# Quantifying Credit Portfolio sensitivity to asset correlations with interpretable generative neural networks

Sergio Caprioli[1,*,†], Emanuele Cagliero[2,†] and Riccardo Crupi[2,†]

[1]Intesa Sanpaolo S.P.A., Milano MI - 20121, Italy
[2]Intesa Sanpaolo S.P.A., Torino TO - 10138, Italy

## Abstract

In this research, we propose a novel approach for the quantification of credit portfolio Value-at-Risk (VaR) sensitivity to asset correlations with the use of synthetic financial correlation matrices generated with deep learning models. In previous work Generative Adversarial Networks (GANs) were employed to demonstrate the generation of plausible correlation matrices, that capture the essential characteristics observed in empirical correlation matrices estimated on asset returns. Instead of GANs, we employ Variational Autoencoders (VAE) to achieve a more interpretable latent space representation. Through our analysis, we reveal that the VAE latent space can be a useful tool to capture the crucial factors impacting portfolio diversification, particularly in relation to credit portfolio sensitivity to asset correlations changes.

## Keywords

Variational Autoencoder, VAE, Credit Portfolio Model, Concentration risk, Interpretable neural networks, Generative neural networks

## 1. Introduction

### 1.1. Credit Portfolio concentration risk

One of the most adopted models to measure the credit risk of a loan portfolio was proposed in [1] and it is currently a market standard used by regulators for capital requirements [2]. This model provides a closed-form expression to measure the risk in the case of asymptotic single risk factor (ASRF) portfolios. The ASRF model is portfolio-invariant, i.e., the capital required for any given loan only depends on the risk of that loan, regardless of the portfolio it is added to. Hence the model ignores the concentration of exposures in bank portfolios, as the idiosyncratic risk is assumed to be fully diversified. Under the Basel framework, Pillar I capital requirements for credit risk do not cover concentration risk hence banks are expected to autonomously estimate such risk and set aside an appropriate capital buffer within the Pillar II process [3].

A commonly adopted methodology of measuring concentration risk, in the more general case of a portfolio exposed to multiple systematic factors and highly concentrated on a limited number of loans, is to use a Monte Carlo simulation of the portfolio loss distribution under the assumption reported in [4]. The latter states that the standardized value of the $i$-th counterparty, $V_i$, is driven by a factor belonging to a set of macroeconomic Gaussian factors $\{Y_j\}$ and by an idiosyncratic independent Gaussian process $\varepsilon_i$:

$$V_i = \rho_i Y_j + \sqrt{1 - \rho_i^2}\, \varepsilon_i = \sum_f \rho_i \alpha_{j,f} Z_f + \sqrt{1 - \rho_i^2}\, \varepsilon_i \tag{1}$$

through a coefficient $\rho_i$. The systematic factors $\{Y_j\}$ are generally assumed to be correlated, with correlation matrix $\Sigma$. The third term of Eq. 1 makes use of the spectral decomposition $\Sigma = \alpha \alpha^T$ to express $V_i$ as a linear combination of a set of uncorrelated factors $\{Z_f\}$, allowing for a straightforward Monte Carlo simulation.

The bank's portfolio is usually clustered into sub-portfolios that are homogeneous in terms of risk characteristics (i.e. industrial sector, geographical area, rating class or counterparty size). A distribution of losses is simulated for each sub-portfolio and the Value at Risk (VaR) is calculated on the aggregated loss.

The asset correlation matrix $\Sigma$ is a critical parameter for the estimation of the sub-portfolio loss distribution, that is the core component for the estimation of the concentration risk. Therefore it is worth assessing the credit portfolio VaR sensitivity to that parameter.

## 1.2. Sampling Realistic Financial Correlation Matrices

As reported in [5],

> "markets in crisis mode are an example of how assets correlate or diversify in times of stress. It is essential to see how markets, asset classes, and factors change their correlation and diversification properties in different market regimes. [...] It is desirable not only to consider real manifestations of market scenarios from history but to simulate new, realistic scenarios systematically. To model the real world, quants turn to synthetic data, building artificially generated data based on so-called market generators."

Marti [6] proposed Generative Adversarial Networks (GANs) to generate plausible financial correlation matrices. The author shows that the synthetic matrices generated with GANs present most of the properties observed on the empirical financial correlation matrices estimated on asset returns. In line with [6] we generated synthetic asset correlation matrices verifying some "stylized facts" of financial correlations.

We used a different type of neural network, Variational Autoencoders (VAE), to map historical correlation matrices onto a bidimensional "latent space", also referred to as the bottleneck of the VAE. After training a VAE on a set of historical asset correlation matrices, we show that it is possible to explain the location of points in the latent space. Furthermore, analyzing the relationship between the VAE bidimensional bottleneck and the VaR values computed by the Credit Portfolio Model using different historical asset correlation matrices, we show that the distribution of the latent variables encodes the main aspects impacting portfolio's diversification as presented in [7].

## 2. Sensitivity to the Asset Correlation matrix

### 2.1. Data

The dataset contains $n = 206$ correlation matrices of the monthly log-returns of $M = 44$ equity indices, calculated on their monthly time series from February 1997 to June 2022, using overlapping rolling windows of size 100 months. Historical time series considered are Total Market (Italy, Europe, US and Emerging Markets) and their related sector indices (Consumer Discretionary, Energy, Basic Materials, Industrials, Consumer Staples, Telecom, Utilities, Technology, Financials, Health Care), the source is Datastream.

### 2.2. Variational Autoencoder design

An autoencoder is a neural network composed of a sequence of layers ("encoder" E) that perform a compression of the input into a low-dimensional "latent" vector, followed by another sequence of layers ("decoder" D) that approximately reconstruct the input from the latent vector. The encoder and decoder are trained together to minimize the difference between original input and its reconstructed version.

Variational Autoencoders [8] consider a probabilistic latent space defined as a latent random variable $z$ that generated the observed samples $x$. Hence the "probabilistic decoder" is given by $p(x|z)$ while the "probabilistic encoder" is $q(z|x)$. The underlying assumption is that the data are generated from a random process involving an unobserved continuous random variable $z$ and it consists of two steps: (1) a value $z_i$ is generated from some prior distribution $p_\theta^*(z)$, (2) a value $\hat{x}_i$ is generated from some conditional distribution $p_\theta^*(x|z)$. Assuming that the prior $p_\theta^*(z)$ and the likelihood $p_\theta^*(x|z)$ come from parametric families of distributions $p_\theta(z)$ and $p_\theta(x|z)$, and that their PDFs are differentiable almost everywhere w.r.t. both $\theta$ and z, the algorithm proposed by [8] for the estimation of the posterior $p_\theta(z|x)$ introduces an approximation $q_\phi(z|x)$ and minimizes the Kullback-Leibler (KL) divergence of the approximate $q_\phi(z|x)$ from the true posterior $p_\theta(z|x)$. Using a multivariate normal as the prior distribution, the loss function is composed of a deterministic component (i.e. the mean squared error MSE) and a probabilistic component (i.e. the Kullback-Leibler divergence from the true posterior):

$$KL = -\frac{1}{2\bar{n}} \sum_{i=1}^{\bar{n}} \sum_{k=1}^{2} (1 + log(\sigma_k^{i\,2}) - \mu_k^{i\,2} - \sigma_k^{i\,2})$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{\bar{n}} \parallel \mathbf{x}_i - \bar{\mathbf{x}}_i \parallel_2^2 = \frac{1}{n} \sum_{i=1}^{\bar{n}} \parallel \mathbf{x}_i - D(E(\mathbf{x}_i)) \parallel_2^2$$

$$\text{Loss} = \text{MSE} + \beta \cdot \text{KL}$$

(2)

where $E$ and $D$ are the encoding and decoding map respectively, $E : \mathbf{x} \in \mathbf{R}^{M \times M} \longrightarrow \theta = \{\mu_1, \mu_2, \sigma_1, \sigma_2\} \in \mathbf{R}^4$, $D : \mathbf{z} \in \mathbf{R}^2 \longrightarrow \bar{\mathbf{x}} \in \mathbf{R}^{M \times M}$, $\mathbf{z} = \mu + \sigma \odot \varepsilon$, $\mu = \{\mu_1, \mu_2\}, \sigma = \{\sigma_1, \sigma_2\}, \varepsilon$ is a bivariate standard Gaussian variable, and $\bar{n} < n$ is the number of samples in the training set.

In this equation, $\mu_k^i$ and $\sigma_k^i$ represent the mean and standard deviation of the $k$-th dimension of the latent space for the sample $\mathbf{x}_i$. The loss function balances the MSE, reflecting the
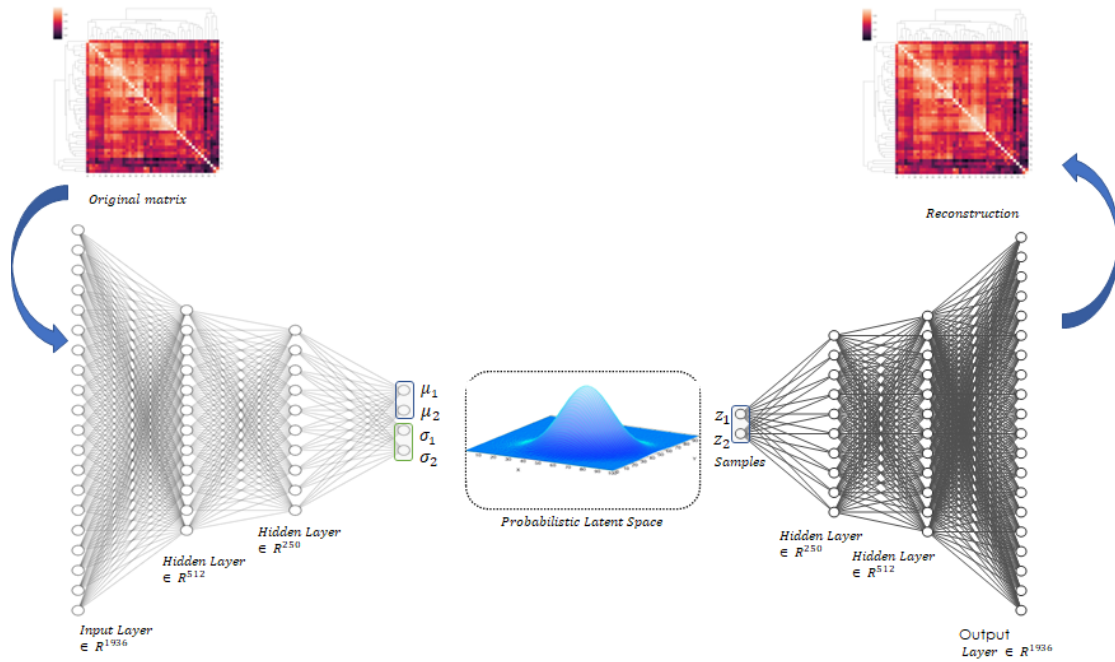
**Figure 1:** VAE Framework: The input layer comprises 1936 nodes, corresponding to the $44 \times 44$ matrix input. Subsequently, there are layers with 512, 250, and a central hidden layer with 4 nodes. These values represent the mean and variance of a bivariate gaussian distribution. The decoder receives as input two values sampled from the latent space and it is asked to reconstruct the input. Hence, the architecture is symmetrically mirrored until the output layer, which also has 1936 nodes.

reconstruction quality, with $\beta$ times the KL divergence, enforcing a distribution matching in the 2-dimensional latent space. The KL divergence can be viewed as a regularizer of the model and $\beta$ as the strength of the regularization.

We trained the VAE for 80 epochs using a learning rate of 0.0001 with an Adam optimizer. The structure of the VAE is shown in Fig. 1. We randomly split the dataset described in section 2.1 in a training sample, used to train the network, and a validation set used to evaluate the performance. We used 30% of the dataset as validation set.

Variational Autoencoders were employed in previous works for financial applications. In particular Brugier and Turinici [9] proposed a VAE to compute an estimator of the Value at Risk for a financial asset. Bergeron et al. [10] used VAE to estimate missing points on partially observed volatility surfaces. Sokol [11] applied VAEs for interest rate curves simulation.

## 2.3. Comparison with linear models

We compared the performances of the Variational Autoencoders with the standard Autoencoder (AE) and with the linear autoencoder (i.e. the autoencoder without activation functions).

The linear autoencoder is equivalent to apply PCA to the input data in the sense that its output is a projection of the data onto the low dimensional principal subspace [12]. As shown in Fig. 2b the autoencoder performs better than VAE (Fig. 2a), while linear models have lower
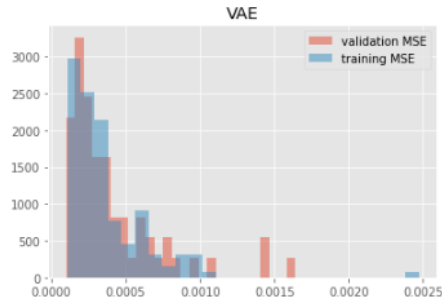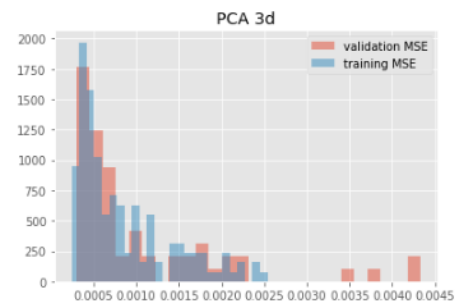
(a) Variational Autoencoder (VAE)
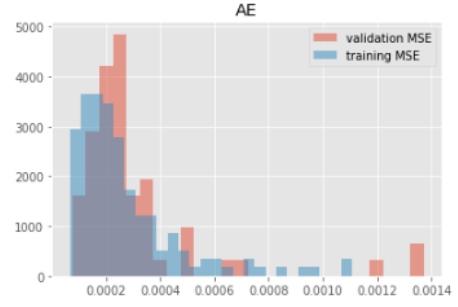
(b) Autoencoder (AE)

**Figure 2:** Histogram of Mean squared error (MSE) of the Autoencoder and Variational Autoencoder on the historical correlation matrices, split into train and validation set.



(a) 2-dimensional linear autoencoder (PCA 2d)

(b) 3-dimensional linear autoencoder (PCA 3d)

**Figure 3:** Histogram of Mean squared error (MSE) of the linear autoencoders on the historical correlation matrices, split into train and validation set.

performances (Fig. 3a) even increasing the dimensions of the latent space (Fig. 3b). Hence, neural networks actually bring an improvement in minimizing the reconstruction error. The generative probabilistic component of VAE decreases the performance when compared to a deterministic autoencoder. On the other hand, it allows to generate new but realistic correlation matrices in the sense of stylized facts.

## 2.4. Latent space interpretability

According to Miller [13] and Lipton [14]:

*Interpretable* is a model such that an observer can understand the cause of a decision.

*Explanation* is one mode in which an observer may obtain understanding in the latent space, for instance, building a simple surrogate model that mimics the original model to gain a better understanding of the original model's underlying mechanics.

For the sake of our analysis, we refer to the "interpretability" of the VAE as the possibility to understand the reason underlying the responses produced by the algorithm in the latent space. The Variational Autoencoder projected the 206 historical correlation matrices on a two
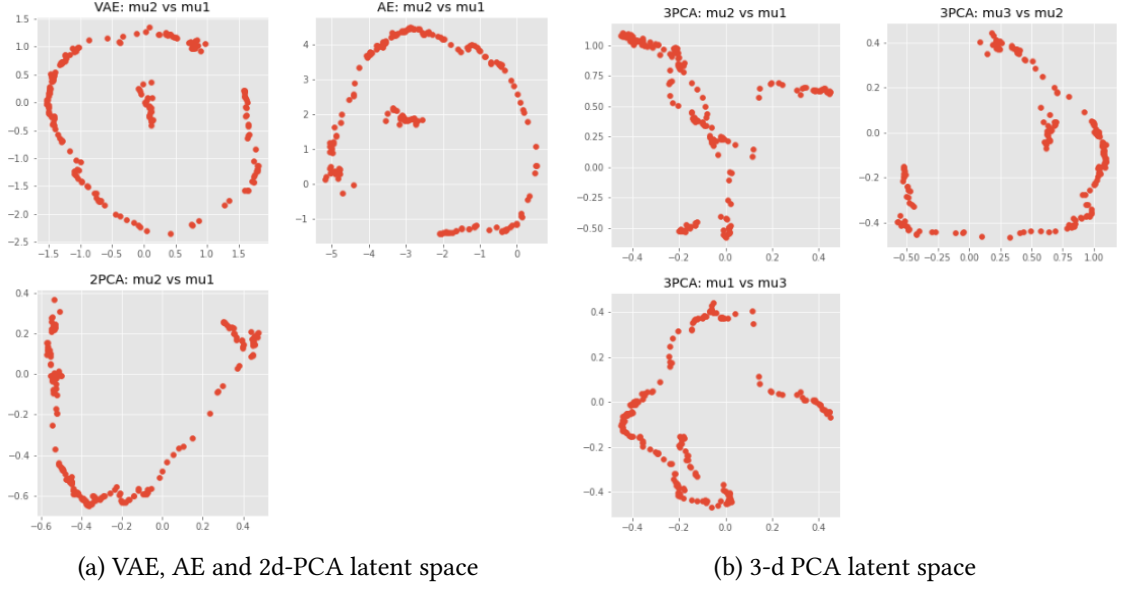
(a) VAE, AE and 2d-PCA latent space        (b) 3-d PCA latent space

**Figure 4:** Comparison of the latent space generated with different models. The latent space generated by the VAE and AE are similar while the cluster of points in the middle is recovered only by the 3-dimensional linear autoencoder.

dimensional probabilistic latent space represented by a bivariate normal distribution. As shown in Fig. 4a, the latent space generated by the VAE and AE are similar while the cluster of points in the middle is recovered only by the 3-dimensional linear autoencoder (Fig. 4b).

In order to understand the rationales underlying such representation, we analysed the relationship of the encoded values of the original correlation matrices with respect to their eigenvectors $\{v_i \mid i = 1 : M\}$ and eigenvalues $\{\lambda_i \mid i = 1 : M\}$. It turned out that the first component of the latent space ($\mu_1$) is strongly negatively correlated to the first eigenvalue (Fig. 5).

As pointed out in [15]

> "the largest eigenvalue of the correlation matrix is a measure of the intensity of the correlation present in the matrix, and in matrices inferred from financial returns tends to be significantly larger than the second largest. Generally, this largest eigenvalue is larger during times of stress and smaller during times of calm."

Hence, the first dimension of the latent space seems to capture the information related to the rank of the matrix i.e. to the "diversification opportunities" on the market. The interpretation of the second dimension ($\mu_2$) of the latent space turned out to be related to the eigenvectors of the correlation matrix. In order to understand the other dimension we consider the cosine similarity $\alpha_t^i$ between the $i$-th eigenvector at time $t$ and its average over time. Formally:

$$\alpha_{i,t} = \frac{\frac{1}{n}(\sum_{t=1}^{n} v_{i,t})^T \cdot v_{i,t}}{\| \frac{1}{n}(\sum_{t=1}^{n} v_{i,t})^T \| \| v_{i,t} \|} \tag{3}$$

where $i$ is the number of the eigenvector and $t$ the index of the matrix in the dataset.
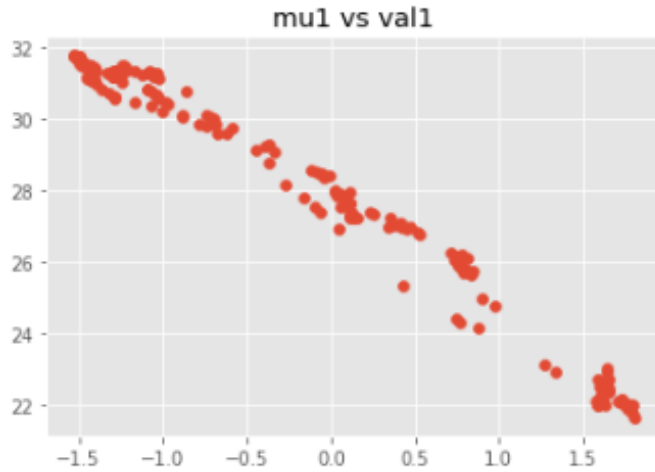
**Figure 5:** Scatterplot of the first eigenvalue $\lambda_1$ versus the first component of the latent space $\mu_1$, showing a clear negative correlation.

Let us define $\alpha_1 = \{\alpha_{1,t}\}_{t=1,\dots n}$, $\alpha_2 = \{\alpha_{2,t}\}_{t=2,\dots n}$. The data point subgroups observed in the space $(\alpha_1, \alpha_2, \lambda_1)$ can be traced to corresponding subgroups in the latent space $(\mu_1, \mu_2)$, as shown in Fig. 6.

As pointed out in [7], each eigenvector can be viewed as a portfolio weights of stocks that defines a new index which is uncorrelated with the other eigenvectors. It follows that a change in eigenvectors can impact portfolio diversification. We can conclude that the VAE latent space effectively captures, in two dimensions, the main factors impacting the financial correlations, which is determinant for portfolio diversification.

### 2.5. Generating synthetic correlation matrices

As explained in Section 2.2, the probabilistic decoder of the VAE allows to generate a "plausible" correlation matrix starting from any point of the latent space. Hence, we defined a grid of 132 points of the latent space that cover approximately homogeneously an area centered around the origin and including the historical points. For each point on the grid, we used the decoder (described in Section 2.2) to compute the corresponding correlation matrix. Along the lines of [6], we checked whether the following stylized facts of financial correlation matrices hold for both the historical and the synthetic matrices.

- The distribution of pairwise correlations is significantly shifted towards positive values.
- Eigenvalues follow the Marchenko–Pastur distribution, except for a very large first eigenvalue and a couple of other large eigenvalues.
- The Perron-Frobenius property holds true (first eigenvector has positive entries).
- Correlations have a hierarchical structure.
- The Minimum Spanning Tree (MST) obtained from a correlation matrix satisfies the scale-free property.

We verified that the distributions of pairwise correlations are shifted to the positive and that the distributions of the eigenvalues (each averaged respectively over the historical and synthetic matrices) are very similar to each other and can be approximated by a Marchenko-Pastur distribution, but for a first large eigenvalue and a couple of other eigenvalues. Regarding the Perron-Frobenius property, we verified that the eigenvector corresponding to the largest eigenvalue has strictly positive components. Inspecting the dendrogram of the correlation matrices we confirmed the presence of a hierarchical structure. Finally, the distribution of the degrees of the Minimum Spanning Tree (calculated on the mean of the matrices) is compatible with the scale-free property, i.e. very few nodes have high degrees while most nodes have degree 1. It is worth noting that the correlation matrices analyzed for our purposes were calculated starting from 44 equity indices (as explained in Section 2.1) instead of single stocks as shown in [6], hence a higher degree of concentration was expected.

### 2.6. Quantifying the sensitivity to asset correlations

For each matrix generated with the VAE probabilistic decoder, we estimated the corresponding VaR according to the multi-factor Vasicek model described in section 1.1. We used the VaR metric to show a proof of concept of the methodology and to be aligned to the Economic Capital requirements, but the same rationale can be followed adopting a different risk metric. As mentioned in section 1.1, Vasicek multi-factor model cannot be solved in closed form solution, hence it is necessary to run a Monte Carlo simulation for each generated matrix. We used a stratified sampling simulation with 1 million runs. In each estimation, the parameters of the model and portfolio exposures are held constant. Running the simulation for every sampled point of the latent space, we derived the VaR surface of Fig. 7.

To obtain an estimate of the sensitivity of the VaR to future possible evolutions of the correlation matrix, we "bootstrapped" (see Fig. 9) the historical time series of the points in the 2-dimensional latent space. We used a simple bootstrapping [16] and a block-bootstrapping technique [17] on the differences' time series of the two components of the VAE latent space, $\mu_1$ and $\mu_2$ (depicted in Fig. 8).

Interpolating the estimated VaR over the randomly sample grid (Fig. 7) we can derive the Value at Risk corresponding to any point of the latent space. Hence, for each point belonging to the distribution of correlations changes over a 1-year time horizon estimated via bootstrap, we can calculate the corresponding VaR without recurring to the Monte Carlo simulation.

In this way, we obtained the VaR distribution related to the possible variations of correlation matrices on a defined time horizon.

## 3. Conclusions

In this work we applied a Variational Autoencoder for generating realistic financial matrices that we used as input for the estimation of credit portfolio concentration risk estimated with a multi-factor Vasicek model. We deviated from the methodology proposed by G. Marti 2020 [6] who adopted a Generative Adversarial Network, in order to obtain an interpretable model by leveraging the dimensionality reduction provided by VAE. Using as a proof of concept a VAE trained on a dataset composed of 206 correlation matrices calculated on the time series of 44

equity indices using a rolling window of 100 months, we showed how it is possible, even using a small data sample, to derive an interpretation of the latent space that seems aligned to the main aspects driving portfolio diversification [7].

We exploited the generative capabilities of the VAE to extend the scope of the model beyond the necessarily limited size of the historical sample, generating a larger set of correlation matrices that retain all the realistic features observed in the market. Therefore, the VAE has primarily been utilised for data augmentation, assessing its efficacy in terms of the quality of the artificially generated matrices, determined by suitably testing the stylized facts known about the financial correlation matrices.

We computed the augmented sample of synthetic correlation matrices on a grid in the 2-dimensional VAE latent space and for each synthetic matrix the corresponding credit portfolio loss distribution (and its VaR at a certain percentile) was obtained via Monte Carlo simulation under a multi-factor Vasicek model. This way we estimated a VaR surface over the VAE latent space.

Analyzing the time series of the encoded version of the correlation matrices (i.e. the two components of the probabilistic latent space) we easily estimated (via bootstrapping) the possible variation of the correlation matrices over a 1-year time horizon. Finally, using the interpolated VaR surface, we were able to estimate the corresponding VaR distribution obtaining a quantification of the impact of the correlations movements on the credit portfolio concentration risk. This approach gives rapid estimation of risk without depending on the extensive computations of Monte Carlo simulation, and it does so in a compressed, easy-to-visualize space that captures several aspects of market dynamics.

Our analysis provides clear indications that the capabilities of realistic data-augmentation provided by Variational Autoencoders combined with the ability to obtain model interpretability can prove useful for risk management purposes, when addressing the sensitivity of models on a structured multidimensional market data as the correlation matrix.

## Disclaimer

The views and opinions expressed within this paper are those of the authors and do not necessarily reflect the official policy or position of Intesa Sanpaolo. Assumptions made in the analysis, assessment, methodology, model and results are not reflective of the position of any entity other than the authors.

## References

[1] O. A. Vasicek, Probability of loss on loan portfolio, KMV, 1987.

[2] Regulation (EU) no 575/2013 of the European Parliament and of the Council of 26 June 2013 on prudential requirements for credit institutions and investment firms and amending Regulation (EU) no 648/2012, Official Journal of the European Union (2015).

[3] C. of European Banking Supervisors, Cebs guidelines on the management of concentration risk under the supervisory review process (gl31) (2010).

[4] P. Grippa, L. Gornicka, Measuring concentration risk-A partial portfolio approach, International Monetary Fund, 2016.

[5] J. Papenbrock, P. Schwendner, M. Jaeger, S. Krügel, Matrix evolutions: synthetic correlations and explainable machine learning for constructing robust investment portfolios, The Journal of Financial Data Science 3 (2021) 51–69.

[6] G. Marti, Corrgan: Sampling realistic financial correlation matrices using generative adversarial networks, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 8459–8463.

[7] H. T. Nguyen, P. N. Tran, Q. Nguyen, An analysis of eigenvectors of a stock market cross-correlation matrix, in: Econometrics for Financial Applications, Springer, 2018, pp. 504–513.

[8] D. P. Kingma, M. Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114 (2013).

[9] P. Brugière, G. Turinici, Deep learning of value at risk through generative neural network models: The case of the variational auto encoder, MethodsX 10 (2023) 102192.

[10] M. Bergeron, N. Fung, J. Hull, Z. Poulos, A. Veneris, Variational autoencoders: A hands-off approach to volatility, The Journal of Financial Data Science (2022).

[11] A. Sokol, Autoencoder market models for interest rates, Available at SSRN 4300756 (2022).

[12] E. Plaut, From principal subspaces to principal components with linear autoencoders, arXiv preprint arXiv:1804.10253 (2018).

[13] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, Artificial intelligence 267 (2019) 1–38.

[14] Z. C. Lipton, The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery., Queue 16 (2018) 31–57.

[15] T. Millington, M. Niranjan, Construction of minimum spanning trees from financial returns using rank correlation, Physica A: Statistical Mechanics and its Applications 566 (2021) 125605.

[16] S. Abney, Bootstrapping, in: Proceedings of the 40th annual meeting of the Association for Computational Linguistics, 2002, pp. 360–367.

[17] M. Mader, W. Mader, L. Sommerlade, J. Timmer, B. Schelter, Block-bootstrapping for noisy data, Journal of neuroscience methods 219 (2013) 285–291.
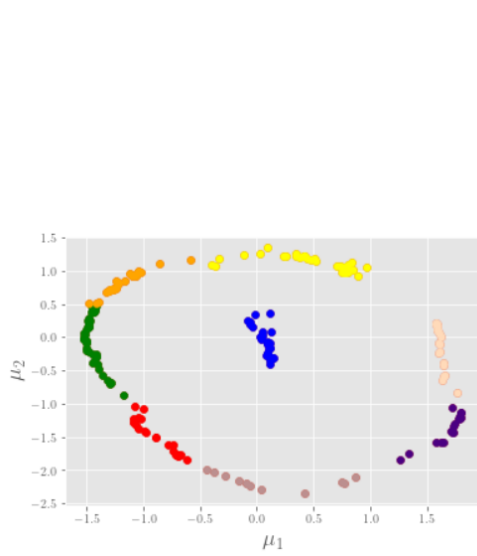
(a) The points on the latent space $(\mu_1, \mu_2)$ representing the historical correlation matrices. The latent space was conventionally partitioned in nine subgroups of data points identified by different colors.
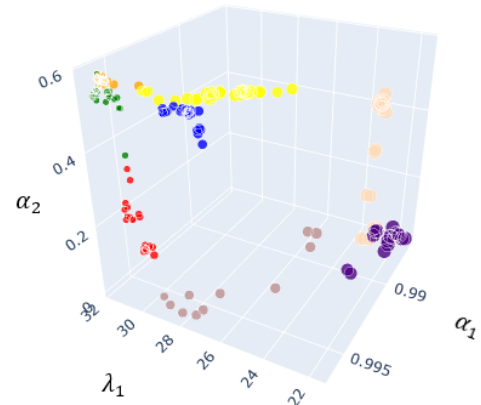


(b) The data points of Fig. (a) represented in the parameter space defined by $\alpha_1$, $\alpha_2$, and $\lambda_1$ (also, the size of each dot corresponds to the value of $\lambda_2$). The proximity of these data points consistently mirrors the subgroups illustrated in Fig. (a), with matching color schemes. There is a noticeable separation between different subgroups, and this separation is well-defined in most cases.
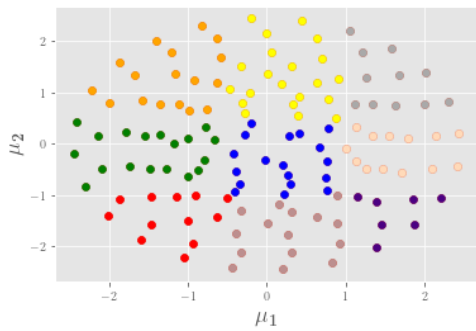


(c) Sampling in the latent space $(\mu_1, \mu_2)$: each point can be decoded into a synthetic correlation matrix. The latent space was conventionally partitioned in nine regions identified by different colors, with the same convention as Fig (a).
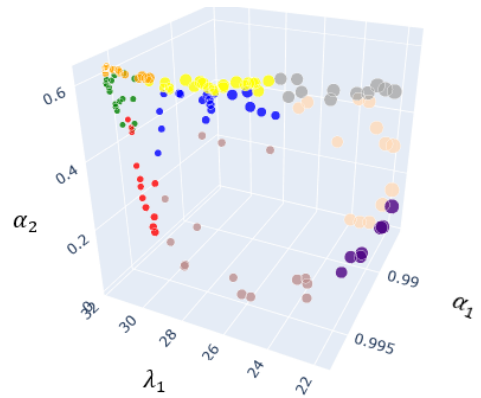


(d) The sampled points of Fig. (c), plotted with the same color schemes in the space formed by $\alpha_1$, $\alpha_2$ and $\lambda_1$ (also, the size of each dot corresponds to the value of $\lambda_2$). Similar observations and considerations can be drawn here as those made for Fig. (a) and Fig. (b).

**Figure 6:** The distribution of the distance of the first two eigenvectors from their respective historical average and the distribution of the first eigenvalue characterize the regions of the latent space.
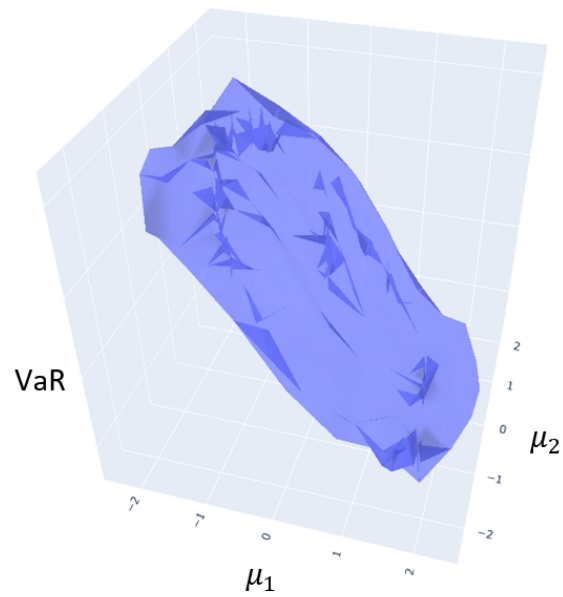
**Figure 7:** The surface generated from Value at Risk with respect to the points of the 2-dimensional latent space.
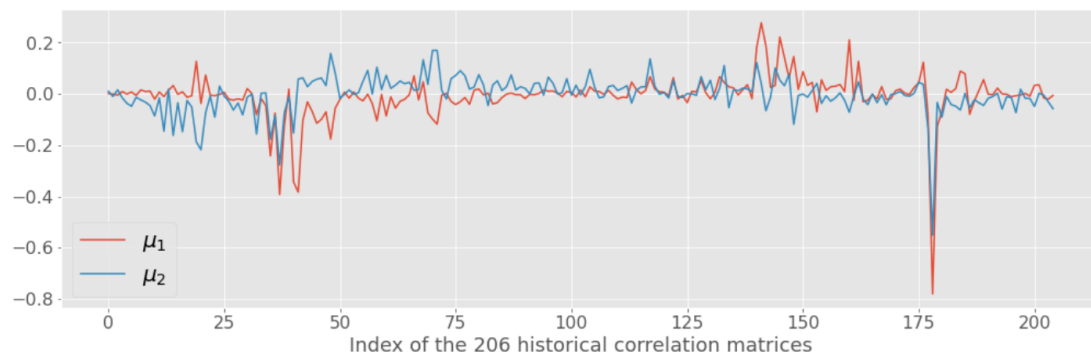


**Figure 8:** The time series of the variations of $\mu_1$ and $\mu_2$, projections of the 206 historical correlation matrices in the 2-dimensional latent space.
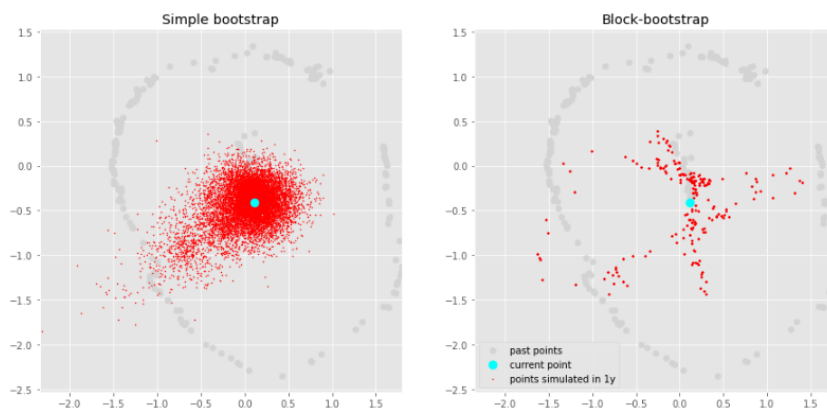
**Figure 9:** Using simple bootstrap (on the left) or block-bootstrap (with 11 monthly steps) on the "compressed" representation of the correlation matrices, we estimated the distribution of the possible variation of the current matrix on 1-year time horizon.