# Monitoring time-varying systemic risk in sovereign debt and currency markets with generative AI

Sabuhi Khalili,

Faculty of Economics and Business, Riskcenter and UB-School of Economics, University of Barcelona (UB),
Avinguda Diagonal 690-696,08034,
Contact: *sabuhi.khalili@ub.edu*

Jorge M. Uribe,

Faculty of Economics and Business, Universitat Oberta de Catalunya (UOC),
Rambla de Poblenou 156, 08018,
Contact: *juribeg@uoc.edu*

# Monitoring time-varying systemic risk in sovereign debt and currency markets with generative AI


## Abstract

We propose generative artificial intelligence to measure systemic risk in the global markets of sovereign debt and foreign exchange. Through a comparative analysis, we explore three novel models to the economics literature and integrate them with traditional factor models. These models are: Time Variational Autoencoders, Time Generative Adversarial Networks, and Transformer-based Time-series Generative Adversarial Networks. Our empirical results provide evidence in support of the Variational Autoencoder. Results here indicate that both the Credit Default Swaps and foreign exchange markets are susceptible to systemic risk, with a historically high probability of distress observed by the end of 2022, as measured by both the Joint Probability of Distress and the Expected Proportion of Markets in Distress. Our results provide insights for governments in both developed and developing countries, since the realistic counterfactual scenarios generated by the AI, yet to occur in global markets, underscore the potential worst-case scenarios that may unfold if systemic risk materializes. Considering such scenarios is crucial when designing macroprudential policies aimed at preserving financial stability and when modeling the effectiveness of the implemented policies.



**Keywords:** Twin Ds, Sovereign Debt, Credit Risk, TimeGANs, Transformers, TimeVAEs, Autoencoders, Variational Inference,

**JEL Codes:** C45, C53, F31, F37

## 1. Introduction

We use sate-of-the-art generative Artificial Intelligence (AI) to assess time-varying systemic risk in global markets for sovereign debt and foreign exchange, both of which are susceptible to what the literature identifies as the "Twin Ds - Default and Devaluation" risks.[1]. We model the time series of sovereign Credit Default Swaps (CDS) spreads and foreign exchange rates against the US dollar (FX) of 35 countries. These two markets have become a top priority for central banks and regulators due to concerns about sustainability of historically high levels of debt in emerging and low-income countries, coupled with rising interest rates and inflation above its long-run target, as well as currency depreciation against the US dollar driven by market expectations of further monetary policy tightening.

It is crucial to gain a deeper economic understanding of currencies and sovereign debt markets, as current global financial conditions affecting large financial intermediaries in advanced economies might nonlinearly translate to the dynamics of both markets (Gilchrist et al., 2022). This may lead to systemic risk episodes with global reach, potentially triggering contagion and spillovers not only within sovereign debt markets (e.g., Longstaff et al, 2011; Yacine Ait-Sahalia et al., 2014; Wu et al, 2016; Arellano et al., 2017; Dawood et al., 2017; Augustin et al., 2022) or FX markets (e.g., Chuliá et al., 2018), but also between the two markets (e.g., Du and Schreger, 2016). The existing literature lacks a comprehensive empirical framework that enables policymakers to effectively monitor risks in these two global markets, which we address with our contribution.

Importantly though, we approach this research question from a systemic-risk point of view. This means that although our research is clearly related to the large literature on covered and uncovered interest rate parity (e.g. Lustig and Verdelhan, 2007; Kalemli-Özcan and Varela, 2021; Borio et al., 2016; Du et al., 2018) and asset pricing approaches to the joint risks (Duffie et al., 2003; Carr et al. 2007; Lettau et al., 2014; Na et al., 2018; Augustin et al., 2018; Augustin et al., 2020; Chernov et al., 2023), the main analytical toolkit here is innovative; in the sense that it acknowledges an overlooked fact by the past literature: debt and currency markets cannot be fully understood in isolation or within a linear framework. They interact in complex and nonlinear systems that need to be correctly modeled if one wants to accurately predict the probability of systemic risk realizations.

[1] Starting with Garcia and Lowenkros (2005), other authors have also referred to the combined risk of currency and default as 'cousin risks.'

To this end, we resort to three cutting-edge AI models: Time-series-Generative Adversarial Networks (GANs), Transformers GANs, and Variational Autoencoders (VAEs), as well as traditional factor models in econometrics, in a two-step approach that is inspired by Fan et al. (2021) and Fan et al. (2022). Our objective is to generate realistic scenarios of the markets on a high frequency basis, for both CDS and FX markets, and for the two markets combined.

In the first step, we use Principal Component Analysis (PCA) to model the linear factors underlying the time series variation of all the series in our relatively large set of prices in each market. In the second step, we apply generative AI to the idiosyncratic components of the series to generate joint market scenarios. These simulations allow us to calculate statistics of systemic risk, in line with prior literature, such as the Joint Probability of Markets in Distress (JPD) and Expected Proportion of Markets in Distress (EPD). See, for instance, Hartmann et al. (2006) and Oh and Patton (2018).

In essence, we surpass the limitations of modeling sovereign debt and currency markets individually within a linear paradigm, a practice common in previous literature. Moreover, we pioneer the integration of these two markets within a comprehensive multivariate framework. The limitations that we overcome are particularly relevant in the twin markets where nonlinearities play a crucial role. For example, debt thresholds for safe government deficits, sudden stops, currency crises, interactions between market and funding liquidity of global intermediaries, margin-calls above certain risk thresholds, and fiscal crises are all expected to produce nonlinearities in the dynamics of these markets.

Our overall contribution is twofold: we propose a new method for tracking systemic risk on a high-frequency and accurate basis using a new generation of AI models, which have produced impressive advancements not only within traditional AI subfields like computer vision and natural language processing but also in entire fields such as medicine and biochemistry, yet remain uncharted in the field. Secondly, we present the first estimates in the literature of systemic risk in the Twin markets for policy assessment, which emphasize the historically high vulnerability in these markets, especially in the emerging world.

Our findings highlight significant increases in sovereign risk and currency markets during periods following the European debt crisis in 2016, the Covid-19 pandemic in 2020, and global monetary policy tightening in 2022. These instances notably influenced the probability of distress in both markets, evident in our indicators. Indeed, according to the TimeVAE, our preferred model, the joint probability of distress in the CDS market peaks during the Covid-19 pandemic and remains historically high by the end of our sample period. The record-high JPD in 2020 for CDS markets

aligns with a similarly elevated JPD in FX markets, underscoring the interconnectedness between default and depreciation risk in the global markets.

Section 2 provides an overview of generative AI. In Section 3, we describe our methodology, including the minimax game that forms the basis of GANs and the VAE frameworks, in the context of multivariate time series generation. Section 4 covers our data sources and their description. The main results are presented in Section 5 and Section 6 concludes.

## 2. Generative AI

Generative AI is a set of models used to produce realistic synthetic samples of data given a large set of actual data samples. It was first developed in the field of computer vision, where it started as a way to generate synthetic images by training on a large set of images. These images share general abstract features that can be learned by the models, e.g., they are images of people, cats, dogs, cartoons, handwritten digits etc. The original challenge was to produce synthetic images that could not be distinguished from the real ones. One of the most successful models in this field is Generative Adversarial Networks (GANs), which was introduced in DL by Goodfellow et al. (2014).

GANs consist of two neural networks - a generator and a discriminator. The generator's objective is to create synthetic data samples (e.g., images), while the discriminator aims to differentiate between real and synthetic samples. These objectives are opposing each other since the better the generator's output, the harder it becomes for the discriminator to distinguish it from real data. Conversely, as the discriminator becomes better at identifying synthetic data, it becomes increasingly difficult for the generator to create realistic data that can fool it. Adversarial training, where the two networks' parameters are estimated simultaneously, is used to achieve this goal. GANs have been successful in generating high-quality images (Richardson et al., 2021) and have expanded to generate other types of data, such as text, audio, and video frames (Pascual et al., 2017; Engel et al., 2019; Skorokhodov et al., 2022). GANs have also been extended recently to generate multivariate time series, which is of particular interest to macroeconomists (Yoon et al., 2019).

Another type of generative AI model that competes with GANs in unsupervised learning is Variational Autoencoders (VAEs), as proposed by Kingma and Welling (2019). While the original autoencoders were a generalization of Principal Component Analysis (PCA) to a nonlinear setting (Hinton and Salakhutdinov, 2006), they lacked the ability to generate random synthetic samples. In

contrast, VAEs can generate synthetic samples that are as realistic as those produced by GANs. In finance, autoencoders have been applied to asset pricing, specifically to estimate factors underlying the cross-section of returns (e.g., Gu et al., 2020). This is an area where PCA has traditionally been used for a long time (Connor and Korajczyk, 1988). With VAEs, the incorporation of a sampling distribution enables the generation of synthetic data samples.

In macroeconomics and finance, understanding the cross-sectional dependence of a large number of variables and their joint dynamics over time is crucial. This is particularly important in forecasting applications where dynamic factor models (DFMs) have been used with relative success in recent times (see Bai and Ng (2008) and Bai and Wang (2015) for reviews). DFMs explicitly consider both dimensions and can be easily applied to a large number of series in a panel-like structure. More recently, researchers have expanded DFMs to generate realistic counterfactual scenarios of the series being studied, which have not yet occurred (Cahan et al., 2023; Xiong and Pelger, 2023). One way to interpret the factors generated by DFMs is by associating them with the series within the sample set that were used in the DFM's estimation. This can be done by conducting pair-regressions between each factor and each series and analyzing the factor loads (McCracken and Ng, 2016; 2021).

While DFMs have attractive features such as simplicity and well-studied properties, they are linear models and thus cannot capture the nonlinear dynamics in the system—an area where artificial intelligence, particularly machine learning (ML), excels.

Indeed, ML has become part of the toolbox for empirical research among economists and policy makers. It has been applied for various purposes such as forecasting (e.g., Medeiros et al., 2021; Easley et al., 2021; Bianchi et al., 2021; Bianchi et al., 2022; Lee et al., 2022; Dong et al., 2022; Klein et al.,2023), nowcasting (e.g., Babii, et al., 2021), classification and scoring (e.g., Chen et al., 2019; Bandiera et al., 2020; Burke et al., 2022), estimation of latent factors of asset prices (e.g., Kozak et al., 2020; Gu et al., 2021; Leippold et al., 2022), and evaluation of causal effects for policy purposes (e.g., Deryugina et al., 2019; Chiang et al., 2021; Chernozhukov, et al., 2022a,b) including causal mediation (e.g., Farbmacher et al., 2022a). Furthermore, text analysis of sentiments and intentions (e.g., Angelico et al., 2022; Farbmacher et al., 2022b) and, even computer vision for market assessment (e.g., Bose et al., 2022; Obaid and Pukthuanthong, 2022) have been explored[2].

[2] See Athey and Imbens (2019) and Athey (2019) for a review of the earlier applications.

However, generative AI as implemented here, an active area of Deep Learning (DL), has been largely overlooked in the field, despite its direct and multiple applications for informing macro-prudential policies and for systemic risk monitoring by central banks and market regulators[3].

## 3. Methodology

Our methodology involves a dynamic factor model to capture the conditional mean of the series, generative AI to create synthetic market scenarios using the residuals of the dynamic factor model and, lastly, systemic risk statistics to condense all the generated information into an easily interpretable indicator.

### *3.1. Factor Models*

Let $N$ be the number of cross-sectional units and $T$ be the number of time series observations. For $i = 1, \dots, N$ and $t = 1, \dots, T$, a dynamic factor model can be defined as:

$$y_{it} = \lambda_i(L) f_t + x_{it}, \tag{1}$$

where $\lambda_i(L) = (1 - \lambda_{i1}L -, \dots, -\lambda_{is}L^S)$ is a vector of dynamic factor loadings of order $s$. When $s$ is finite, we refer to it as a DFM. In contrast, a Generalized-DFM allows $s$ to be infinite. Stock and Watson (2011, 2012) provide examples of the former and Forni and Reichlin (1998) and Forni et al. (2000, 2005) introduce the latter. In both cases, the (dynamic) factors $f_t$ evolve according to:

$$f_t = C(L)\varepsilon_t, \tag{2}$$

where $\varepsilon_t$ are $iid$ errors. The dimension of $f_t$, denoted $q$, is the same as that of $\varepsilon_t$ and it refers to the number of dynamic or primitive factors (Bai and Ng, 2007).

The model stated in (2) can be rewritten in static form, simply by redefining the vector of factors to contain the dynamic factors and their lags, and the matrix of loads accordingly, as:

$$\underset{(N \times T)}{Y} = \underset{(N \times r)(r \times T)}{\Lambda F} + \underset{(N \times T)}{X}, \tag{3}$$

where $Y = (Y_1, \dots, Y_N)$ and $F = (F_1, \dots, F_T)$. Clearly, $F$ and $\Lambda$ are not separately identifiable. For any arbitrary $(r \times r)$ invertible matrix $H$, $F\Lambda' = FHH^{-1}\Lambda' = F^*\Lambda'^*$, where $F^* = FH$ and $\Lambda^* =$

[3] Chen et al. (2023) recently used a generative adversarial network to identify portfolio states with unexplained pricing information in an asset pricing application unrelated to ours.

$\Lambda H^{-1}$, the factor model is observationally equivalent to $Y = F^*\Lambda'^* + X$. Therefore $r^2$ restrictions are required to uniquely fix $F$ and $\Lambda$ (Bai and Wang, 2015). Note that the estimation of the factors by principal components or singular value decomposition imposes the normalization that $\frac{\Lambda'\Lambda}{N} = I_r$ and $F'F$ is diagonal, which is sufficient to guarantee identification (up to a column sign change).

The model above can be extended as to incorporate lagged values of the predicted variable in the following way:

$$y_{it} = \beta_0 + \beta_i(L)y_{it} + \lambda_i(L)f_t + x_{it} \quad , \tag{4}$$

where the lag-polynomial in front of $y_{it}$ and $f_t$ do not necessarily have the same order.

### *3.2. Synthetic data generation*

We generate $D = 10{,}000$ simulations using TimeVAEs (Desai et al., 2021), TimeGANs (Yoon et al., 2019), and TTS GANs (Li et al., 2022) based on the residual series from equation 4. Then, we add the synthetic residuals to the fitted values of the model, resulting in $10{,}000$ simulations for the $N$ series at each of the $T$ periods. To evaluate the accuracy of the synthetic residuals generated by these models, we use visual metrics such as t-SNE (Maaten and Hinton, 2008) and PCA, as well as quantitative metrics such as the discriminative and predictive score calculated by a 2-layer Long-Short-Term-Memory (LSTM) post-hoc time-series classification model (Yoon et al., 2019). Additionally, we visually compare the distribution of the actual residuals to that of the generated residuals. Our goal is to determine how closely the distribution of the generated residuals resembles the sample distribution of the actual residuals.

### *3.3. Generative models*

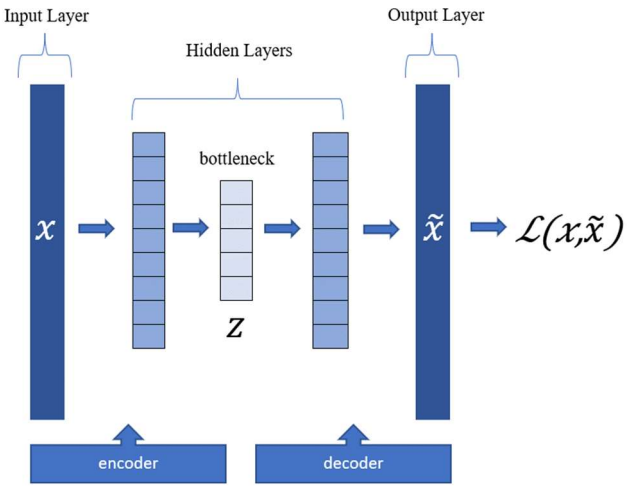
#### *i. Time Variational Autoencoders*

Starting from a raw (unlabelled) dataset $x$ consisting of FX or spread series, the goal is to estimate $\tilde{x}$, which is the reconstruction of the units in the original dataset. The reconstruction can be done by minimizing the reconstruction error $\mathcal{L}(x, \hat{x})$, which measures the difference between the original data and the reconstructed units. The hidden layers in the middle of Figure 1 (which can be a single layer or a group of layers, three in this case) are fundamental for knowledge representation because they force the system to discard non-essential information and represent the data in a lower-

dimensional latent vector. This latent vector, commonly referred to as the bottleneck and represented by $z$ in the figure, captures the essential structure of the data.

A good AEN must balance sensitivity and insensitivity, which depends on the choice of the dimensionality of latent space. In other words, it should accurately describe the inputs to reconstruct the original data, but also it must be insensitive enough to filter out noise and avoid overfitting to the training data.

**Figure 1. Generic Autoencoder, AEN**

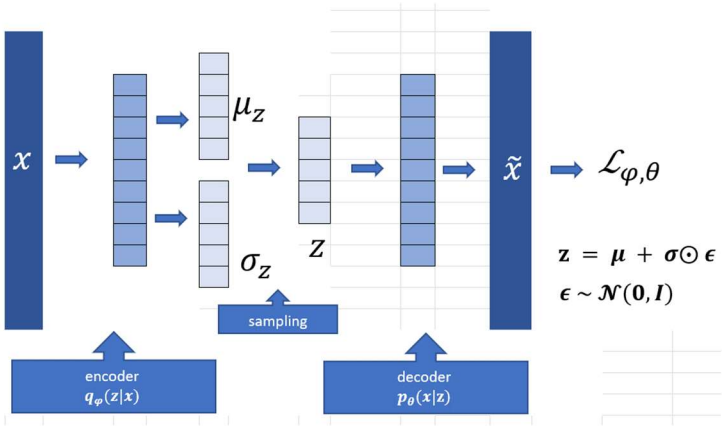



Source: Own elaboration.

VAEs are a generalization of AENs (Kingma and Welling, 2019). They provide a *probabilistic* framework to describe observations in a latent space. Thus, instead of estimating an encoder that generates a unique value that describes each latent attribute (as shown in Figure 1), the VAE model consists of an encoder that describes a probability distribution for each latent attribute. Figure 2 presents a generic VAE. It can be observed that instead of directly producing values for the latent state as it would be the case in a standard AEN, the encoder model of a VAE, generates parameters that describe a distribution for each dimension in the latent space (in our example in Figure 2, it is a Normal distribution with a vector of means ($\mu$) and a vector of variances ($\sigma$)). Thus, the encoder network generates a hidden vector of $z$ by sampling from these predefined

distributions and only then the decoder proceeds to reconstruct the original input units, essentially enforcing a continuous, smooth latent space representation of the original data.

**Figure 2. Generic Variational Autoencoder, VAE**




Source: Own elaboration.

TimeVAE, as implemented by Desai et al. (2021) combines a traditional VAE with some interpretable temporal series components, namely, level, trend and seasonality. The objective function of a VAE is derived from the following minimization problem where the aim is to estimate a posterior distribution $q(z|x)$ which is close to the true posterior distribution $p(z|x)$) by finding:

$$argmin_q \, D_{KL}(q_\varphi(z|x)||p_\theta(z|x)) = E_{q_\varphi(z|x)}[\ln(q_\varphi(z|x)/p_\theta(z|x)]. \quad (5)$$

Here $x$ represents the data we are trying to reconstruct by learning its underlying latent representation $z$. $\varphi$ and $\theta$ the parameters of the posterior distribution, which we omit for now from the equation, alongside the subscripts. Since the true posterior distribution is usually intractable as the marginal likelihood of $x$ is intractable, Kingma and Welling (2013) suggests rearranging equation 5 in the following way:

$$\ln p(x) = D_{KL}(q(z|x)|p(z|x)) - E_{q(z|x)}[\ln q(z|x)] + E_{q(z|x)}[\ln p(z,x)], \quad (6)$$

since KL divergence in equation 6 is always non-negative and the term on the Left-Hand-Side (LHS) is at most 0, the following holds true:

$$\ln p(x) \geq -E_{q(z|x)}[\ln q(z|x)] + E_{q(z|x)}[\ln p(z,x)]\,, \tag{7}$$

Right-Hand-Side, RHS, of equation 7 defines the Evidence Lower Bound (ELBO) or $L_{\theta,\varphi}$. By maximizing ELBO, we can minimize the KL divergence in equation 5. ELBO could be expressed in the following way, where we have changed the signs to present it as a minimization problem:

$$L_{\theta,\varphi} = -E_{q_\varphi(\mathrm{z}|\mathrm{x})}[\ln p_\theta(\mathrm{x}|\mathrm{z})] + D_L(q_\varphi(\mathrm{z}|\mathrm{x})||p_\theta(\mathrm{z}))\,. \tag{8}$$

Here, the first term on the RHS represents the negative log-likelihood of the reconstructed data given samples from the latent space *z* or the reconstruction loss, while the second term on the RHS represents the KL divergence or a regularizer term. $q_\varphi(\mathrm{z}|\mathrm{x})$ represents an approximate posterior distribution - the probabilistic encoder in VAE, where φ are trained parameters – weights and biases, $p_\theta(\mathrm{x}|\mathrm{z})$ is the probabilistic decoder of VAE whose trained weight and bias parameters are denoted by θ . $p_\theta(\mathrm{z})$ is the prior distribution of the latent space, which is chosen to be Gaussian.

Since ELBO is not differentiable with respect to φ and using Monte Carlo simulations is impractical due to high variance of the gradient of *L*, Kingma and Welling (2013) propose to use a reparameterization trick. This involves redefining $q_\varphi(\mathrm{z}|\mathrm{x})$ as a multivariate Gaussian with a diagonal covariance structure, where the mean and the variance are the outputs of an encoder network and function of *x* and φ, as follows:

$$\begin{gathered} \ln q(z|x) = ln\mathcal{N}(z,\mu,\sigma^2\boldsymbol{I}), \\ \text{where } \mathrm{z} = \mu + \sigma\odot\epsilon\,, \\ \epsilon \sim \mathcal{N}(0,\boldsymbol{I})\,, \end{gathered} \tag{9}$$

This allows the loss function to be differentiable with respect to φ , with the following representation:

$$L_{\theta,\varphi} = -\frac{1}{2}\,(1+\ln(\sigma^2)-\mu^2-\sigma^2) + E_{q_\varphi(\mathrm{z}|\mathrm{x})}[\ln p_\theta(\mathrm{x}|\mathrm{z})]. \tag{10}$$

Depending on the type of the data, one can use any differentiable equation to calculate the reconstruction loss -the second term on the RHS of equation 10. We are using mean squared error since the task of data generation involves continuous time series data.

The parameters of the encoder are trained using a 3-layer one-dimensional convolutional network and a dense layer with Rectified-Linear-Unit (RELU) activation function. Time-VAE, as

implemented by Desai et al. (2021) combines a traditional VAE with some interpretable temporal series components, namely, level, trend and seasonality by adding parallel blocks to the decoder network. This is particularly useful in time series applications, and it is what we refer to as TimeVAEs. Their model does it by training a different set of $\theta$s for each of these blocks, simultaneously, and summing up the values obtained at each block, to get reconstructed series $x$. Particularly, the level block measures the average of the series, while the trend block calculates the trend as a polynomial function up to degree $p$ which is defined as a hyper-parameter. In addition to these interpretable layers, TimeVAE could be used to generate a residual component which learns the representation of the part of the data unexplained by the interpretable layers. Two dense layers at each of the interpretable decoder blocks are used to train the parameters of the decoder. The residual block uses a 3-layer one-dimensional convolutional network and a dense layer with RELU activation function to deconvolve and reconstruct the decoder outputs.

In our empirical results, we only activate trend and residual blocks in addition to the default level block, as we already have modeled the conditional mean of the series using the DFM. The final reconstructed series have the following form:

$$\overbrace{x_{recons}}^{T\times S\times N} = \overbrace{\theta_{level}}^{T\times 1\times N} \otimes \overbrace{\boldsymbol{J}}^{1\times S\times 1} + \underbrace{\left(\overbrace{\theta_{trend}}^{T\times N\times P} \overbrace{R}^{P\times S}\right)'}_{T\times S\times D} + \overbrace{residual}^{T\times S\times N}. \tag{11}$$

For $T$ samples with sequence length $S$ and number of features $N$, $\theta_{level}$ represents parameters trained by the neural network for the level block. The tensor product of $\theta_{level}$ with a matrix of ones $\boldsymbol{J}$ is the output of the level block. The second term on the RHS represents the output of the trend block, which consists of the transpose of the dot product of the trend parameters $\theta_{trend}$ with the matrix R. The rows of $R$ include linear time vectors of $r$ raised to power of 1 to $P$, where each of $r$ has the form $[0, 1, 2 \ldots S - 1]/S$.

The data are scaled to the interval 0 to 1 using min-max scaler as in equation 12, as the model uses a RELU activation in its convolutional networks, which assumes any input value less than or equal to zero is zero. A small value is added to the denominator to avoid division by zero:

$$x_{scaled} = \frac{x - min(x)}{max(x) - min(x) + 0.000001}, \tag{12}$$

*ii. Alternative models for data generation - GANs*

Although our main model is the TimeVAEs, we also implemented GANs for comparison purposes. GANs consist of a generator and a discriminator. The generator generates fake data samples that resemble real data, while the discriminator tries to differentiate between real and fake samples. During training, both the generator and the discriminator are optimized in an adversarial manner. The generator produces samples that the discriminator cannot distinguish from real data, and the discriminator tries to become better at distinguishing fake from real samples. GANs can also be structured using the same principles of cross encoder-decoder architectures explained above. We use TimeGAN by Yoon et al. (2019), a recent adaptation of GANs, in this section.

TimeGAN is designed to generate sequences of data. It extends the standard GAN framework to deal with sequences of data by incorporating the time dimension. The basic architecture of a TimeGAN consists of two components: a generator network and a discriminator network. The generator network takes a random noise vector, $z$, as input and produces a generated sequence of data, $G(z)$. The discriminator network takes either a real sequence of data, $x$, or a generated sequence of data, $G(z)$, as input and outputs a scalar representing the probability that the input sequence is real.

The objective of the TimeGAN is to find the optimal generator network that can produce sequences of data that are similar to the real data, and to find the optimal discriminator network that can distinguish between real and generated sequences. This can be formulated as a minimax game between the generator and the discriminator:

$$min_G \, max_D \, V(D,G) \; = \; E_x[log \, D(x)] \; + \; E_z[log(1 \; - \; D(G(z)))] \, , \tag{13}$$

where $V(D,G)$ is the value function, $D(x)$ is the output of the discriminator for a real sequence of data, $x$, and $D(G(z))$ is the output of the discriminator for a generated sequence of data, $G(z)$. In a TimeGAN, the discriminator network is designed to incorporate information over the time by using recurrent neural networks (RNNs), such as LSTM networks, which can process sequences of data. This allows the discriminator to take into account the temporal relationships between the elements in the sequence, and to make a decision based on the entire sequence, rather than just individual cross-sections.

TimeGAN consists of four components: an embedding function, a recovery function, a sequence generator, and a sequence discriminator. The auto-encoding components (embedding and recovery) are trained together with the adversarial components (generator and discriminator), so that

TimeGAN learns features encoding, representations generation, and simultaneously, iterates in time. The embedding network creates the latent space where the adversarial network works. A supervised loss synchronizes the latent dynamics of synthetic and real data (Yoon et al., 2019).

### *iii.* *Time* GANs with Attention

Transformer-based Time-Series Generative Adversarial Network by Li et al. (2022) is the first GANs model replacing the RNN network with a transformer network to be applied in time series and sequential data. The authors combine the ideas used for image and Natural Language Processing (NLP) for time series data by redefining width of an image as the number of sequences in a time series sample and by defining channels, which are normally used for RGB color scheme, as features of time series data.

## *3.4. Systemic risk indicators*

After adding the $D$ synthetic error series generated by the DL models described above to the output of the model in equation 4, we end up with $N$ matrices of $D \times T$ for each market, where $D = 10{,}000$ is the number of simulations. The simulated series are used to construct the systemic risk statistics adapted from the work of Oh and Patton (2018). First, we identify that a given instrument signals distress if the simulated series at day $t$ is higher than a selected critical value in the following way:

$$S_{i,t+250} \equiv \mathbf{1}\{x_{i,t+250} > c^{*}_{i,t+250}\}. \tag{14}$$

For every original CDS or currency return, the first year (250 days) is taken to define the 95$^{th}$ percentile. This is the level of distress above which the return is in the most extreme 5% on the right tail. $c^{*}$ is dynamic and calculated using a moving window of one day. Then, for every series of returns a threshold series as in equation 14 is constructed. Once we calculate the distress signal for the selected threshold, we end up with a matrix of zeros and ones for each instrument, where the value 1 for date $t$ and simulation $j$ indicates the simulated series is above the threshold or, in other words, is signaling distress, while 0 indicates no distress.

*i. Joint probability of distress*

We use the probability that a large proportion of markets are in distress as a measure of systemic risk. Following Oh and Patton (2018) we define the joint probability of distress (JPD) as follows:

$$\mathrm{JPD}_{t,k} \equiv \mathrm{Pr}_t \left[\left(\frac{1}{N}\sum_{i=1}^{N} S_{i,t+250}\right) \geq \frac{c}{N}\right], \tag{15}$$

where $c$ is an arbitrary threshold of a large proportion of the series. To calculate the index, we add all ones and zeros in the matrices, obtaining a matrix where every element is the number of markets in distress on simulation $i$ on day $j$. After defining a critical value of distress, we convert the values in the matrix to ones and zeros if the number of instruments in distress is greater or equal than the threshold. Then, by column, we sum the number of ones and divide by $D$, the number of simulations. We end up with an estimate of the probability of distress given $c$, i.e., the probability that at least $c\ of\ N$ instruments in the market are in distress.

*ii. Expected proportion in distress*

$$\mathrm{EPD}_{i,t} \equiv \mathrm{E}_t \left[\frac{1}{N}\sum_{j=1}^{N} S_{j,t+250} \mid S_{i,t+250} = 1\right]. \tag{16}$$

Equation 16 shows the expected proportion in distress (EPD). EPD is the proportion of the markets in distress given that one of the markets is in distress. We estimate this proportion not only for the markets of sovereign debt and currencies but also between sub-groups, e.g., emerging market versus advanced markets, FX versus CDS markets, etc. For each instrument *i* and for each time *t,* we calculate the proportion of market in distress at time $t + 250$ for the simulation, given the instrument *i* signals distress at time $t + 250$. In continuation, we calculate the average of these proportions by dividing them by the number of cases where the instrument $i$ signals distress at time $t + 250$. The EPD values lay between 1/N and 1, since the instrument $i$ is included in the sum. EPD may take a value of 0 for the selected instrument for date $t$ only if that instrument doesn't signal distress in any simulation for that date. Once we get the EPD for each instrument, we plot the median, 20 and 80 percentiles for each day for each market. In section A.1 of the Appendix we describe some additional practical considerations useful for replication of our results.

The EPD allows us to identify which instruments and markets are the most systemic, that is, which house higher spillovers to the rest of the markets. The dynamic nature of the model allows us to observe how the systemic relationship among the participants of the market changes over time by comparing EPD estimates at different time periods. We put a special emphasis on crisis periods and demonstrate snapshots of three crisis periods with highest EPDs. Thus, we augment the findings of Greenwood-Nimmo et al. (2023), not only by considering FX markets in addition to sovereign CDS instruments but also, because these authors primary focus is on bivariate relationships, while our EPD values capture the influence of an instrument on the entire market.

We may also draw parallels for the estimates of EPD with centrality estimates of Le et al. (2022), particularly outdegree centrality ($C_{outdegree}$) which measures the ratio of the number of countries affected by each market in the sample to the total number of countries. Unlike these authors' approach based on Granger causality and VARs, our approach allows us to fully consider non-linearities in the propagation of shocks.

## 4. Data

The frequency of our data is daily, spanning the period January 1, 2011 – October 27, 2022, for a total of 3,083 business days. We work with two datasets: sovereign CDS spreads and foreign exchange rates.

Sovereign CDS are financial contracts where a protection seller takes regular premium fees from an investor and compensates them for the face value of a reference obligation in case the sovereign entity defaults its debt obligations (Packer and Suthiphongchai, 2003). The spread is measured as the percentage of the obligation, provided in basis points (bp), and represents the difference between yield and risk-free rate, i.e. the US Treasury rate in the sovereign debt market. We have retrieved the CDS spread data from Refinitiv for 35 countries spanning 5-year CDS contracts based on senior unsecured sovereign debt instruments. The maturity is selected as 5 years as the CDS contracts with this maturity are considered the most liquid for many countries (Mihai and Neagu, 2011; Oh and Patton, 2018). Exchange rate data are all in local currency per the US dollar such that an increase in the level refers to depreciation of the local currency. FX data are retrieved from Refinitiv as well.

FX and CDS data each include instruments for 35 economic entities including Euro for Eurozone as a single entity. The list of entities for the FX market was constructed based on the flexibility of a

country's exchange rate regime. We follow Ilzetzki et al (2019) to select the currencies with flexible regimes at least at the level that can be classified as "pre-announced crawling band that is wider than or equal to +/-2%". The list of countries for CDS spreads has been made based on data availability, i.e. we have interpolated a few missing points as long as the sequence of missing points was not more than two months and linear interpolation seemed as a reasonable choice given the pattern of the series before and after the missing data points. There is no missing data for the FX market.

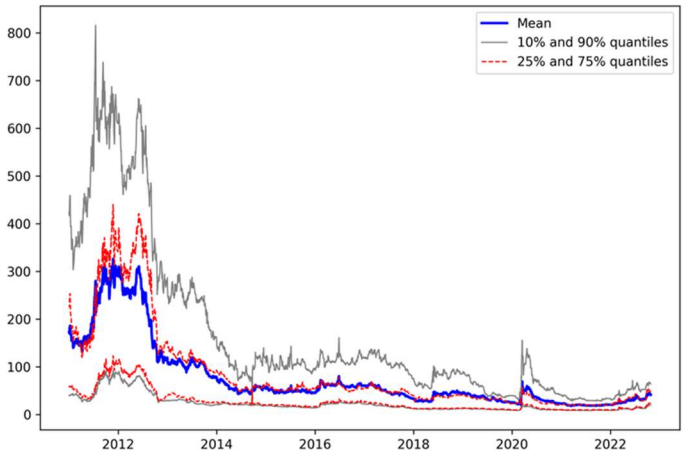
Finally, for the estimation of our systemic risk statistics we use the IMF classification of countries in three categories: advanced economies (AE), emerging markets (EM) and low-income developing countries (LDC) (see People and Back, 2022). The CDS market data include 15 advanced economies, 18 emerging market economies and 2 low-income developing countries. The FX market data include 12 advanced economies (including the Eurozone as a single entity), 16 emerging market economies and 7 low-income developing countries. See Table A1 in the appendix for a list of all CDS instruments and currencies used.

Figure 3a shows that at the beginning of the sample, the CDS spreads for advanced economies were hovering at around 300 bp, on average. This is consistent with the Euro area crisis gaining momentum, in the aftermaths of the global financial crisis, as discussed for instance by Heinz and Sun (2014). Later, we observe that the average CDS spread has decreased towards 50 bp and spiked to 100 bp during COVID-19 pandemic and again in 2022 with the increased risk perceptions following the war in Ukraine and expectations towards an upcoming global recession.

In Figure 3b, CDS data are shown for emerging markets. Both levels and dynamics are greatly dissimilar in the two cases, as expected. In particular, spreads in EM are around 150 bp, on average, except during high volatile periods, for instance some months of 2012, 2016, 2020 and the end of 2022, when the spread averaged more than 200 bp, and the $90^{th}$ percentiles reached 400 bp. Spikes in the spread of EM are much more frequently observed than in AE, and are linked to turmoil in traditional financial markets and in commodity markets, which EM economies often export.

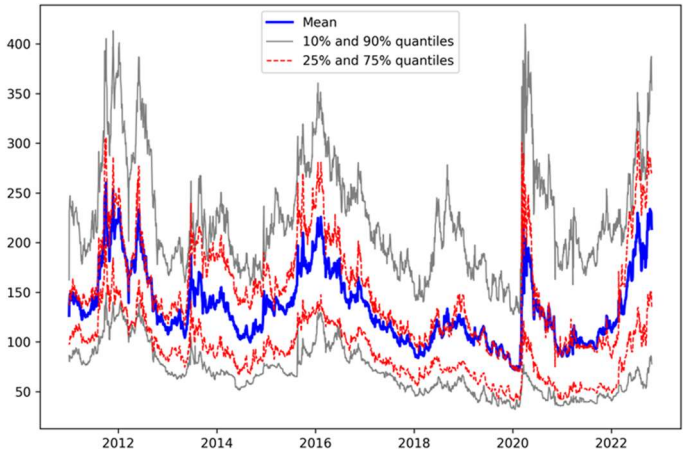
In Figures 3c and 3d we show the series of exchange rates. These are converted to natural logarithm and standardized to 100 at the initial point of the sample, which eases comparisons. All exchange rates are in local currency (LCY) per the US dollar (USD) such that an increase in value implies depreciation. We observe that the exchange rates of almost all emerging market economies with flexible regimes have their currencies depreciating faster during 2020 and 2022 as well as during the commodity price crisis of 2016 compared with other periods.

**Figure 3a: The CDS spreads, Advanced Economies (basis points)**




Note: Mean, 10%, 25%, 75%, and 90% quantiles of the CDS spreads for advanced economies: January 2011- October 2022.

**Figure 3b: The CDS spreads, Emerging Markets (basis points)**

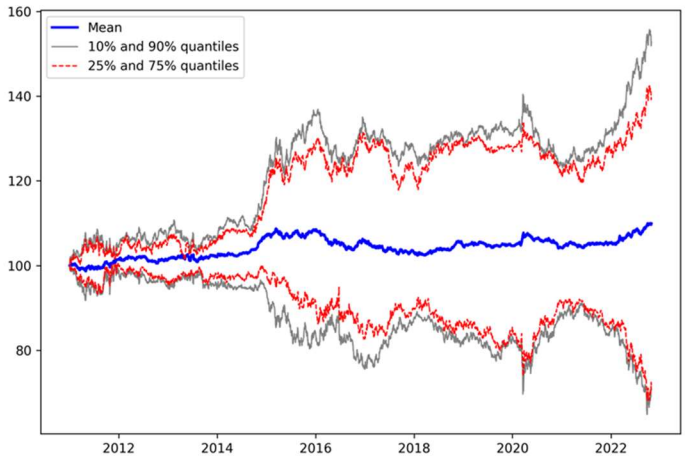



Note: Mean, 10%, 25%, 75%, and 90% quantiles of the CDS spreads for emerging markets: January 2011- October 2022

Although advanced economies also depreciate on average against the US dollar during the crisis periods, the patterns in AE and EM are contrasting. Interestingly, while in AE the $10^{th}$ and $90^{th}$ percentiles move in opposite directions, this is not the case for EM. This is consistent with the fact that depreciation is more of a risk concern for emerging and low-income developing countries than it is for advanced economies.

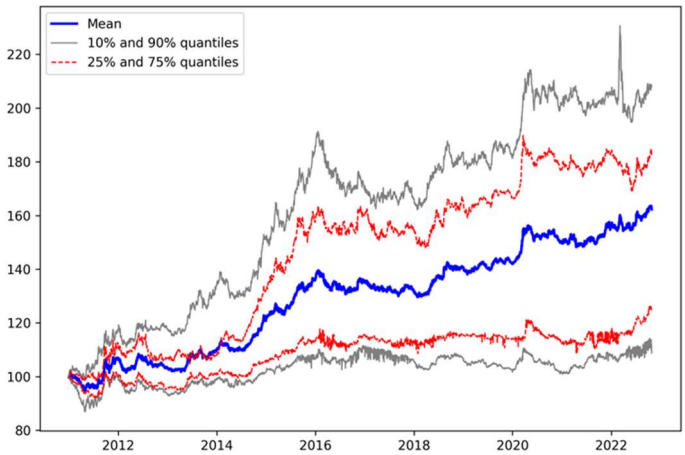
In our estimations we use log-differences of the series to remove the unit roots present in the data, as it is standard practice in the factor models literature. Indeed, augmented-Dickey-Fuller tests applied to the original series suggest the presence of a unit root in 31 out of 35 series for CDS market data and all 35 series in FX, at 5% significance level.

**Figure 3c: The exchange rates, AE (LCY per USD, 100*log)**




Note: Mean, 10, 25, 75, and 90 percentiles of advanced economies exchange rates against the US Dollar from January 2011 (=100) to October 2022.

**Figure 3d: The exchange rates, EM (LCY per USD, 100*log)**




Note: Mean and 10, 25, 75, and 90 percentiles across emerging markets exchange rates from January 2011 (=100) to October 2022.

The summary statistics of the log-differences series are presented in Tables 1a and 1b. CDS spreads seem to be more volatile and to be more autocorrelated than FX returns series. Furthermore, CDS series house heavier tails than FX series and are right-skewed, while the distribution of FX series is more symmetric and leptokurtic.

**Table 1a. Summary statistics for log-differences of daily CDS spreads**

| | **Mean** | **5%** | **25%** | **Median** | **75%** | **95%** |
|---|---|---|---|---|---|---|
| *Mean* | -1.024 | -295.318 | -96.746 | 0.755 | 93.941 | 287.173 |
| *Std dev* | 172.333 | 263.992 | 170.972 | 147.555 | 193.874 | 308.514 |
| *1st-order autocorrel* | 0.267 | 0.334 | 0.251 | 0.217 | 0.327 | 0.397 |
| *Skewness* | 1.810 | -2.444 | -1.162 | 2.752 | 3.968 | 2.928 |
| *Kurtosis* | 19.113 | 14.766 | 19.827 | 32.438 | 30.53 | 15.912 |
| *5%* | -220.238 | -748.035 | -380.178 | -184.276 | -33.365 | 2.151 |
| *25%* | -74.309 | -397.912 | -160.429 | -25.585 | 0 | 86.795 |
| *Median* | -11.030 | -248.071 | -58.649 | 0.000 | 25.085 | 206.551 |
| *75%* | 53.786 | -125.401 | -1.921 | 7.670 | 126.726 | 389.759 |
| *95%* | 248.823 | -3.178 | 41.147 | 198.820 | 429.847 | 819.885 |

Note: Summary statistics of the log-diffs of daily CDS spreads, in basis points. Columns 2 to 7 represent the mean and quantiles of the cross-sectional distribution of the measures in the first row.

**Table 1b. Summary statistics for log-differences of daily exchange rates**

| | **Mean** | **5%** | **25%** | **Median** | **75%** | **95%** |
|---|---|---|---|---|---|---|
| *Mean* | 1.549 | -94.932 | -26.108 | 1.224 | 29.174 | 99.214 |
| *Std dev* | 22.090 | 49.714 | 23.094 | 18.116 | 25.612 | 54.552 |
| *1st-order autocor* | 0.049 | 0.258 | 0.118 | 0.081 | 0.189 | 0.234 |
| *Skewness* | 0.275 | -1.573 | -1.724 | 0.330 | 2.017 | 1.753 |
| *Kurtosis* | 1.615 | 5.377 | 5.527 | 4.392 | 7.252 | 5.697 |
| *5%* | -32.743 | -188.843 | -71.645 | -26.830 | 0.573 | 36.679 |
| *25%* | -12.099 | -118.255 | -36.510 | -7.036 | 12.056 | 62.410 |
| *Median* | 0.664 | -84.415 | -20.898 | 0.000 | 23.045 | 86.147 |
| *75%* | 14.148 | -60.134 | -10.115 | 8.994 | 39.087 | 123.977 |
| *95%* | 37.704 | -35.680 | 0.000 | 30.847 | 77.112 | 198.956 |

Note: Summary statistics of the log-diffs of daily exchange rates measured, in basis points. Columns 2 to 7 represent the mean and quantiles of the cross-sectional distribution of the measures in the first row.

## 5. Results

Our results consist of three parts. In the first part, we describe our linear factor model estimates, in the second part we compare the accuracy of the simulation by the three generative AI models we employ and, in the third part, we show our systemic risk statistics and interpret them.

### *5.1. Linear Factor Model*

We estimate three models as described in equation 4, using 5 lags for the outcome variable, for CDS, for FX and for both datasets. All the series were converted to log-diffs and were standardized to zero mean and unit variance before estimation. Latent factors were retrieved through PCA.

We use the $IC_p$ criteria of Bai and Ng (2002) to decide the number of static factors to include in the model. The maximum number of factors was set as Schwert (1989), who suggested using $8 \times round\left\{\left[\frac{min(T,N)}{100}\right]^{1/4}\right\}$. Information criteria suggest using 4 factors to model the conditional mean of the CDS, FX and the two markets combined, which explain around 61.99%, 43.07% and 45.14% of the total variation, respectively.

In Tables 2a-2c we present the main results of our linear factor model. We have regressed each original series on each of the four estimated factors, and report the marginal R-squared, $mR2(\#\text{factor})$, of the regressions in the second row of each table. Then we report the largest factor loads on each factor, aiming to identify what series are better represented by what factors. As can be observed, the first factors in the CDS market represent medium-size and large emerging markets like Colombia (CO5Y), Paraguay (PA5Y), India (ID5Y) and South Africa (ZA5Y); the second factor represent spread co-movements in advanced economies such as Belgium (BE5Y) or Austria (AT5Y).

Regarding the FX market, the first factor is mainly associated with strong currencies like those of Singapur (SGD), Australian Dollar (AUD or Norway Crown (NOK). When the two markets are combined, the first factor is dominated by CDS series, while only Australian Dollar and the Norwegian Crown make it to the top 10 according to the factor loads.

In Figures 4a-4d we present box-plots of the factor loads in the model of CDS and FX markets, distinguishing between advanced economies and emerging market economies. As can be observed, regarding the CDS market (Figures a and b) there is a larger heterogeneity between emerging markets than between advanced economies. This heterogeneity is even larger regarding FX markets.

**Table 2a. CDS Factors: Total Variation Explained, 0.6199**

| | Factor 1 | | Factor 2 | | Factor 3 | | Factor 4 | |
|---|---|---|---|---|---|---|---|---|
| | mR2(1) | 0.37 | mR2(2) | 0.11 | mR2(3) | 0.09 | mR2(4) | 0.05 |
| Factor Load | CO5Y | 0.55 | BE5Y | 0.26 | MX5Y | 0.28 | PL5Y | 0.32 |
| | PA5Y | 0.54 | AT5Y | 0.23 | PE5Y | 0.26 | HU5Y | 0.27 |
| | ID5Y | 0.53 | IE5Y | 0.22 | BR5Y | 0.26 | RO5Y | 0.27 |
| | ZA5Y | 0.52 | DE5Y | 0.22 | CO5Y | 0.24 | KZ5Y | 0.15 |
| | PE5Y | 0.51 | FR5Y | 0.21 | PA5Y | 0.24 | PT5Y | 0.15 |
| | MX5Y | 0.50 | NL5Y | 0.21 | MY5Y | 0.23 | ES5Y | 0.15 |
| | MY5Y | 0.50 | DK5Y | 0.18 | TH5Y | 0.23 | IT5Y | 0.14 |
| | CL5Y | 0.49 | SE5Y | 0.17 | CL5Y | 0.22 | IE5Y | 0.02 |
| | PH5Y | 0.49 | GB5Y | 0.17 | PH5Y | 0.22 | SE5Y | 0.02 |
| | CN5Y | 0.45 | FI5Y | 0.15 | CN5Y | 0.21 | FR5Y | 0.02 |

**Table 2b. Factor Estimates from FX: Total Variation Explained, 0.4307**

| | mR2(1) | 0.30 | mR2(2) | 0.06 | mR2(3) | 0.04 | mR2(4) | 0.03 |
|---|---|---|---|---|---|---|---|---|
| Factor Load | SGD | 0.77 | EUR | 0.28 | MYR | 0.34 | UGX | 0.24 |
| | BND | 0.72 | ISK | 0.22 | PHP | 0.17 | UYU | 0.12 |
| | AUD | 0.68 | COP | 0.20 | THB | 0.10 | ZMW | 0.09 |
| | NOK | 0.63 | JPY | 0.19 | MXN | 0.10 | JPY | 0.09 |
| | NZD | 0.61 | STN | 0.12 | JPY | 0.07 | MGA | 0.09 |
| | SEK | 0.59 | SEK | 0.12 | BRL | 0.07 | PYG | 0.08 |
| | HUF | 0.57 | MYR | 0.11 | KRW | 0.07 | RUB | 0.07 |
| | EUR | 0.56 | CLP | 0.11 | RUB | 0.05 | BND | 0.04 |
| | CAD | 0.54 | HUF | 0.10 | ZAR | 0.04 | ISK | 0.04 |
| | ZAR | 0.53 | MXN | 0.09 | VUV | 0.04 | SGD | 0.03 |

**Table 2c. Factor Estimates from CDS & FX: Total Variation Explained, 0. 4514**

| | mR2(1) | 0.24 | mR2(2) | 0.10 | mR2(3) | 0.06 | mR2(4) | 0.04 |
|---|---|---|---|---|---|---|---|---|
| Factor Load | CO5Y | 0.58 | SGD | 0.36 | BE5Y | 0.27 | MX5Y | 0.19 |
| | PA5Y | 0.56 | EUR | 0.35 | AT5Y | 0.22 | PH5Y | 0.17 |
| | MX5Y | 0.56 | BND | 0.33 | DE5Y | 0.22 | PE5Y | 0.17 |
| | PE5Y | 0.54 | NZD | 0.29 | IE5Y | 0.22 | CO5Y | 0.17 |
| | ZA5Y | 0.54 | AUD | 0.28 | FR5Y | 0.21 | MY5Y | 0.16 |
| | CL5Y | 0.52 | SEK | 0.27 | NL5Y | 0.21 | BR5Y | 0.16 |
| | BR5Y | 0.50 | HUF | 0.26 | SE5Y | 0.18 | TH5Y | 0.15 |
| | TR5Y | 0.42 | MY5Y | 0.26 | DK5Y | 0.17 | PA5Y | 0.15 |
| | NOK | 0.41 | TH5Y | 0.25 | GB5Y | 0.17 | CL5Y | 0.15 |
| | AUD | 0.39 | PH5Y | 0.23 | FI5Y | 0.15 | CN5Y | 0.15 |

Note: The table report the marginal R-squared, $mR2(\#\text{factor})$ and the highest factor loads on each factor from the second row on. See Table A1 in the appendix for a description of the CDS contracts and currencies in the first column of the tables.

**Figure 4a: The CDS factor loadings, AE**

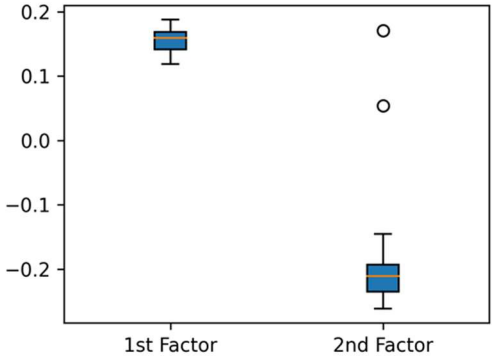



**Figure 4b: The CDS factor loadings, EM**

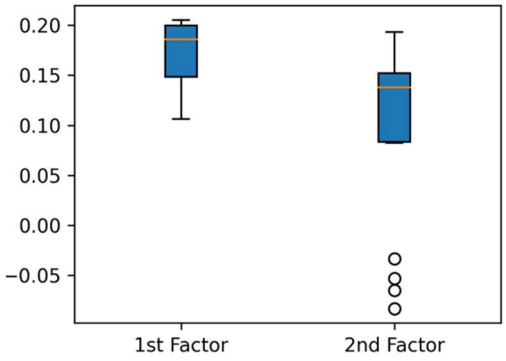



**Figure 4c: The FX factor loadings, AE**

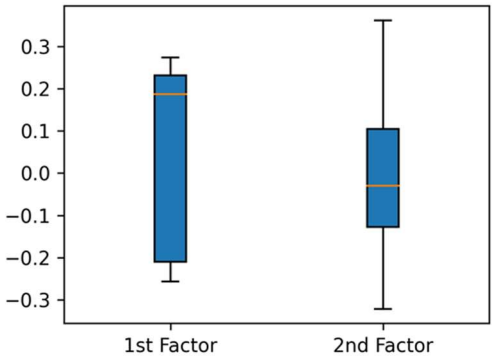
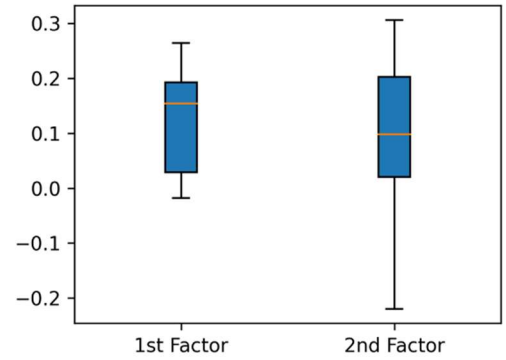



**Figure 4d: The FX factor loadings, EM**




Note: Box plots of estimated factor loadings corresponding to the first and second common factors in both CDS (top) and FX (bottom) markets, and for Advanced (left) and Emerging (right) market economies.

*5.2. Comparison of Generative models to simulate idiosyncratic components*

Table 3 shows the comparative performance of three generative models: TimeVAE, TimeGAN and TTSGAN, for different choices of hyper-parameters. Interestingly, the discriminative score (Table 3a) does not suggest a significant difference between the models, with similar values ranging between 0.4987 and 0.5 for different number of hidden layers (i.e., 50-300 for the TimeVAE and using GRU, LSTM and LSTMLN for the TimeGAN) and different latent space dimensions (i.e. 15 or 30), for the three models. On the other hand, the predictive score (Table 3b) provides evidence in favor of the TimeVAE. In this case, there is a reduction in the score value amounting to about 0.1, when we compare TimeVAE with the second-best model, irrespective of using CDS, FX or the combined markets data.

When we set the latent space dimension to 30, model performance increases in most of the cases, except for TimeGAN and TTSGAN for FX markets. Based on the results shown in Tables 3a and 3b, we set hyper-parameters to maximize the model's performance in all three cases.

Finally, visual metrics (Figure A1 in the Appendix) suggest that TimeGAN and TTSGAN perform similarly, although TimeVAEs is the only one that seems to extract visible patterns from the original data, providing further evidence in favor of TimeVAEs.

**Table 3a. Discriminative Score**

| | CDS | | FX | | CDS & FX | |
|---|---|---|---|---|---|---|
| Model Type | Latent Dimension | | | | | |
| | 15 | 30 | 15 | 30 | 15 | 30 |
| **TimeVAE** | | | | | | |
| *Hidden Layer Sizes* | | | | | | |
| 50,100,200 | 0.4998 | 0.4992 | 0.4997 | 0.4998 | 0.5 | 0.4998 |
| 150,250,300 | 0.4997 | 0.4995 | 0.5 | 0.4997 | 0.5 | 0.5 |
| **TimeGAN** | | | | | | |
| *Hidden Layer Types* | | | | | | |
| gru | 0.4984 | 0.4987 | 0.4993 | 0.4997 | 0.4995 | 0.5 |
| lstm | 0.4995 | 0.498 | 0.4995 | 0.499 | 0.4998 | 0.5 |
| lstmLN | 0.4987 | 0.5 | 0.4996 | 0.4989 | 0.4995 | 0.5 |
| **TTSGAN** | | | | | | |
| | 0.485 | 0.4962 | 0.4845 | 0.4923 | 0.5 | 0.4993 |

Note: Performance of generative models evaluated by a 2-layer LSTM post-hoc time-series classification network. The discriminative score shows classification accuracy minus 0.5 (lower the better).

As the TimeVAE removes noise while generating synthetic data (Desai et al, 2021), it is important to note that the variance of the generated synthetic series is smaller than that of the actual series. This may not be true for TimeGAN or TTSGAN since these models do not include separate blocks in their structure for noise reduction, as in Denoising GAN (Chen et al, 2020), for instance. The variance of synthetic residuals generated by GANs is indeed very close to the variance of actual residuals while the variance of VAE generated residuals is much lower. We expect that the noise constitutes a significant portion of the total variance in the residuals obtained from the regression, thus TimeVAE scores a better predictive accuracy than GAN models in this specific exercise by simply performing better noise reduction.

**Table 3b. Predictive Score**

| | | CDS | | FX | | CDS & FX | |
|---|---|---|---|---|---|---|---|
| Model Type | | Latent Dimension | | | | | |
| | | 15 | 30 | 15 | 30 | 15 | 30 |
| **TimeVAE** | | | | | | | |
| *Hidden Layer Sizes* | | | | | | | |
| | 50,100,200 | 0.4103 | 0.3835 | 0.4704 | 0.4698 | 0.4208 | 0.4011 |
| | 150,250,300 | 0.3941 | 0.3847 | 0.4698 | 0.4825 | 0.4297 | 0.3988 |
| **TimeGAN** | | | | | | | |
| *Hidden Layer Types* | | | | | | | |
| | gru | 0.4927 | 0.4475 | 0.5544 | 0.5808 | 0.4876 | 0.4649 |
| | lstm | 0.5219 | 0.4629 | 0.5173 | 0.5534 | 0.5512 | 0.5079 |
| | lstmLN | 0.4079 | 0.3853 | 0.5366 | 0.5748 | 0.4883 | 0.4386 |
| **TTSGAN** | | | | | | | |
| | | 0.6534 | 0.5954 | 0.7698 | 0.7633 | 0.8699 | 0.6123 |
| **Original** | | | | | | | |
| | | 0.3494 | 0.3494 | 0.3494 | 0.3494 | 0.3494 | 0.3494 |

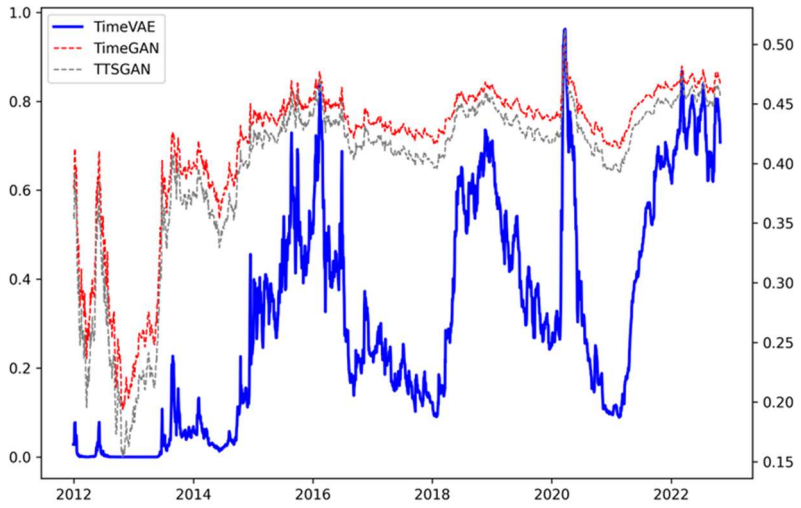
Note: Performance comparison of generative models as evaluated by a 2-layer LSTM post-hoc time-series classification model. The predictive score represents mean absolute error (lower the better). As a benchmark, we present the predictive score of the classification model, where the original data is used to calculate the metrics.

### *5.3. Systemic risk indicators*

Estimated Joint Probabilities of Distress are presented in Figures 5a and 5b for CDS and FX market, respectively. We show the results for all models setting to 50% the threshold that defines a systemic risk situation (i.e. a large proportion of the markets in distress).

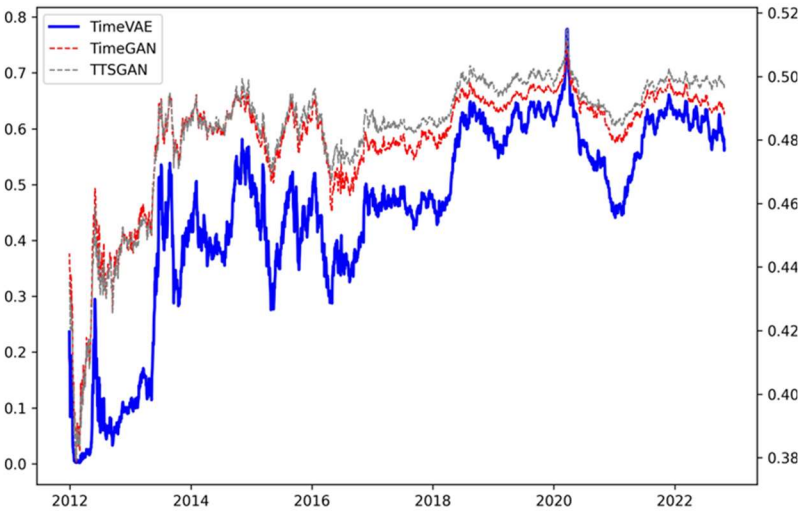
The results with TimeGAN and TTSGAN resemble each other, as expected, since they both belong to the GANs family. For both CDS and FX markets, JPD estimated according to all three models, appears to capture instances of widespread market turmoil. Notably, these instances include the periods following the European debt crisis in 2016, the Covid-19 pandemic in 2020, and the global monetary policy tightening in 2022. These three situations clearly affected the probability of distress in the two markets, and this was reflected by pronounced increases in the JPD indicators.

**Figure 5a: Joint probability of distress, CDS**




Note: JPD of the CDS markets for the threshold=0.5. The simulations combine PCA and generative models: TimeVAE (left axis), TimeGAN and TTSGAN (right axis).

**Figure 5b: Joint probability of distress, FX**



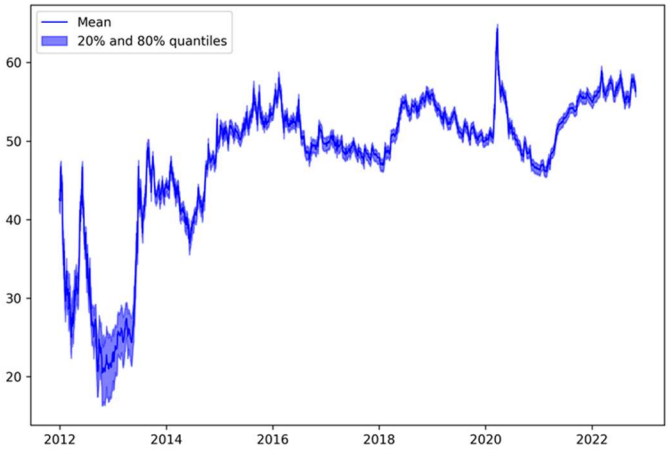

Note: JPD of the FX markets for the threshold=0.5. The simulations combine PCA and generative models: TimeVAE (left axis), TimeGAN and TTSGAN (right axis).

In summary, TimeVAE seems to better capture the overall temporal dynamics of systemic distress in both markets, as it shows clearer and more differentiated dynamics over the time compared to GANs. Indeed, according to the TimeVAE, the joint probability of distress in the market for CDS

reaches the peak during the Covid19 pandemic, and remains at historically high levels by the end of our sample period. The record high JPD in 2020 for the CDS markets coincides with a record high JPD in the FX markets, which emphasizes the relationship between default and depreciation of the Twin markets.

Figure 6a and 6b show the Expected Proportion in Distress given one of the markets is in distress. EPD is presented alongside 20% and 80% quantiles for CDS and FX markets in top and bottom panels of the figure, respectively. For CDS markets, the narrow band around the average value suggests low uncertainty in the estimate of the average value of markets in distress, which suggests a large commonality between the series of CDS spreads.

**Figure 6a: CDS expected proportion in distress, given one economy in distress**

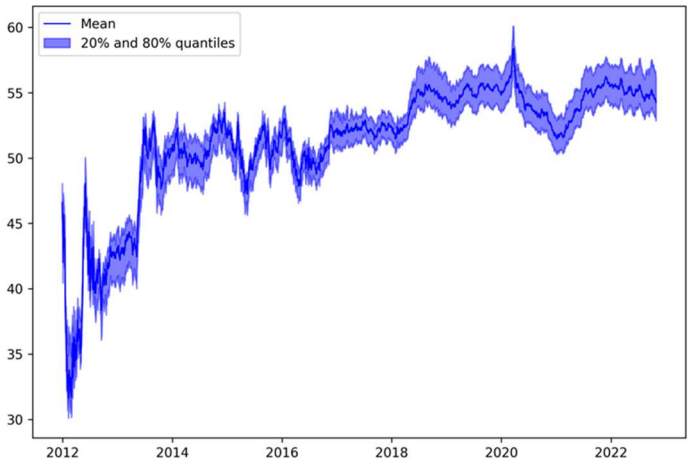



Note: Expected proportion (in percent) of CDS market in distress given one market is in distress, averaged out across markets.

Regarding the FX market on the other hand, a more heterogeneous dependence structure is observed, which could already be foreseen from the raw distribution of exchange rate variations provided in the data section. We can observe for both markets the band became narrower in the wake of the Covid19 crisis in 2020, when all the economies faced lockdowns, and fears for sudden stops and fiscal instabilities plagued debt markets in emerging and low income countries (Kose et al. 2021). A more heterogeneous dependence structure allows us to see more visibly the changes in densities of EPD estimates. The results confirm the pattern we observe at the CDS market as the mean of EPD estimates move up while its density concentrates more around the mean. For global

events which should affect both markets in a similar way such as the economic crisis induced by COVID-19 lockdowns, we see EPD estimates and its density move following the same pattern.

**Figure 6b: FX expected proportion in distress, given one economy in distress**




Note: Expected proportion (in percent) of FX market in distress given one market is in distress, averaged out across markets.

Tables 4a and 4b present individual estimates of systemic risk for the economies with the largest and lowest EPDs, at three selected dates, which in turn, correspond to the highest average EPDs across all the markets over the sample. As for CDS instruments, Turkey, South Africa and Qatar appear at least twice among the most systemic markets for the selected dates. This indicates that one of these economies being in distress signals a high distress risk in the market as a whole. On the other hand, Ireland and Romania are among the least systemic in our sample, as the low EPD for these economies implies the realization of distress in these economies being more of idiosyncratic phenomena than a realization of systemic risk. Nonetheless, the difference between the highest and lowest statistics is relatively small, which once again points out to high commonality between CDS markets, which is in line with the previous literature (Gomez-Gonzalez et al. 2023).

As for FX markets, the difference between the most systemic and the least systemic currencies is higher at around 5 pp, even during crisis periods. The currencies of Japan and Great Britain are among the most systemic ones. SCR and SGD are also on the list since the relevant monetary authorities allow them to move within a certain band against major currencies like USD and EUR,

thus the systemic relation should be attributed more to these major currencies. On the other hand, the currencies of Philippine, Ghana and Korea are among the least systemic.

All in all, our systemic risk statistics provide a precise quantification of the intuitive result that currencies of less developed countries would not necessarily signal distress for the whole global FX market, while distress in strong currencies would indeed signal a distress for the entire FX market, mainly consisting of emerging market economies.

**Table 4a. Estimates of systematic risk, CDS**

| | 11 February 2016 | | 23 March 2020 | | 07 March 2022 | |
|---|---|---|---|---|---|---|
| | EPD | Instrument | EPD | Instrument | EPD | Instrument |
| Most Systemic | 57.33 | ZA5Y | 64.13 | QA5Y | 58.08 | ZA5Y |
| 2 | 57.33 | TR5Y | 63.54 | TR5Y | 57.82 | SA5Y |
| 3 | 57.12 | IT5Y | 63.35 | TH5Y | 57.75 | QA5Y |
| 33 | 55.29 | RO5Y | 61.84 | RO5Y | 56.32 | BR5Y |
| 34 | 55.12 | DE5Y | 61.8 | MY5Y | 56.2 | IE5Y |
| Least Systemic | 55.12 | IE5Y | 61.4 | IE5Y | 55.84 | DE5Y |

Note: The table presents three instruments in the CDS market with the largest and smallest expected proportion in distress (in percent) for three select dates with the highest expected proportion of distress estimates. See Table A1 in the appendix for a description of the full lists of CDS contracts and currencies in the table.

**Table 4b. Estimates of systematic risk, FX**

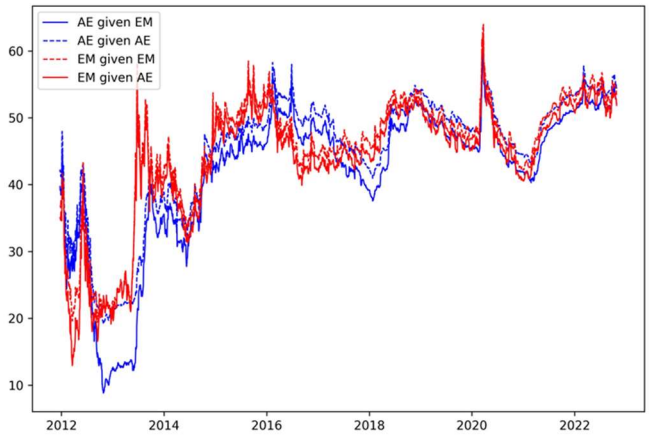
| | 20 January 2016 | | 23 March 2020 | | 26 September 2022 | |
|---|---|---|---|---|---|---|
| | EPD | Instrument | EPD | Instrument | EPD | Instrument |
| Most Systemic | 55.19 | JPY | 60.08 | SCR | 56.69 | SGD |
| 2 | 54.47 | ZAR | 60.08 | GBP | 56.64 | SCR |
| 3 | 54.46 | MGA | 59.79 | AUD | 56.42 | GBP |
| 33 | 51.16 | KRW | 57.23 | KRW | 53.67 | GHS |
| 34 | 51.01 | GHS | 57.15 | GHS | 53.61 | PHP |
| Least Systemic | 50.38 | PHP | 57.13 | PHP | 53.36 | KRW |

Note: This table presents three instruments in the CDS market with the largest and smallest expected proportion in distress (in percent) for three select dates with the highest expected proportion of distress estimates. See Table A1 in the appendix for a description of the full list of CDS contracts and currencies in the table.

Figures 7a and 7b divide the results regarding the expected proportion of markets in distress between advanced economies and emerging markets. In this way, we evidence the existing

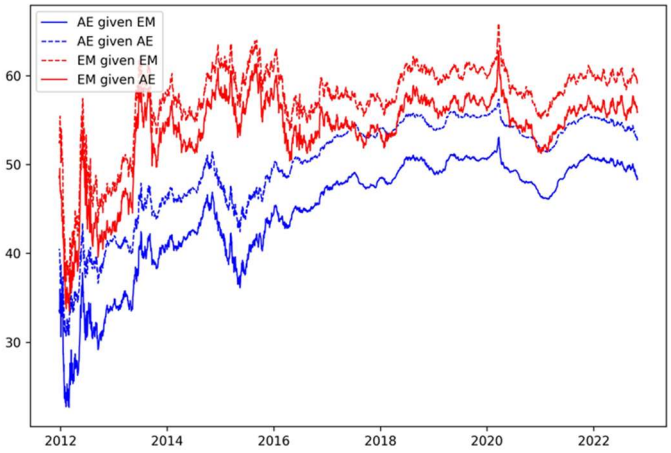
interdependence between the two groups, which help to understand how market spillovers occur. Results indicate that spillover effects within groups are stronger than between groups for both advanced and emerging market economies. This confirms the findings of Le et al. (2022) which suggests that countries in the same economic groupings have stronger connectedness in the sovereign credit market. In our study, we demonstrate that the same is true for the FX market as well where the difference is more highlighted.

**Figure 7a: CDS expected proportion in distress, given an AE or EM in distress**




Note: CDS expected proportion in distress given one market is in distress

**Figure 7b: FX expected proportion in distress, given an AE or EM is in distress**

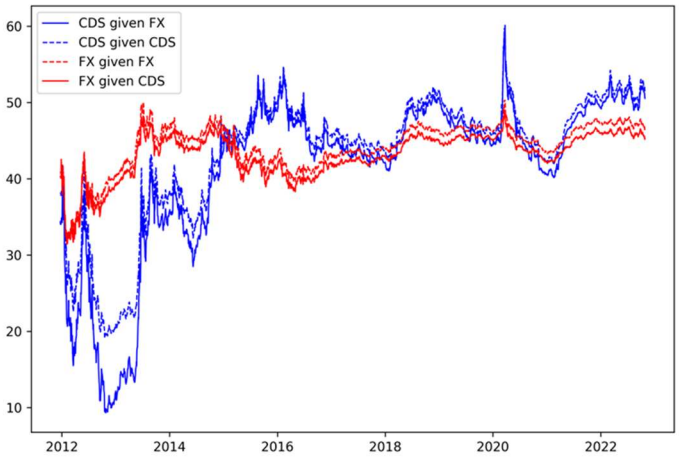



Note: FX expected proportion in distress given one market is in distress.

Interestingly, in the case of FX markets, results clearly suggest asymmetric effects from AE to EM compared to spillovers from EM to AE, being the former clearly stronger than the latter. For the CDS markets, the view is more balanced, as spillovers seem to transit with equal strength in both directions, although with a slightly higher effect from AE to EM. Dovern & Roye (2014) provide evidence for stronger transmission mechanisms from AE to EM. Our EPD estimates support those findings and yet claim that the transmission in the opposite direction is not negligible either.

Figure 7c displays spillovers between the CDS and FX markets, as measured by the crossed EPD using the model that incorporates all the series of CDS and FX. Notably, EPD in both markets has consistently increased over time, peaking at the beginning of 2020 and currently sitting at a record high. This highlights the potential occurrence of the Twin Ds given the current scenario of high debt and monetary policy tightening. The significant monetary easing observed globally during 2021 appeared to have temporarily reduced the expected proportion in distress in the two markets when one market is in distress, but it did not have a significant impact on reducing the EPD in each market separately. However, both EPDs remain at very high levels, especially in the CDS markets.

**Figure 7c: CDS and FX EPD given one CDS or FX market in distress**




Note: CDS or FX expected proportion of market in distress given one CDS or FX market is in distress. The results are based on a model that combines the CDS-FX datasets.

## 6. Conclusions and Policy Implications

Recent advances in deep learning for artificial intelligence provide rich possibilities for simulating market scenarios with multivariate time series that exhibit intricate patterns. In this study, we demonstrate the potential of generative AI in estimating systemic risk statistics, comparing the performance of two GAN-based models with Variational Autoencoders, and provide evidence in favor of the latter, as VAE achieves higher predictive accuracy by performing better noise reduction.

Our work contributes to the estimation of the probability of distress in the FX and CDS markets, which has not been previously estimated. Our results indicate that by the end of our sample period, both the joint probability of distress and the expected proportion of markets in distress for the two markets are at historically record highs. This highlights the real possibility of a systemic risk event occurring in these markets.

Our systemic risk statistics can be employed to track the impact of monetary and macro-prudential policies in real-time. For example, we observe a positive effect of monetary policy easing following the pandemic, as reflected in the reduction of the crossed EPDs of the markets during that period. However, this effect was temporary, as the EPD rebounded when policies tightened by the end of 2022.

Our findings provide valuable insights for monetary policy in developed economies, where the potential for spillover effects to create global instability exists, despite historical insensitivity of monetary policy in these economies to market instability in emerging market and low-income developing economies. Furthermore, our results are particularly important for central banks in emerging market and low-income developing economies, where macro-prudential regulations can be employed to mitigate the negative spillovers generated by policies in central economies.

We find that spillovers within economic groups are stronger than in-between these groups. Additionally, the transmission is stronger from advanced economies to emerging and developing market economies than the opposite. This is true for both the CDS and FX markets.

Our results underscore significant upticks in both sovereign risk and currency risk following the European debt crisis in 2016, the Covid-19 pandemic in 2020, and the tightening of global monetary policy in 2022. These occurrences notably impacted the likelihood of distress in these markets, as evidenced by substantial increases in our JPD indicators.

As per the TimeVAE analysis, the joint probability of distress in the CDS market reaches its peak during the Covid-19 pandemic and maintains historically high levels at the conclusion of our sample period. The peak JPD in 2020 for CDS markets connects with a similarly heightened JPD in FX markets, highlighting the interconnected nature of default and depreciation in the twin markets.

All in all, our generative models generate realistic counterfactual scenarios that are yet to occur in financial markets but are likely to materialize in worst-case scenarios.

## References


Acharya, VV, Pedersen, LH, Philippon, T, Richardson, M (2017) Measuring systemic risk, The Review of Financial Studies, 30, 2-47.

Adrian, T, Brunnermeier, MK (2016) CoVaR, American Economic Review, 106, 1705-1741.

Ait-Sahalia, Y, Laeven, RJ, Pelizzon, L (2014) Mutual excitation in euro- zone sovereign CDS. Journal of Econometrics, 183, 151-167.

Angelico, C, Marcucci, J, Miccoli, M, Quarta, F (2022) Can we measure inflation expectations using Twitter? Journal of Econometrics, 228, 259-277.

Arellano, C, Bai, Y, Lizarazo, S (2017) Sovereign risk contagion, NBER Working Paper Series, 24031.

Athey, S (2019). 21. The Impact of Machine Learning on Economics. *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans and Avi Goldfarb, Chicago: University of Chicago Press, 507-552.

Athey, S, Imbens, GW (2019) Machine Learning Methods That Economists Should Know About, Annual Review of Economics, 11, 685-725.

Augustin, P (2018) The term structure of CDS spreads and sovereign credit risk, Journal of Monetary Economics, 96, 53-76.

Augustin, P, Chernov, M, Song, D (2020) Sovereign credit risk and exchange rates: Evidence from CDS quanto spreads, Journal of Financial Economics, 137, 129-151.

Augustin, P, Sokolovski, V, Subrahmanyam, MG, Tomio, D (2022) How sovereign is sovereign credit risk? Global prices, local quantities, Journal of Monetary Economics, 131, 92-111.

Babii, A, Ghysels, E, Striaukas, J (2021) Machine Learning Time Series Regressions With an Application to Nowcasting, Journal of Business and Economic Statistics, 40, 1094-1106.

Bai, J, Wang, P (2015) Identification and Bayesian Estimation of Dynamic Factor Models, Journal of Business and Economic Statistics, 33, 221-240.

Bai, J, Ng, S (2002) Determining the Number of Factors in Approximate Factor Models, Econometrica, 70, 191-221.

Bai, J, Ng, S (2007). Determining the number of primitive shocks in factor models, Journal of Business and Economic Statistics,52-60.

Bai, J, Ng, S (2008). Large dimensional factor analysis. Foundations and Trends in Econometrics, 3, 89–163.

Bandiera, O, Prat, A, Hansen, S, Sadun, R (2020) CEO Behavior and Firm Performance, Journal of Political Economy, 128: 1325-1369.

Bianchi, D, Buchner, M, Tamoni, A (2021) Bond Risk Premiums with Machine Learning, The Review of Financial Studies, 34: 1046-1089.

Bianchi, F, Ludvigson, SC, Ma, S (2022) Belief Distortions and Macroeconomic Fluctuations, American Economic Review, 112:2269-2315.

Borio, C, McCauley, R, McGuire, P, Sushko, V (2016) Covered interest parity lost: understanding the cross-currency basis, BIS Quarterly Review.

Bose, D, Cordes, H, Nolte, S, Schneider, JC, Camerer, CF (2022) Decision Weights for Experimental Asset Prices Based on Visual Salience, The Review of Financial Studies, 35, 5094-5126.

Burke, J, Jamison, J, Karlan, D, Mihaly, K, Zinman, J (2022) Credit Building or Credit Crumbling? A Credit Builder Loans Effects on Consumer Behavior and Market Efficiency in the United States, The Review of Financial Studies, 0.1093/rfs/hhac060

Cahan E, Bai, J, Ng, S (2023) Factor-based imputation of missing values and covariances in panel data of large dimensions, Journal of Econometrics, 233, 113-131.

Carr, P, Wu, L (2007) Theory and evidence on the dynamic interactions between sovereign credit default swaps and currency options, Journal of Banking & Finance, 31, 2383–2403.

Chen, L, Pelger, M, Zhu, J (2023) Deep Learning in Asset Pricing, Management Science, https://doi.org/10.1287/mnsc.2023.4695.

Chen, MA, Wu, QX, Yang, BZ (2019) How Valuable Is FinTech Innovation?, The Review of Financial Studies, 32, 2062-2106.

Chen, Z., Zeng, Z., Shen, H., Zheng, X., Dai, P., & Ouyang, P. (2020). DN-GAN: Denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images. Biomedical Signal Processing and Control, 55, 101632.

Chernov, M, Creal, D, Hördahl, P (2023) Sovereign credit and exchange rate risks: Evidence from Asia-Pacific local currency bonds, Journal of International Economics, 140, 103692

Chernozhukov, V, Escanciano, JC, Ichimura, H, Newey, WK, Robins, JM (2022) Locally Robust Semiparametric Estimation, Econometrica, 90, 1501-1535.

Chernozhukov, V, Newey, WK, Singh, R (2022) Automatic Debiased Machine Learning of Causal and Structural Effects, Econometrica, 90, 967-1027.

Chiang, HD, Kato, K, Ma, YK, Sasaki, Y (2021)Multiway Cluster Robust Double/Debiased Machine Learning, Journal of Business and Economic Statistics, 40, 1046-1056

Chuliá, H, Fernández, J, Uribe, JM (2018) Currency downside risk, liquidity, and financial stability, Journal of International Money and Finance, 89, 83–102

Connor, G, Korajczyk, RA (1988) Risk and return in an equilibrium APT: Application of a new test methodology, Journal of Financial Economics, 21, 255–289.

Dawood, M, Horsewood, N, Strobel, F (2017) Predicting sovereign debt crises: An Early Warning System approach, Journal of Financial Stability, 28, 16-28.

Deryugina, T, Heutel, G, Miller, NH, Molitor, D, Reif, J (2019) The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction, American Economic Review, 109, 4178-4219.

Desai, A, Freeman, C, Wang, Z, Beaver, I (2021) TimeVAE: A Variational Auto-Encoder for Multivariate Time Series Generation. ArXiv preprint arXiv:2111.08095.

Dong, X, Li, Y, Rapach, DE, Zhou, GF (2022) Anomalies and the Expected Market Return, Journal of Finance, 77, 639-681.

Dovern, J., & van Roye, B. (2014). International transmission and business-cycle effects of financial stress. Journal of Financial Stability, 13, 1-17.

Du, W, Schreger, J (2016) Local currency sovereign risk, Journal of Finance, 71, 1027–1070.

Du, W, Tepper, A, Verdelhan, A (2018) Deviations from covered interest rate parity, Journal of Finance, 73, 915–957.

Duffie, D, Pedersen, LH, Singleton, KJ (2003) Modeling sovereign yield spreads: a case study of Russian debt, Journal of Finance, 58, 119–159

Easley, D, de Prado, ML, O'Hara, M, Zhang, ZB (2021) Microstructure in the Machine Age, The Review of Financial Studies, 34, 3316-3363.

Engel, J, Agrawal, KK, Chen, S, Gulrajini, I, Donahue, C, Roberts, A (2019) GANSynth: Adversarial Neural Audio Synthesis, Proceedings of the International Conference on Learning Representations- ICLR 2019.

Fan, J, Jiang, B, Sun, Q (2022) Bayesian factor-adjusted sparse regression, Journal of Econometrics, 230, 3-19.

Fan, J, Masini, R, Medeiros, M (2021) Bridging factor and sparse models, arXiv:2102.11341 [econ.EM]. Accessed on March 4 2023.

Farbmacher, H, Huber, M, Laffers, L, Langen, H, Spindler, M (2022) Causal mediation analysis with double machine learning, Econometrics Journal, 25, 277-300.

Farbmacher, H, Low, L, Spindler, M (2022) An explainable attention network for fraud detection in claims management, Journal of Econometrics, 228, 244-258.

Forni, M, Hallin, M, Lippi, M, Reichlin, L (2000) The generalized dynamic-factor model: identification and estimation, The Review of Economics and Statistics, 82, 540–554.

Forni, M, Hallin, M, Lippi, M, Reichlin, L (2005) The generalized dynamic factor model: one-sided estimation and forecasting, Journal of the American Statistical Association, 100, 830–840.

Forni, M, Reichlin, L (1998). Let's get real: factor to analytical cycle approach dynamics disaggregated business, The Review of Economic Studies, 65, 453–473.

Giesecke, K, Kim, B (2011) Systemic risk: What defaults are telling us, Management Science, 57(8), 1387-1405.

Gilchrist, S, Wei, B, Yue, VZ, Zakrajšek, E (2022) Sovereign risk and financial risk, Journal of International Economics, 136, 103603.

Gomez-Gonzalez J.E., Uribe, J.M., Valencia, O (2023), Risk spillovers between global corporations and Latin American sovereigns: global factors matter" Applied Economics 55 (13): 1477-1496.

Goodfellow, I, Pouget-Abadie, J, Mirza, M, Xu, B, Warde-Farley, D, Ozair, S, Courville, A, Bengio, Y (2014) Generative Adversarial Nets. 27th Conference on Neural Information Processing Systems.

Greenwood-Nimmo, M., Nguyen, V. H., & Shin, Y. (2023). What is mine is yours: Sovereign risk transmission during the European debt crisis. Journal of Financial Stability, 65, 101103.

Gu, SH, Kelly, B, Xiu, DC (2020) Empirical Asset Pricing via Machine Learning, The Review of Financial Studies, 33, 1879-1890.

Gu, SH, Kelly, B, Xiu, DC (2021) Autoencoder asset pricing models, Journal of Econometrics, 222, 429-450.

Hartmann, P, Straetmans, S, de Vries, C (2006) Banking system stability: A cross-atlantic perspective, in M. Carey and R.M. Stulz (eds), *The Risks of Financial Institutions*, University of Chicago Press, Chicago, 133-192.

Heinz, FF, Sun, YM (2014) Sovereign CDS spreads in Europe: The role of global risk aversion, economic fundamentals, liquidity, and spillovers, International Monetary Fund, Working Paper No. 2014/017

Hinton, GE, Salakhutdinov, RR (2006) Reducing the dimensionality of data with neural networks, Science, 313 (5786), 504–507.

Ilzetzki, E, Reinhart, C Rogoff, KS (2019) Exchange Arrangements Entering the 21st Century: Which Anchor Will Hold?, Quarterly Journal of Economics, 134, 599-646.

Kalemli-Özcan, Ş, Varela, L (2021) Five facts about the UIP premium, NBER Working Paper Series, 28923

Kingma, D P, Welling, M (2019). An introduction to variational autoencoders. Foundations and Trends in Machine Learning, 12(4), 307–392.

Kingma, DP, Welling, M (2013) Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

Klein, N., Smith, M. S., & Nott, D. J. (2023). Deep distributional time series models and the probabilistic forecasting of intraday electricity prices. Journal of Applied Econometrics.

Kose, M., Ohnsorge, F., Sugawara, N. 2021. A Mountain of Debt: Navigating the Legacy of Pandemic. Policy Research Working Paper 9800, World Bank Group.

Kozak, S, Nagel, S, Santosh, S (2020) Shrinking the cross-section, Journal of Financial Economics, 135, 271-292.

Le, C., Dickinson, D., & Le, A. (2022). Sovereign risk spillovers: A network approach. Journal of Financial Stability, 60, 101006.

Lee, JH, Shi, ZT, Gao, Z (2022) On LASSO for predictive regression, Journal of Econometrics, 229, 322-349.

Leippold, M, Wang, Q, Zhou, WY (2022) Machine learning in the Chinese stock market, Journal of Financial Economics, 145, 64-82.

Lettau, M., Maggiori, M., Weber, M. (2014). Conditional risk premia in currency markets and other asset classes, Journal of Financial Economics, 114, 197-225.

Li, X, Metsis, V, Wang, H, Ngu, A H (2022) TTIMEGAN: A Transformer-based Time-Series Generative Adversarial Network. arXiv preprint arXiv:2202.02691.

Longstaff, F. A., Pan, J., Pedersen, L. H., Singleton, K. J. (2011). How sovereign is sovereign credit risk?, American Economic Journal: Macroeconomics, 3, 75-103.

Lucas, A, Schwaab, B, Zhang, X (2014) Conditional Euro Area Sovereign Default Risk, Journal of Business and Economic Statistics, 32, 271-284.

Lustig, H., Verdelhan, A. (2007). The cross section of foreign currency risk premia and consumption growth risk, American Economic Review, 97, 89-117.

Lustig, H., Roussanov, N., Verdelhan, A. (2014). Countercyclical currency risk premia, Journal of Financial Economics, 111, 527-553.

McCracken, M, Ng, S (2020) FRED-QD: A quarterly database for macroeconomic research, National Bureau of Economic Research, Working Paper 26872.

Medeiros, MC, Vasconcelos, GFR, Veiga, A, Zilberman, E (2021) Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods, Journal of Business and Economic Statistics, 39, 98-119.

Mihai, I., Neagu, F. (2011) CDS and government bond spreads-how informative are they for financial stability analysis? IFC Bulletins chapters, 34, 415-429.

Na, S, Schmitt-Grohé, S, Uribe, M, Yue, V (2018) The twin Ds: optimal default and devaluation. American Economic Review, 108, 1773–1819.

Obaid, K, Pukthuanthong, K (2022) picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news, Journal of Financial Economics, 144, 273-297.

Oh, D. H., Patton, A. J. (2018) Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads, Journal of Business & Economic Statistics, 36, 181-195.

Packer, F, Suthiphongchai, C (2003) Sovereign credit default swaps. BIS Quarterly Review, December.

Pascual, S, Bonafonte, A, Serra, J (2017) SEGAN: Speech enhancement generative adversarial network. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2017-August (D), 3642–3646.

People, H. Back, B (2022) Fiscal Monitor. IMF e-library, 38. https://www.elibrary.imf.org/downloadpdf/book/9798400212741/9798400212741.pdf

Richardson, E, Alaluf, Y, Patashnik, O, Nitzan, Y, Azar, Y, Shapiro, S, Cohen-Or, D, (2021) Encoding in Style: A StyleGAN Encoder for Image-to-Image Translation, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2287-2296.

Skorokhodov, I, Tulyakov, S, Elhoseiny, M (2022) StyleGAN-V: A Continuous Video Generator With the Price, Image Quality and Perks of StyleGAN2, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3626-3636.

Smith, LN (2018) A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay. arXiv preprint arXiv:1803.09820.

Stock, J H, Watson, MW (2012) Forecasting using principal components from a large number of predictors, Journal of the American Statistical Association, 97, 1167–1179.

Stock, JH, Watson, MW (2011) Dynamic factor models. In M.P. Clements & D.F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting,* (pp. 35-59). *Oxford: Oxford University Press*.

Van der Maaten, L, Hinton, G (2008) Visualizing data using t-sne, Journal of Machine Learning Research, 9, 2579–2605.

Vaswani, N, Shazeer, N, Parmar, N, Uszkoreit, J, Jones L, Gomez, A, Kaiser, L, Polosukhin I (2017) Attention is all you need, 31st Conference on Neural Information Processing Systems.

Wu, E, Erdem, M, Kalotychou, E, Remolona, E (2016) The anatomy of sovereign risk contagion, Journal of International Money and Finance, 69, 264-286.

Xiong, R, Pelger, M (2023) Large dimensional latent factor modeling with missing observations and applications to causal inference, Journal of Econometrics, 233, 271-301.

Yoon, J, Jarrett, D, Van der Schaar, M (2019) Time-series generative adversarial networks. 32nd Conference on Neural Information Processing Systems.

# Appendix

*A.1. Some additional practical consideration*

The residuals data consist of a $T$ by $N$ matrix, where $N$ represents the number of instruments present in the market and $T$ refers to the number of business days. Then, the data are separated in monthly blocks of cardinality $S$. To generate multiple samples, we use a moving window of 1 day and rearrange our data in a $(T - S) \times S \times N$ matrix. Based on the trained model on the data above, we then generate $D/S$ random samples, each as a $S \times N$ matrix. The purpose of augmenting samples in this way is to preserve sequential features extracted from the trained data.

*Selecting hyper-parameters*

The main hyperparameters in the TimeVAE are: the size of latent space dimension, the number of iterations, batch size, number of trend polynomials and size of hidden layers. For the TimeGAN, the list includes the size of latent space, selected module (gru, lstm, lstmLN), number of layers, batch size and number of iterations. For TTSGAN, we adjust the hyperparameters of the latent space dimension, number of iterations and batch size.

We explain below the reasoning behind the choice of all the major hyperparameters. We keep unchanged hyperparameters such as the learning rate for discriminator and generator and beta values for Adam optimizer, since they have already been tuned to optimal values in benchmark results and we consider them as a part of the model itself. Overall, the grounds for making the choices are very similar, thus we tend to apply the same choices for the same hyperparameters for all models to keep the results more comparable across our implementations. The performance metrics, however, demonstrate that the choice of model matters more than the choice of hyperparameters in most of the cases.

To select the optimal size of the *latent space*, we generate data using different latent sizes up to the sequence length, i.e. the number of days in each sample. Yoon et al. (2019) recommend using latent space equal to half of the size of input features or equal to the sequence length as mentioned in the published source code. In terms of the prediction score for TimeVAE, the difference between selecting a high number of dimension sizes at 30 versus a low number of dimension sizes at 2 is very small, around 0.005 in favor of the higher dimension size. However, visual metrics suggest better performance with a higher dimension size, thus we end up using a latent space equal to 30.

Smith (2018) shows in their tests of different *batch sizes* that at larger the batch size may improve the result until a certain size, but the impact becomes negligible thereafter. However, increasing

batch size would increase the number of epochs for the same number of iterations and this would make the computational burden on memory much higher. We cannot freely test different batch sizes for all the models due to different hardware requirements of each model. For TimeVAE and TimeGAN, this is less of an issue, however, TTSGAN, which heavily relies on image processing structure of transformer models, requires much more memory and it becomes infeasible beyond a certain level while keeping the number of iterations above 10,000. Another drawback of financial time series for the batch size selection is the number of samples being relatively small. Our database includes 3,077 samples which already suggests we may not simply divide our data to higher batches similar to the datasets with hundreds of thousands of samples. For our dataset, we could test all the models with batch size equal to 64 or less. The metrics suggest that the benefit of having batch size of 16 vs 64 is very little for TimeVAE, and almost non-existent for TTSGAN and TimeGAN.

We keep the *number of iterations* at 10,000, as it is large enough for all the models since the change in losses already becomes zero or close to zero after a few hundred iterations. The *hidden layer sizes*, i.e. the number of filters in TimeVAE is unchanged for CDS and FX at default values, but we adjust them to make the number of layers of the first layer bigger than the feature size (i.e. number of instruments) and incrementally increase thereafter. It is standard design in Variational encoders with convolutional layers that hidden layer sizes increase in the encoder level, while the dimension of original data is reduced in each layer. For CDS and FX separately, we use 3 hidden layers with the number filters equal to 50, 100, 200, for the mixed dataset, we set them equal to 150 (about twice the number of instruments), 250, 350. Increasing the number of filters in hidden layers slightly improves the results. For TimeGAN, we use *three layers* and show the metrics results for all types of layers.

The number of polynomials for TimeVAE is set to 5 to capture polynomial trends up to degree 5. Prediction score shows an improvement of 0.02 from 0 trend polynomials to 5 polynomials and the improvement thereafter is negligible.

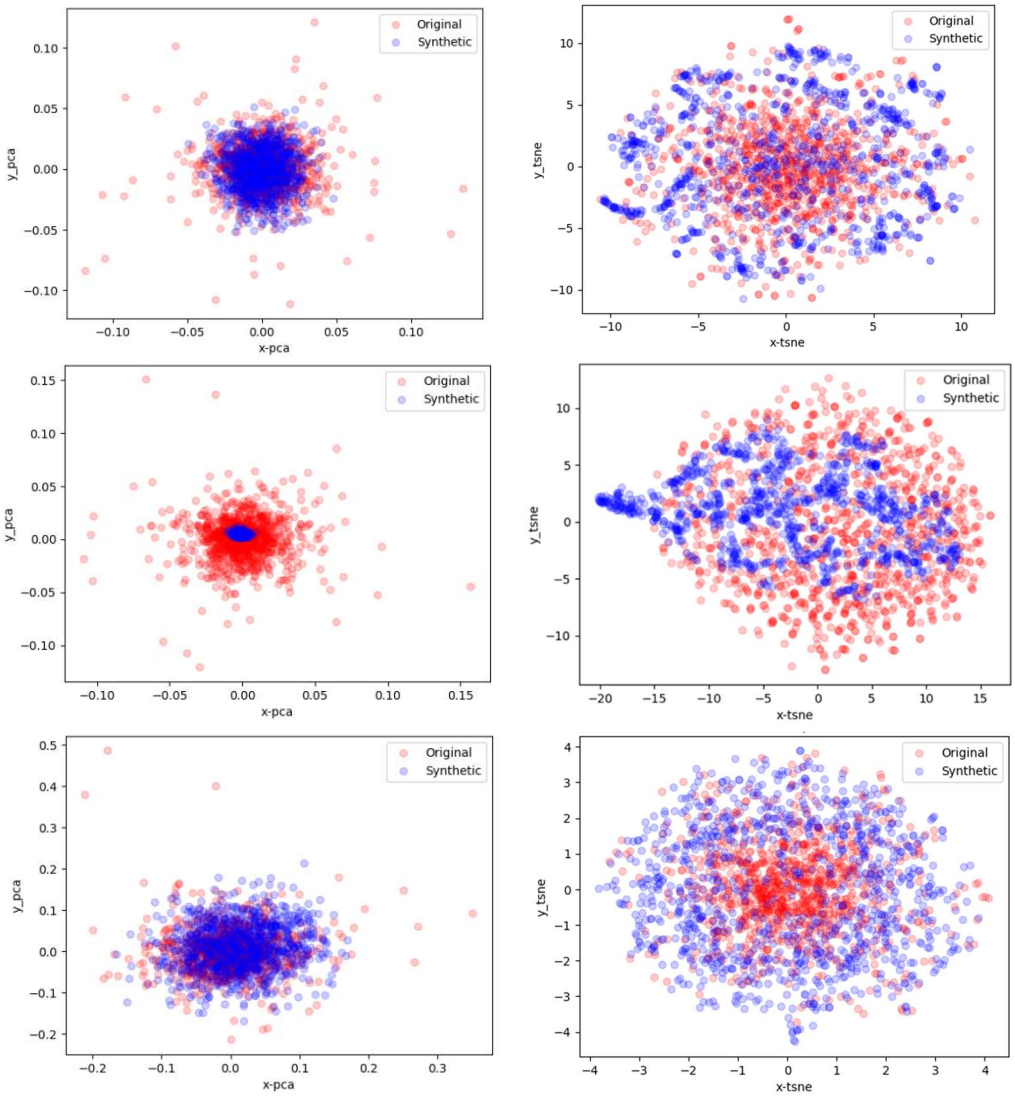
## Table A1. List of currencies and CDS contracts

| Country | Short Name | Group | CDS | Currency | Currency Name |
|---|---|---|---|---|---|
| Australia | AU | Advanced | - | AUD | Australian dollar |
| Austria | AT | Advanced | AT5Y | EUR | Euro |
| Belgium | BE | Advanced | BE5Y | EUR | Euro |
| Brazil | BR | Emerging Market | BR5Y | BRL | Brazilian real |
| Brunei | BN | Emerging Market | - | BND | Brunei dollar |
| Canada | CA | Advanced | - | CAD | Canadian dollar |
| Chile | CL | Emerging Market | CL5Y | CLP | Chilean peso |
| China | CN | Emerging Market | CN5Y | - | - |
| Colombia | CO | Emerging Market | CO5Y | COP | Colombian peso |
| Denmark | DK | Advanced | DK5Y | - | - |
| Finland | FI | Advanced | FI5Y | EUR | Euro |
| France | FR | Advanced | FR5Y | EUR | Euro |
| Germany | DE | Advanced | DE5Y | EUR | Euro |
| Ghana | GH | Least Developed Economy | - | GHS | Ghanaian cedi |
| Hungary | HU | Emerging Market | HU5Y | HUF | Hungarian forint |
| Iceland | IS | Advanced | - | ISK | Icelandic króna |
| Indonesia | ID | Emerging Market | ID5Y | - | - |
| Ireland | IE | Advanced | IE5Y | EUR | Euro |
| Israel | IL | Advanced | - | ILS | Israeli new shekel |
| Italy | IT | Advanced | IT5Y | EUR | Euro |
| Japan | JP | Advanced | JP5Y | JPY | Japanese yen |
| Kazakhstan | KZ | Emerging Market | KZ5Y | - | - |
| Madagascar | MG | Least Developed Economy | - | MGA | Malagasy ariary |
| Malaysia | MY | Emerging Market | MY5Y | MYR | Malaysian ringgit |
| Mexico | MX | Emerging Market | MX5Y | MXN | Mexican peso |
| Netherlands | NL | Advanced | NL5Y | EUR | Euro |
| New Zealand | NZ | Advanced | - | NZD | New Zealand dollar |
| Norway | NO | Advanced | - | NOK | Norwegian krone |
| Panama | PA | Emerging Market | PA5Y | - | - |
| Paraguay | PY | Emerging Market | - | PYG | Paraguayan guaraní |
| Peru | PE | Emerging Market | PE5Y | - | - |
| Philippines | PH | Emerging Market | PH5Y | PHP | Philippine peso |
| Poland | PL | Emerging Market | PL5Y | - | - |
| Portugal | PT | Advanced | PT5Y | EUR | Euro |
| Qatar | QA | Emerging Market | QA5Y | - | - |
| Republic of Korea | KR | Advanced | KR5Y | KRW | South Korean won |
| Romania | RO | Emerging Market | RO5Y | - | - |
| Russia | RU | Emerging Market | - | RUB | Russian ruble |
| São Tomé and Príncipe | ST | Least Developed Economy | - | STN | São Tomé and Príncipe dobra |
| Saudi Arabia | SA | Emerging Market | SA5Y | - | - |

| | | | | | |
|---|---|---|---|---|---|
| Seychelles | SC | Emerging Market | - | SCR | Seychellois rupee |
| Singapore | SG | Advanced | - | SGD | Singapore dollar |
| South Africa | ZA | Emerging Market | ZA5Y | ZAR | South African rand |
| Spain | ES | Advanced | ES5Y | EUR | Euro |
| Sweden | SE | Advanced | SE5Y | SEK | Swedish krona |
| Thailand | TH | Emerging Market | TH5Y | THB | Thai baht |
| Tonga | TO | Emerging Market | - | TOP | Tongan pa'anga |
| Turkey | TR | Emerging Market | TR5Y | TRY | Turkish lira |
| Uganda | UG | Least Developed Economy | - | UGX | Ugandan shilling |
| United Kingdom | GB | Advanced | GB5Y | GBP | Pound sterling |
| Uruguay | UY | Emerging Market | - | UYU | Uruguayan peso |
| Vanuatu | VU | Emerging Market | - | VUV | Vanuatu vatu |
| Vietnam | VN | Least Developed Economy | VN5Y | - | - |
| Yemen | YE | Least Developed Economy | - | YER | Yemeni rial |
| Zambia | ZM | Least Developed Economy | - | ZMW | Zambian kwacha |

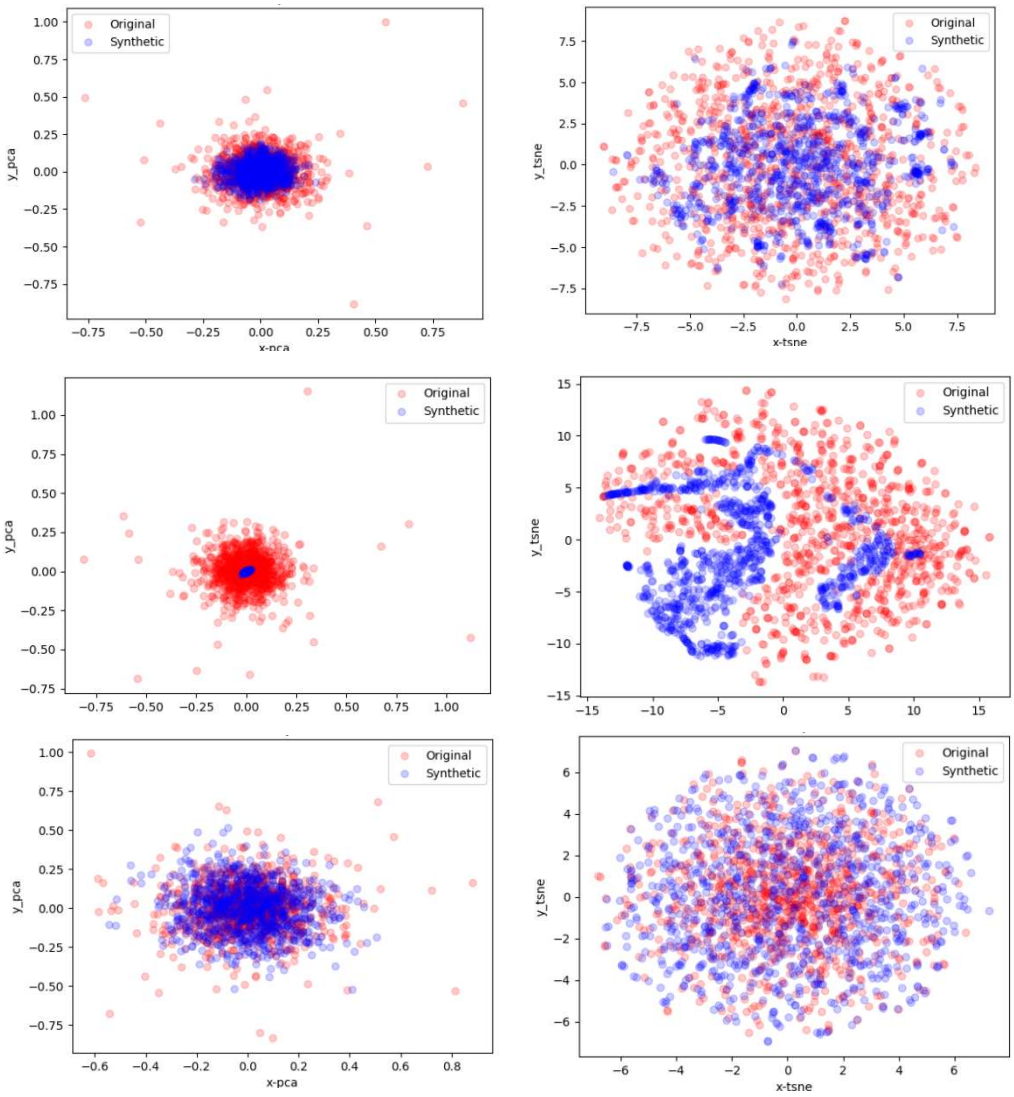
Note: CDS and currencies used in the study.

Figure A1a. Comparison of generative models by visual metrics (CDS market)



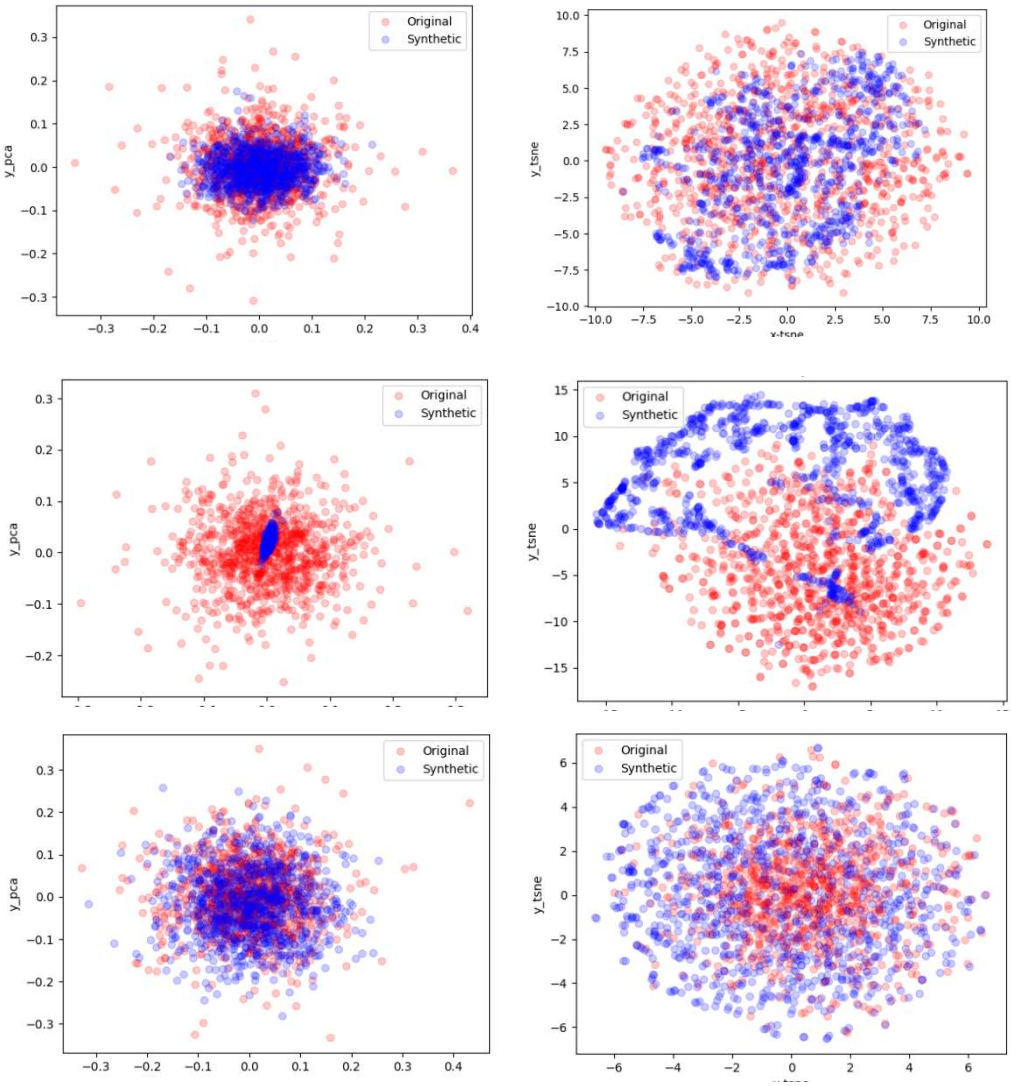

*Note: These charts show the comparison of performance of generative models by visual metrics PCA (left) and TSNE (right). The generative models are ordered from the first row to the last in the following fashion: TimeGAN, TimeVAE, TTSGAN*

Figure A1b. Comparison of generative models by visual metrics (FX market)




*Note: These charts show the comparison of performance of generative models by visual metrics PCA (left) and TSNE (right). The generative models are ordered from the first row to the last in the following fashion: TimeGAN, TimeVAE, TTSGAN*

Figure A1c. Comparison of generative models by visual metrics (CDS & FX combined market)




*Note: These charts show the comparison of performance of generative models by visual metrics PCA (left) and TSNE (right). The generative models are ordered from the first row to the last in the following fashion: TimeGAN, TimeVAE, TTSGAN*