data*; Working Paper; Financial Systems and Bank Examination Department; Bank of Japan: Tokyo, Japan, 2006.
9. Norris, J.R. *Markov Chains*; Number 2008 in Cambridge Series in Statistical and Probabilistic Mathematics; Cambridge University Press: Cambridge, United Kingdom, 1998.
10. Asmussen, S. Stochastic Modelling and Applied Probability. In *Applied Probability and Queues*; Springer: New York, NY, USA, 2008.
11. Bielecki, T.; Rutkowski, M. *Credit Risk: Modeling, Valuation and Hedging*; Springer Finance; Springer: Berlin, Germany, 2013.
12. Culver, W.J. On the Existence and Uniqueness of the Real Logarithm of a Matrix. *Proc. Am. Math. Soc.* **1966**, *17*, 1146–1151.
13. Singer, B.; Spilerman, S. The Representation of Social Processes by Markov Models. *Am. J. Sociol.* **1976**, *82*, 1–54.
14. Israel, R.B.; Rosenthal, J.S.; Wei, J.Z. Finding Generators for Markov Chains via Empirical Transition Matrices, with Applications to Credit Ratings. *Math. Financ.* **2001**, *11*, 245–265.
15. Elfving, G. Zur Theorie der Markoffschen Ketten. *Acta Soc. Sci. Fenn.* **1937**, *8*, 1–17.
16. Kingman, J.F.C. The Imbedding Problem for Finite Markov Chains. *Z. Wahrscheinlichkeitstheorie Verwandte Geb.* **1962**, *1*, 14–24.
17. Guerry, M.A. On the Embedding Problem for Discrete-Time Markov Chains. *J. Appl. Probab.* **2013**, *50*, 918–930.
18. Cuthbert, J. The Logarithmic Function for Finite-State Markov Semi-Groups. *J. Lond. Math. Soc.* **1973**, *6*, 524–532.
19. Johansen, S. Some Results on the Imbedding Problem for Finite Markov Chains. *J. Lond. Math. Soc.* **1974**, *2*, 345–351.
20. Frydman, H. The Embedding Problem for Markov Chains with Three States. *Math. Proc. Camb. Philos. Soc.* **1980**, *87*, 285–294.
21. Carette, P. Characterizations of Embeddable 3×3 Stochastic Matrices with a Negative Eigenvalue. *N. Y. J. Math.* **1995**, *1*, 120–129.
22. Guerry, M.A. On the Embedding Problem for Three-State Markov Chains. In Proceedings of the World Congress on Engineering 2014, London, UK, 2–4 July 2014; Volume II.
23. Chung, K. *Markov Chains with Stationary Transition Probabilities*; Die Grundlehren der mathematischen Wissenschaften; Springer: Berlin, Germany, 1960.
24. Grimmett, G.; Stirzaker, D. Probability and random processes. In *Probability and Random Processes*; OUP Oxford: Oxford, UK, 2001.
25. Goodman, G.S. An Intrinsic Time for Non-Stationary Finite Markov Chains. *Z. Wahrscheinlichkeitstheorie Verwandte Geb.* **1970**, *16*, 165–180.

26. Jarrow, R.A.; Lando, D.; Turnbull, S.M. A Markov Model for the Term Structure of Credit Risk Spreads. *Rev. Financ. Stud.* **1997**, *10*, 481–523.
27. Anderson, W. *Continuous-Time Markov Chains: An Applications-Oriented Approach*; Springer: New York, NY, USA, 2012.
28. Metz, A.; Cantor, R. Moody's Credit Policy—Introducing Moody's Credit Transition Model. Available online: <http://www.moodyanalytics.com/media/Brochures/Credit-Research-Risk-Measurement/Quantative-Insight/Credit-Transition-Model/Introductory-Article-Credit-Transition-Model.pdf> (accessed on 29 August 2016).



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).