Simulating Economic Heavy-Tailed Distributions with Quantum Boltzmann Machines*

Noorain Noorani[1], Valerio Astuti[2], Giuseppe Bruno[2], Lerby M. Ergun[3], and Vladimir Skavysh[3]

[1]*University of Maryland*
[2]*Bank of Italy*
[3]*Bank of Canada*

Monday 19th May, 2025

Abstract

Accurately estimating tail risk in financial markets is crucial yet difficult, especially when data are scarce. This paper investigates the use of Quantum Boltzmann Machines (QBMs), a class of quantum generative models capable of capturing complex, high-dimensional, and heavy-tailed distributions through quantum superposition and entanglement. Compared to classical Restricted Boltzmann Machines, QBMs offer greater expressive power and sampling efficiency. We apply this framework to assess risk for newly listed firms with limited historical data and find that QBM-augmented estimates significantly improve the prediction of long-term risk measures, including Value-at-Risk and expected shortfall, making QBMs a promising tool for financial modeling.

Keywords: Quantum Boltzmann Machines (QBMs), Fat-Tailed Distributions, Economic Modeling, Data Augmentation, Generative Models, Energy-Based Models, Restricted Boltzmann Machines (RBMs), Quantum Computing in Economics, Financial Market Simulation, Heavy-Tailed Distributions

*Corresponding author: skav@bankofcanada.ca

The views expressed in the paper are solely those of the authors and do not necessarily represent the views of the Bank of Italy or the Bank of Canada.

1 Introduction

In economic and financial systems, accurately modeling distributions that exhibit heavy tails, where extreme events occur more frequently than predicted by normal distributions, is crucial for understanding market dynamics and assessing risk. Traditional models often fall short in capturing these heavy-tailed behaviors, leading to underestimation of the likelihood and impact of rare but significant events. A better description and understanding of such distributions has occupied an important role in recent economic, financial, and statistical research (for a short review see Gabaix 2016). By their very nature, tail events are rare, limiting the number of available observations and complicating statistical inference. The resulting data sparsity leads to high estimation variance, particularly for tail-related measures. These challenges are further amplified in inherently data-scarce settings, such as short time series, where reliably characterizing the tails becomes even more difficult.

Existing approaches to this problem typically involve either fitting a parametric distribution to the limited available data, using methods such as maximum likelihood estimation or moment matching, or resampling the data through techniques like bootstrapping. While these methods are easy to implement, they have notable limitations. Parametric approaches require the specification of a functional form in advance, which can lead to model misspecification if the true distribution deviates from the assumed family. Resampling methods, in contrast, tend to underestimate the severity of tail events, particularly in heavy-tailed settings. This is less of a concern under Gaussian assumptions, where tail behavior is more closely tied to the center of the distribution. However, in power-law contexts, the maximum observed value is highly dependent on sample size, making it unlikely that resampled data will capture the full extent of future extremes.

The issue of small sample sizes in the estimation of risk measures such as Value-at-Risk (VaR) and Expected Shortfall (ES) has long been recognized in financial econometrics. Studies such as Danielsson and Vries (2005); Duffie and Pan (1997); Embrechts et al. (2005); Hoga (2022); McNeil and Frey (2000) document how tail-based risk measures are prone to high estimation error, especially when sample sizes are limited and the underlying distributions are heavy-tailed. More recent contributions have further explored these limitations in applied settings, showing how small-sample bias affects both unconditional and conditional risk estimates. For instance, Gao et al. (2022) investigate finite-sample distortions in ES estimation using extreme value theory based methods, while Patton et al. (2019) develop and evaluate non-parametric and semi-parametric estimators of conditional ES in low-data settings. Similarly, Du et al. (2018) and Martins and Ziegel (2021) highlight the challenges in forecasting and back-testing tail risk in volatile or illiquid markets, where data sample period is short. Regulatory frameworks such as Basel III and the Fundamental Review of the Trading Book attempt to address these issues through conservative buffers and stressed-scenario measures, which serve as proxies for robustness rather than improving the core estimation process.

To address the modeling challenges posed by heavy-tailed distributions, recent research has turned to generative models that simulate such distributions in a non-parametric way (Ramzan et al. (2024)), thereby avoiding rigid structural assumptions. In this paper, we extend this line of work by exploring the use of quantum Boltzmann machines (QBMs), a class of quantum generative machine learning models inspired by statistical physics, for simulating heavy-tailed distributions in low-data environments. QBMs offer a promising approach due to their capacity

to model complex, high-dimensional probability distributions with a compact quantum representation (Tüysüz et al. 2024). We demonstrate how QBM-based approaches can improve the risk assessment of young or early-stage firms, where performance data is limited but accurately capturing downside risk is essential for investment and policy decisions.

Boltzmann Machines (BM), particularly their restricted variants (RBM), have been explored in the literature as generative models for learning complex probability distributions. RBMs, with their simplified bipartite structure, make learning more tractable compared to general BMs, introducing strong simplifications with respect to the original model but still retaining the meaningful predictive capacity (Smolensky et al. 1986; Hinton 2002; Hinton and Salakhutdinov 2006). However, when applied to large-scale problems or high-dimensional data, RBMs face significant scalability challenges. Training these models involves computationally expensive sampling techniques, such as Gibbs sampling or contrastive divergence, which become infeasible as the size of the network grows. Moreover, their reliance on approximations for gradient-based learning often results in suboptimal performance, particularly for capturing complex patterns. These limitations have led to a decline in the use of classical BMs, paving the way for alternative generative approaches, including quantum variants. QBMs, extending classical RBMs into the quantum domain, leverage quantum states and transformations to represent complex distributions that classical RBMs may not satisfactorily model (Anschuetz and Cao 2019; Song et al. 2019).

The advent and ongoing advancement of quantum computing have accelerated research into solving problems related to finance and economics, providing innovative methodologies and computational efficiencies that surpass classical approaches (Herman et al. 2023). Specific applications under active exploration include portfolio optimization, credit risk modeling, derivatives pricing, and systemic risk assessment, where classical methods often struggle with non-convex optimization landscapes, high-dimensional state spaces, and heavy-tailed probability distributions. Quantum optimization techniques have been applied to portfolio allocation problems with combinatorially many asset configurations (Buonaiuto et al. 2023). Meanwhile, quantum amplitude estimation enables more efficient calculation of risk measures (e.g., Value-at-Risk) (Woerner and Egger 2019; Skavysh et al. 2023) and option pricing via accelerated Monte Carlo simulations and Hamiltonian Simulation (Stamatopoulos et al. 2020; Stamatopoulos and Zeng 2024).

The application of quantum generative models in financial applications has largely been studied using gate-based Noisy Intermediate Scale Quantum (NISQ) devices and algorithms adapted to this kind of systems, such as for example Quantum Circucit Born Machines (Liu and Wang 2018) and Quantum Generative Adversarial Networks (Dallaire-Demers and Killoran 2018), employing parameterized quantum circuits to generate data. These generative techniques have been implemented and tested on real hardware (Zhu et al. 2022) garnering attention for the exploration of real applications in finance (Coyle et al. 2021). These applications have been further modified to help enhance financial analysis such as portfolio optimization, risk analysis, time series analysis and anomaly detection (see for example Orlandi et al. (2024); Ganguly (2023); Zhou et al. (2024); Bhasin et al. (2024); Stein et al. (2024)). Despite the research into these NISQ models, little research has been conducted on using QBM trained on quantum annealing hardware for financial applications.

To evaluate the practical value of our quantum generative modeling approach, we apply it to a challenging empirical setting: estimating financial risk for newly listed firms with limited

historical return data. Using daily return data from 400 U.S. firms obtained through the Centre for Research in Security Prices (CRSP) database, we focus on measuring risk based on the first year of returns following their initial public offering. This short sample is used to compute standard financial risk metrics—including standard deviation, VaR, ES, skewness, kurtosis, and tail index. These first-year risk measures are then augmented with 400 observations synthesized by a trained QBM, which learns from the observed return distribution. To assess whether the augmented data improve risk estimation, we compare the original one-year estimates and QBM-augmented estimates against the corresponding metrics computed over a full five-year horizon. This extended horizon serves as a more stable benchmark for a firm's risk profile, offering a natural target for evaluating predictive performance. We also benchmark our results against a classical RBM to isolate the contribution of quantum structure in the generative process.

Our results show that QBM-generated data significantly enhance the prediction of longer-term risk measures across several dimensions. The added value is clearest for metrics that rely on extreme outcomes, such as VaR, ES, kurtosis, and the tail index, where small samples typically struggle. For example, the QBM-augmented one-year VaR yields a significant coefficient, whereas its RBM-augmented counterpart does not. These improvements align with the expectation that tail sensitive measures benefit most from data augmentation and highlight QBM's ability to capture complex, non-Gaussian structures in the data. In contrast, standard deviation and skewness, measures that are less sensitive to the tails of the distribution, show smaller or no gains. Taken together, these results suggest that QBM offers a promising tool for augmenting financial risk models in settings where traditional approaches are constrained by data availability.

The rest of the paper is structured as follows: Section 2 introduces Restricted Boltzmann Machines, and Section 3 the Quantum counterpart; Section 4 describes the simulation analysis. Subsequently, Section 5 reports the empirical analysis for young firms, followed by the conclusions.

2 Restricted Boltzmann Machine

Boltzmann machines (BMs) are a class of neural network models that have played a pivotal role in the advancement of unsupervised learning and probabilistic modeling. Introduced by Geoffrey Hinton, Terrence Sejnowski and others in the first half of the 80s (Fahlman et al. 1983; Ackley et al. 1985) as an evolution of Hopfield models (Hopfield 1982), these networks are named after the Boltzmann distribution in statistical mechanics, which models their probabilistic behavior. BMs are characterized by their network architecture, with units symmetrically connected and operating in a stochastic manner. This design enables them to model complex probability distributions over high-dimensional data, making them particularly suited for tasks such as dimensionality reduction and feature learning. However, their computational complexity and training challenges have historically limited their widespread use, leading to the development of more tractable variants like Restricted Boltzmann Machines (RBMs). The roots of RBMs trace back to Smolensky et al. (1986), where the concept was originally introduced, and Hinton (2002), in which an efficient training algorithm was described.

RBMs share the theoretical underpinnings of BMs: they are rooted in energy-based models and

Gibbs sampling, providing a wide set of tools for understanding their probabilistic functioning. Over the years, they have undergone refinements and adaptations, with researchers harnessing their capabilities for diverse applications, ranging from collaborative filtering to feature learning and dimensionality reduction.

This section delves into the theoretical underpinnings of RBMs, exploring their architecture, probabilistic modeling, training procedures, and applications. By unraveling the intricacies of RBMs, we aim to provide a comprehensive understanding of their historical evolution and significance in the broader landscape of machine learning and motivations for their quantum counterpart.

2.1 Basic Architecture

RBMs are a type of generative stochastic artificial neural network that consists of two layers: a visible layer and a hidden layer. The visible layer represents the input data, while the hidden layer captures higher-level features or representations learned from the input. These layers are interconnected by weights, and each node in one layer is connected to every node in the other layer. Notably, there are no connections within layers, creating a "restricted" connectivity pattern, as seen in Figure 1.

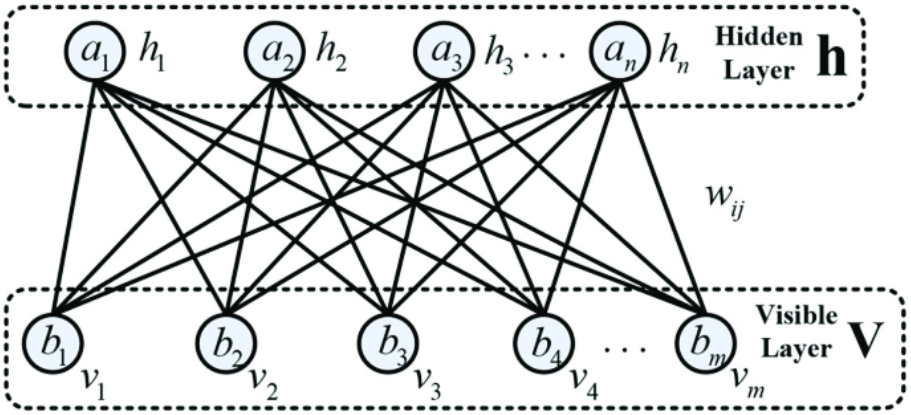

Figure 1: The basic structure of an RBM (Chu et al. 2018), where h_n and v_m represent the hidden and visible units respectively with associated bias vector a_n and b_m. The weights corresponding to the interactions strengths between visible and hidden units are denoted by w_{ij}.

The connectivity pattern in RBMs reflects their bipartite structure, where nodes in the visible layer are only connected to nodes in the hidden layer and vice versa. This restricted connectivity is the feature distinguishing Restricted Boltzmann Machines and Boltzmann Machines. It simplifies the learning process and enables more efficient training compared to fully connected networks. The absence of connections within layers reduces the model's complexity, making it computationally more tractable while still capturing intricate patterns in the data.

One key concept in understanding RBMs is the notion of energy. In the context of RBMs, the energy of a particular configuration (assignment of values to visible and hidden units) is a measure of the compatibility of that configuration with the model. The energy is defined in terms of

the weights and biases of the RBM. Lower energy configurations are more likely to occur, and the learning process aims to adjust the weights and biases to minimize the energy of observed data while at the same time maximizing the energy of unobserved data. This energy-based approach provides a probabilistic interpretation of RBMs, where the model assigns probabilities to different configurations, allowing them to be used for tasks such as sampling and generating new data.

2.2 Probabilistic Model

RBMs are inherently probabilistic models that leverage probability distributions to characterize the relationships between visible and hidden units. The probability of a particular configuration of visible and hidden units is expressed through a joint probability distribution. This distribution encapsulates the likelihood of observing a specific combination of states for both visible and hidden layers.

The joint probability distribution over visible and hidden units is defined in terms of the Boltzmann distribution, introduced and widely used in the field of statistical physics (see for example Kardar 2007 for an introduction). The probability $P(v,h)$ of a specified configuration is given by:

$$P(v,h) = \frac{e^{-E(v,h)}}{Z}. \tag{1}$$

The energy function $E(v,h)$ plays a pivotal role in defining this joint probability distribution. It measures the compatibility of a given configuration of visible units (v) and hidden units (h) within the model. Mathematically, the function is defined as:

$$E(v,h) = -\sum_{i=1}^{N_v} b_i v_i - \sum_{j=1}^{N_h} a_j h_j - \sum_{i=1}^{N_v}\sum_{j=1}^{N_h} w_{ij} v_i h_j. \tag{2}$$

Here N_v represents the number of visible units, N_h is the number of hidden units and w_{ij} denotes the weight between visible unit v_i and h_j. The values b_i and a_j correspond to the biases of visible v_i and hidden h_j units, respectively. The parameters b_i, a_j and w_{ij} are to be learned from the visible data, in order to allow for the hidden nodes to replicate the patterns in the training set. The normalization constant Z in Equation (1) ensures that the probabilities sum up to 1 over all possible configurations. It is defined as the sum of the exponential of negative energies over all possible visible and hidden unit configurations:

$$Z = \sum_v \sum_h e^{-E(v,h)}.$$

The Boltzmann distribution reflects the core probabilistic foundation of RBMs, where configurations with lower energy values are assigned higher probabilities, capturing the underlying relationships within the data. This probability framework enables RBMs to model complex distributions and serves as the basis for tasks such as sampling and generation. From the definition (1) we can define the marginal probabilities over visible units only summing over all possible hidden units configurations:

$$P(v) = \sum_h P(v,h). \tag{3}$$

This can be used, for example, to establish the probability to see a given visible nodes configuration, regardless of the values taken by the hidden nodes. In addition, an important consequence of Equation (1) is the simplifications it implies for the evaluation of the conditional probabilities. Every two node in a layer of the network are independent, once we condition on the nodes of the other layer. This is easily derived from the expression of the energy function (2), and from the structure of the network illustrated in Figure 1: each node in e.g. the hidden layer can be influenced only by the nodes in the visible layer. Conditioning on the latter, two nodes in the hidden layer cannot have any influence on each other. This conditional independence implies that the probability $P(v|h)$ factorizes over the visible nodes, and for each of them we have the factor:

$$P(v_i = 1|h) = \sigma\left(b_i + \sum_{j=1}^{N_h} w_{ij} h_j\right),$$

where $\sigma(x)$ is the logistic function $\sigma(x) = \frac{1}{(1+e^{-x})}$. This form of the conditional probability will be crucial in the training procedure, described in the next section.

2.3 Training Procedure

In this section we focus on the selection of the energy parameters in Equation (2). Training a RBM involves maximizing the log-likelihood of the observed data (or equivalently minimizing the negative log-likelihood) in the space of parameters b_i, a_j and w_{ij}. In the following we denote the set of parameters collectively as $\Theta = \{b_i, a_j, w_{ij}\}$, and introduce explicitly the parameter dependence of the probability (1), $P(v,h) = P(v,h|\Theta)$. From Equation (3) the log-likelihood function ($\mathscr{L}$) for visible data v is given by:

$$\mathscr{L}(\Theta|v) = \log P(v|\Theta), \tag{4}$$

and averaging over all the possible configurations in the training data for the visible nodes we obtain:

$$\mathscr{L}(\Theta) = \frac{1}{N_{\text{obs}}} \sum_v \log P(v|\Theta), \tag{5}$$

where v is summed over all the configurations found in the training data, and N_{obs} is the number of these observations. To maximize the log-likelihood function (5) we use the gradient of the function (Hinton and Salakhutdinov 2006). Hence, evaluating the gradient of the log-likelihood with respect to the parameters Θ is crucial for optimizing the RBM. It provides the direction in which the parameters should be adjusted to increase the log-likelihood.

Note that the log-likelihood (4) can be written as the difference between the two terms:

$$\mathscr{L}(\Theta|v) = \log \sum_h e^{-E(v,h,\Theta)} - \log \sum_{v',h'} e^{-E(v',h',\Theta)}.$$

This implies that the gradient of (5) can be written as:

$$\frac{\partial \mathscr{L}(\Theta)}{\partial \theta} = -\frac{1}{N_{\text{obs}}} \sum_{v,h} P(h|v,\Theta) \frac{\partial E(v,h,\Theta)}{\partial \theta} + \sum_{v',h'} P\left(v',h',\Theta\right) \frac{\partial E(v',h',\Theta)}{\partial \theta}.$$

Finally, the last expression can be written as:

$$\frac{\partial \mathscr{L}(\Theta)}{\partial \theta} = -\left\langle \frac{\partial E(v,h,\Theta)}{\partial \theta} \right\rangle_{data} + \left\langle \frac{\partial E(v,h,\Theta)}{\partial \theta} \right\rangle_{model},$$

where the first average $\langle \bullet \rangle_{data}$ is evaluated with the empirical distribution of visible nodes configurations in the observed data, and the second $\langle \bullet \rangle_{model}$ is obtained from the probability distribution (1) defining the model. The challenge arises in computing these expectation values exactly, because they involve the partition function Z, which requires summing over all possible configurations of visible and hidden units. This summation is intractable due to the exponential number of configurations, making the exact computation infeasible.

2.3.1 Contrastive Divergence

This is where Contrastive Divergence (CD) comes into play. Instead of computing the exact expectations, CD introduces an efficient approximation. The primary idea is to approximate the gradient of the log-likelihood function by comparing the model's response to real data (positive phase) and a "model" generated by Gibbs sampling (negative phase). It initializes the visible layer with a training sample, performs Gibbs sampling for a few steps to create the "model" average, and then computes the gradient using the positive phase (data) and negative phase (model) expectations (see Hinton 2002; Hinton and Salakhutdinov 2006 for details). CD exploits the idea that after a few Gibbs sampling steps, the model sample becomes a reasonable approximation of the true data distribution. The algorithm's efficiency comes from the fact that it avoids the need for exact sampling from the RBM distribution, making it computationally more feasible for training large models. The CD algorithm can be broken down into the following steps:

Initialization
Initialize the visible layer with a training sample (v) and compute the probabilities of the hidden layer being activated:

$$P(h_j = 1|v) = \sigma\left(b_j + \sum_{i=1}^{N_v} v_i w_{ij}\right).$$

We then sample binary values for the hidden layer using these probabilities to get the initial hidden values, $h^{(0)}$.

Negative Phase
Given the sampled hidden states from the initialization step, compute the probabilities of the visible units being activated:

$$P(v_i = 1|h^{(0)}) = \sigma\left(a_i + \sum_{i=1}^{N_h} h_j^{(0)} w_{ij}\right)$$

and sample binary values for the visible layer to get $v^{(1)}$.

Positive Phase
Now given the sampled visible states $v^{(1)}$, compute the probabilities of the hidden units being

activated again

$$P(h_j = 1|v^{(0)}) = \sigma\left(b_j + \sum_{i=1}^{N_v} v_i^{(0)} w_{ij}\right)$$

and sample binary values for the hidden layer to get $h^{(1)}$. Repeat the process, alternating between updating visible and hidden units, for several steps to create a model sample.

Compute Gradients
After repeating the process k times, we can use the two sets of data $(v, h^{(0)})$ and $(v^{(k)}, h^{(k+1)})$ to approximate the two averages $\langle \bullet \rangle_{data}$ and $\langle \bullet \rangle_{model}$. As an example, to update the values of the weights w_{ij} we obtain:

$$\Delta w_{ij} = \eta \frac{\partial \mathscr{L}(\Theta)}{\partial w_{ij}} = \eta \left(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model} \right).$$

The parameter η is the learning rate, which regulates the speed of convergence of the process. By iteratively applying Gibbs sampling and computing these differences, CD efficiently approximates the gradient of the log-likelihood, facilitating the training of RBMs on large datasets. The evaluated gradient can finally be used to change the parameters of the model in order to increase the likelihood of the training data and to better replicate the features of the distribution under study. After the training procedure, the model can be used to generate new data from the learned distribution using Equation (3).

2.4 Advantages and Challenges

RBMs prove adept at modeling intricate relationships and capturing high-dimensional dependencies. One key strength lies in their capacity for representation learning, as RBMs autonomously unveil hierarchical features when trained on unlabeled data (Decelle et al. 2023). This is particularly valuable in scenarios where feature extraction poses challenges. Moreover, the incorporation of non-linearity through the activation functions empowers RBMs to model complex relationships, a trait crucial for tasks where linear models fall short.

RBMs also excel in dimensionality reduction, effectively distilling essential information from high-dimensional datasets and mitigating the challenges associated with the curse of dimensionality. Their generative modeling capability enables the creation of new samples similar to the training data, proving advantageous in tasks such as image synthesis and recommendation systems (Salakhutdinov et al. 2007). Notably, RBMs robustly handle missing data during both training and inference, a practical feature for real-world datasets where incomplete information is common.

Despite their merits, RBMs present certain challenges that have to be taken into account. Notably, the training of RBMs can be computationally demanding, particularly with large datasets and deep architectures. The iterative nature of the contrastive divergence algorithm, commonly used for training RBMs, can lead to slow convergence, making the training process resource-intensive. Efficiently addressing these computational demands remains a focal point to enhance the practicality of RBMs in real-world applications.

Another challenge lies in the sensitivity of RBMs to hyperparameters. Selecting appropriate hyperparameter values is a non-trivial task, and suboptimal choices may result in poor model

performance or hinder convergence during training. Balancing the learning rate, the number of hidden units, and other hyperparameters is crucial, and a lack of clear guidelines adds complexity to the model development process. While RBMs offer significant advantages with respect to traditional BMs, addressing the mentioned challenges is an important step to maximize their performance in various applications.

3 Quantum Boltzmann Machines

The motivation for integrating quantum mechanics into machine learning stems from the recognition that certain computational problems, especially those involving complex optimization and probabilistic modeling, can be addressed more efficiently using quantum principles. Quantum mechanics enables the representation of information through new concepts like state superpositions and quantum entanglement. In machine learning, in turn, tasks such as optimization, matrix operations and sampling distributions often involve computationally hard problems. Quantum computing has the potential to provide speedup in these areas.

3.1 QBM Theory

In the classical RBM, the visible and hidden layers consist of classical bits (1s and 0s). If we were to change this to the fundamental unit of quantum information (qubits or quantum bits), we could now exploit the more powerful mathematical tools of quantum mechanics (Dirac 1930). Quantum mechanics employs matrices (operators) whose dimensionality corresponds to the total number of potential states (2^N), unlike conventional machine learning methods that utilize vectors with a dimensionality equivalent to the number of variables (N). The quantum analog of the energy function defined in Equation (2) is given by a $2^N \times 2^N$ matrix, called the *Hamiltonian* of the system:

$$H = -\sum_{i=1}^{N_v} \Gamma_i \sigma_{v_i}^x - \sum_{j=1}^{N_h} b_j \sigma_{h_j}^z - \sum_{i=1}^{N_v} \sum_{j=1}^{N_h} w_{ij} \sigma_{v_i}^z \sigma_{h_j}^z. \tag{6}$$

Here, Γ_i, b_j, w_{ij} represents weights similar to the ones in classical RBMs. In this case the dimensionality of the system N is given by the sum of the numbers of visible and hidden nodes: $N = N_v + N_h$. The binary values v_i and h_j of the visible and hidden nodes in the classical RBM are here supplanted by the $2^N \times 2^N$ matrices $\sigma_{v_i}^x$, $\sigma_{v_i}^z$ and $\sigma_{h_j}^z$, defined as:

$$\sigma_{v_i}^x = \overbrace{I \otimes \ldots \otimes I}^{v_i - 1} \otimes \sigma_x \otimes \overbrace{I \otimes \ldots \otimes I}^{N_v - v_i} \otimes \overbrace{I \otimes \ldots \otimes I}^{N_h},$$

$$\sigma_{v_i}^z = \overbrace{I \otimes \ldots \otimes I}^{v_i - 1} \otimes \sigma_z \otimes \overbrace{I \otimes \ldots \otimes I}^{N_v - v_i} \otimes \overbrace{I \otimes \ldots \otimes I}^{N_h},$$

$$\sigma_{h_j}^z = \overbrace{I \otimes \ldots \otimes I}^{N_v} \otimes \overbrace{I \otimes \ldots \otimes I}^{h_j - 1} \otimes \sigma_z \otimes \overbrace{I \otimes \ldots \otimes I}^{N_h - h_j}.$$

Here I is the identity matrix, and σ_x and σ_z are the Pauli-X and Pauli-Z matrices.[1] The eigenstates of the Hamiltonian (6) can be represented as linear combinations of the states $|v,h\rangle$, where v and h denote the combinations of eigenvalues of the matrices $\sigma^z_{v_i}$ and $\sigma^z_{h_j}$. Like in the RBM case we define the associated probability distribution as:

$$\rho = \frac{1}{Z} e^{-H}, \tag{7}$$

where ρ is the $2^N \times 2^N$ matrix obtained by matrix exponentiation of H, and is called *density matrix*.[2] The normalization constant Z in (7) is given by $Z = \mathrm{Tr}[e^{-H}]$, the matrix trace of the exponential e^{-H}. The diagonal elements of ρ are the Boltzmann probabilities of the 2^N eigenstates of the Hamiltonian (6). From the density matrix ρ we can obtain the marginal Boltzmann probabilities P_v for a given set of visible states v_i as:

$$P_v = \mathrm{Tr}[\Lambda_v \rho],$$

where Λ_v is the projection operator reducing ρ to the subspace specified by the visible state v. Using the probabilities just defined we can obtain the log-likelihood, which for a given data distribution $p_{data}(v)$ and parameters $\Theta = \{\Gamma_i, b_j, w_{ij}\}$ can be written as:

$$\ell(\Theta) = \sum_{\{v\}} p_{data}(v) \log \mathrm{Tr}[\Lambda_v \rho\,(\Theta)], \tag{8}$$

where $\sum_{\{v\}}$ denotes the sum over all possible configurations of the eigenvalues of $\sigma^z_{v_i}$.

3.2 Training a QBM

To train a QBM one would have to maximize the log likelihood function, which is achieved by following the gradient of the function as for the classical RBM (Amin et al. 2018):

$$\partial_\theta \ell\,(\Theta) = \sum_{\{v\}} p_{data}(v) \left(\frac{\mathrm{Tr}[\Lambda_v\, \partial_\theta e^{-H}]}{\mathrm{Tr}[\Lambda_v\, e^{-H}]} - \frac{\mathrm{Tr}[\partial_\theta\, e^{-H}]}{\mathrm{Tr}[e^{-H}]} \right). \tag{9}$$

with $\partial_\theta \ell = \frac{\partial \ell}{\partial \theta}$. Like in the classical RBM, the goal is to estimate the gradient efficiently by sampling. For QBMs this is non trivial because Λ_v and H are matrices, and they don't commute with each other:

$$\Lambda_v H \neq H \Lambda_v.$$

The second term on the right hand side of Equation (9) can be simplified to:

$$\frac{\mathrm{Tr}[\partial_\theta e^{-H}]}{\mathrm{Tr}[e^{-H}]} = -\mathrm{Tr}\left[\rho\, \partial_\theta H\right],$$

which can be computed by sampling as in the classical case. The first term in Equation (9) however is equal to:

$$\frac{\mathrm{Tr}[\Lambda_v \partial_\theta e^{-H}]}{\mathrm{Tr}[\Lambda_v e^{-H}]} = \int_0^1 \partial t \frac{\mathrm{Tr}[\Lambda_v e^{-tH} \partial_\theta H e^{-(1-t)H}]}{\mathrm{Tr}[\Lambda_v e^{-H}]},$$

[1] The use of the symbol σ for the Pauli matrices is conventional in quantum mechanics, and it is not to be confused with the activation function introduced in the previous section.

[2] We can define matrix exponentiation through Taylor expansion $e^{-H} = \sum_{k=0}^{\infty} \frac{1}{k!}(-H)^k$. For a diagonal Hamiltonian, e^{-H} is a diagonal matrix with its 2^N diagonal elements corresponding to the energy states e^{-E_z}.

and cannot be computed as efficiently. This makes training a QBM more computationally expensive than training an RBM. A solution to this problem consists in introducing an appropriate lower bound to the log-likelihood function, and maximizing it instead of the original function (see Amin et al. (2018) for more details).

3.2.1 Bound-Based QBM

We can define a lower bound for the probabilities using the Golden-Thompson inequality (Golden 1965; Thompson 1965):

$$\mathrm{Tr}[e^A e^B] \geqslant \mathrm{Tr}[e^{A+B}],$$

which holds for any Hermitian matrices A and B, as is the case in this application (obviously the inequality reduces to an identity when A and B commute). We can therefore introduce a lower bound to the log-likelihood function (8) as:

$$\ell(\Theta) = \sum_{\{v\}} p_{data}(v) \log \mathrm{Tr}[\Lambda_v \rho] \geqslant \sum_{\{v\}} p_{data}(v) \log \frac{\mathrm{Tr}[e^{\log \Lambda_v - H}]}{\mathrm{Tr}[e^{-H}]}. \tag{10}$$

Introducing the new Hamiltonian $H_v = H - \log \Lambda_v$, we can write the right-hand side of Equation (10) as:

$$\tilde{\ell}(\Theta) = \sum_{\{v\}} p_{data}(v) \log \frac{\mathrm{Tr}[e^{-H_v}]}{\mathrm{Tr}[e^{-H}]}.$$

This expression is called *clamped* log-likelihood function, because the form of the Hamiltonian H_v implies a null probability whenever the visible nodes configuration is different from v. When calculating the gradient of this function, we get:

$$\partial_\theta \tilde{\ell}(\Theta) = -\sum_{\{v\}} p_{data}(v) \left(\frac{\mathrm{Tr}[e^{-H_v} \partial_\theta H_v]}{\mathrm{Tr}[e^{-H_v}]} - \frac{\mathrm{Tr}[e^{-H} \partial_\theta H]}{\mathrm{Tr}[e^{-H}]} \right),$$

which in turn, defining the clamped density matrix $\rho_v = e^{-H_v} / \mathrm{Tr}\left[e^{-H_v}\right]$, can be simplified as:

$$\partial_\theta \tilde{\ell}(\Theta) = \mathrm{Tr}\left[\rho\, \partial_\theta H\right] - \sum_{\{v\}} p_{data}(v) \mathrm{Tr}\left[\rho_v\, \partial_\theta H_v\right].$$

The explicit gradients of $\tilde{\ell}(\Theta)$ with respect to the function parameters are given by:

$$\partial_{\Gamma_i} \tilde{\ell}(\Theta) = \sum_{\{v\}} p_{data}(v) \mathrm{Tr}\left[\rho_v\, \sigma^x_{v_i}\right] - \mathrm{Tr}\left[\rho\, \sigma^x_{v_i}\right]. \tag{11}$$

$$\partial_{b_j} \tilde{\ell}(\Theta) = \sum_{\{v\}} p_{data}(v) \mathrm{Tr}\left[\rho_v\, \sigma^z_{h_j}\right] - \mathrm{Tr}\left[\rho\, \sigma^z_{h_j}\right]. \tag{12}$$

$$\partial_{w_{ih}} \tilde{\ell}(\Theta) = \sum_{\{v\}} p_{data}(v) \mathrm{Tr}\left[\rho_v\, \sigma^z_{v_i}\, \sigma^z_{h_j}\right] - \mathrm{Tr}\left[\rho\, \sigma^z_{v_i}\, \sigma^z_{h_j}\right]. \tag{13}$$

Equations (12) and (13) can be easily estimated by sampling from the probability distributions obtained from the density matrices ρ and ρ_v, and these estimations can be used to evaluate the

related components of the function $\tilde{\ell}(\Theta)$. The results obtained from this procedure are shown to be compatible with the evaluation of the full gradient (9) in controllable situations, thus we obtained a viable algorithm to maximize the log-likelihood with respect to the two sets of parameters b_j and w_{ij} (see Amin et al. 2018). The evaluation of the simplified Equation (11), however, does not approximate the associated components of the true gradient. Following the gradient defined with Equation (11) leads to parameters configurations in which $\Gamma_i = 0$ for all i, which is an artifact of the lower-bound maximization. For this reason, the values of Γ_i are usually treated as hyperparameters, or trained maximizing the full log-likelihood function. QBMs are trained on machines called *quantum annealers*, specialized quantum computers designed to solve optimization problems by finding the minimum energy state of a system, typically mapped to a Hamiltonian of the form (6). They operate through quantum annealing, a process that leverages quantum tunneling to explore the energy landscape of the system (Falco et al. 1988; Apolloni et al. 1989; Kadowaki and Nishimori 1998). We use a D-Wave quantum annealer (Boothby et al. 2019; Boothby et al. 2021). Finally, once the training procedure is completed, the QBM can be used similarly to the RBM to generate new data.

4 Classical Benchmark

As explained in the introduction, suitable applications for demonstrating the value of QBMs arise in settings where data are limited, such as the rare events in the tails of a distribution. A motivation for the focus on this setting is the widespread, yet often inappropriate, assumption of normality. If the true data-generating process exhibits heavy tails then relying on normality can lead to a severe underestimation of the likelihood and impact of extreme events. Tail events, however, are rare. As a result, it is often difficult to gather enough data to reject the assumption of normality and to accurately describe the sporadic, large outliers in the distribution. Such situations frequently arise in macroeconomic contexts, where variables like GDP are reported infrequently, and they can lead to high estimation uncertainty. These considerations motivate our empirical focus on heavy-tailed distributions, where QBM-generated synthetic observations can help improve the robustness of inference.

A distribution $\rho_X(x)$ is said to have a (right) heavy tail if it decreases more slowly than any exponential. Defining the complementary cumulative distribution function as:

$$F_X(x) = \int_x^{\infty} \rho_X(y) dy.$$

We say that the distribution is heavy-tailed if:

$$lim_{x\to\infty} F_X(x)\, e^{tx} = \infty \quad \forall t > 0.$$

For a large class of heavy-tailed distributions only a finite number of moments are bounded. This is true in particular for all distributions having a right tail decaying as a power law:

$$\rho_X(x) \sim x^{-(\alpha+1)} \quad , \quad \alpha > 0.$$

In this case, the largest integer smaller than α indicates the maximum finite moment of the distribution.

In this section, we directly compare the estimation performance of QBM and RBM models when applied to a common heavy-tailed distribution. We focus on the Student-t distribution with 5 degrees of freedom, for which only the first four moments exist. For finite sample sizes, this implies that the estimation of higher-order moments becomes unreliable. Moreover, although the fourth moment is theoretically finite, its estimation can exhibit substantial uncertainty in most realistic sample sizes, as we will demonstrate.

The comparison is performed on small Student-t samples of 100 observations. For each sample of 100 observations, we first compute the first four moments of the distribution, denoted in Table 1 as x, x^2, x^3, and x^4. The same sample is then used to train both a RBM and a QBM. Once trained, each model is used to generate a synthetic dataset consisting of 10,000 observations. These synthetic samples are then used to re-estimate the first four moments of the distribution. The estimates obtained from the QBM- and RBM-generated data are compared against the original sample estimates as well as the true theoretical moments of the distribution. This repeated sampling procedure, performed over 300 independent samples, allows us to evaluate the consistency, bias, and variability of each method in recovering key distributional features from limited data.[3] In particular, the setup highlights how the accuracy of moment estimation deteriorates as the order of the moment increases, and how generative models can help mitigate this issue under small-sample conditions.

The choice of an RBM as the classical benchmark is straightforward, as the QBM is the immediate quantum extension of the RBM. Both RBM and QBM models can be used to model distributions on finite sets. As such, they need a regularization of the distribution domain: we have to select a finite range of variation of the analyzed data, and discretize the domain. While the latter operation has negligible effects on the resulting distribution, the clipping of the range of variation can have sizable results for a heavy-tailed distribution as a Student-t. To minimize the effects of this choice, we select the range of variation $[-10, 10]$, much larger than the typical domain for our selected sample size. With this selected range, the probability for a sample point to fall outside of this interval is less than 2×10^{-4}.

Figure 2 shows an example of a simulated and augmented sample. With these cutoffs, the Student-t distribution with 5 degrees of freedom has true values of zero for the first and third moments. The true value of the second moment is equal[4] to 1.64 and the fourth moment equal to 15.96. For samples composed of 10.000 data points, the standard deviation of the second moment is 0.04, while the standard deviation of the fourth moment is 1.52.

The results for the sample data, RBM model and QBM model are shown in Table 1. As we anticipated in the beginning of the section, while the first and second moments of the distribution are reasonably stable for the original sample size of 100 points, the oscillations in the fourth moment are of the same order of magnitude of the true value itself. This uncertainty gets amplified when training the generative models. We find, however, a much better estimation using the QBM with respect to the RBM.

As we can see from the table, the estimation performances of the RBM model start degrading

[3] In order to make a meaningful comparison, we first optimize the hyper-parameters set for the RBM. We optimize over an independent set of 1000 samples, using as objective functions the L_1 and L_2 distances between the sample distribution and the synthetic one (both distances result in a similar estimation for the parameters).

[4] These values are obviously lower than for the standard Student-t (5) distribution, given the artificial cutoff on the tail values.

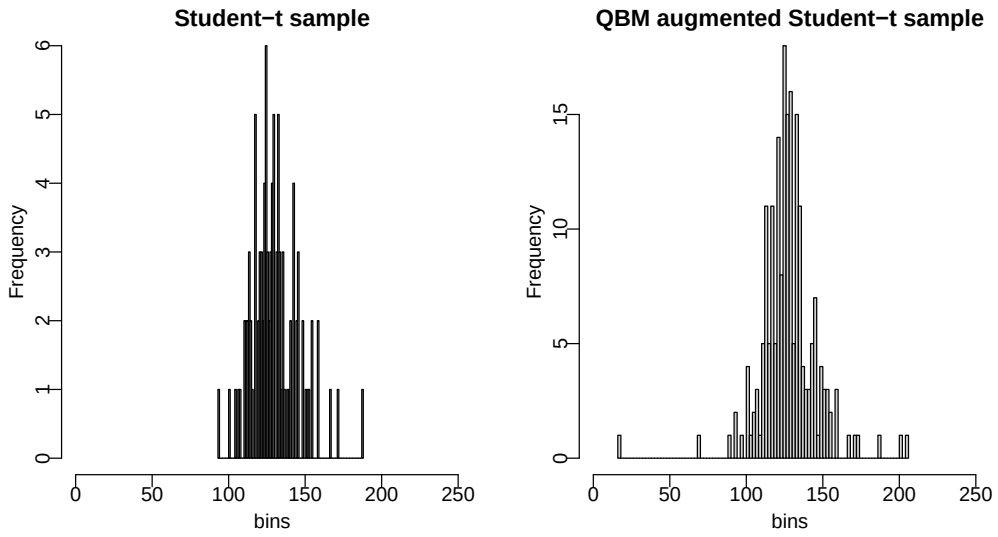

Figure 2: Augmenting Student-t (5) data

This figure shows an example of augmenting data sampled from a Student-t distribution with 5 degrees of freedom. The left panel shows the histogram of a draw of $n = 100$ observations. We have discretized the data by creating equally sized bins of size $(10+10)/(2^8)$, as the support is limited between -10 and 10. After training the QBM on this sample, we generate 100 additional QBM simulated observations and add them to the original data, presented in the right panel.

Table 1: Simulation results

	Sample		RBM			QBM		
Estimator	Average	St. dev.	Average	St. dev.	Corr	Average	St. dev.	Corr
x	0.01	0.14	-0.21	0.45	0.21	-0.08	0.11	-0.58
x^2	1.62	0.33	13.83	2.51	0.16	3.99	0.44	0.23
x^3	0.19	2.23	-12.54	26.66	0.65	-0.02	1.64	-0.68
x^4	15.63	13.42	745.27	169.23	0.15	188.28	23.07	-0.02

This table display the first four moments of the empirical distribution generated from the DGP, the RBM model and QBM model. We report the mean and standard deviation for the moments over the 300 samples drawn from the Student-t (5) distribution. Correlations between the terms $\varepsilon^{\text{sample}}_{\text{true}}$ and $\varepsilon^{\text{synt}}_{\text{sample}}$ for RBM and QBM data.

already from the second moment. For the first moment we obtain an average value which is of the same order of magnitude of the sample standard deviation, which implies we would not have a significantly worse estimation using the RBM generated data. However, for the second, third and fourth moment we obtain averages and standard deviations which are orders of magnitudes larger than the sample (and QBM) ones. This implies that, in trying to augment the sample data with synthetic RBM generated ones, we would introduce significant distortions in the data, and alter the statistical properties of the original distribution.

For the QBM-generated data, the results are more encouraging. The first moment exhibits only a small bias, with a standard deviation even lower than that of the sample-based estimate. The second moment shows a slight positive bias; however, both its average value and standard deviation are of the same order of magnitude as those of the sample. The estimation of the third moment also improves relative to the sample data, both in terms of average value and

dispersion. This clear enhancement can be attributed to two key factors. First, the synthetic data are generated starting from a symmetric distribution over a symmetric interval. Second, this symmetry is not broken by the random fluctuations in the sample. As a result, with a sufficiently large number of synthetic observations, the average of any odd moment tends to zero. This property would not hold if the QBM model were overfitting to the outlier values in the sample, as we observe with the RBM generated data.

In this regard, the QBM model performs particularly well: it efficiently downweights the influence of outlier observations and more reliably captures the core structure of the distribution.[5] For the fourth moment, we begin to observe a larger bias, although the standard deviation of the QBM-based estimate remains roughly comparable to that of the sample-based estimate. Overall, this comparison suggests that when synthetic data are used to augment a small sample, the QBM model offers a significantly more accurate replication of the underlying distribution's statistical properties. In contrast, the RBM tends to introduce distortions that may be unacceptable, even when compensated by a much larger number of generated data points.

In addition to the absolute performance of the estimations, it is instructive to consider the correlation between the errors of the sample estimators and the synthetic ones. More precisely, as shown in Appendix B, let $\varepsilon_{\text{true}}^{\text{sample}}$ denote the errors between the sample estimators and the true values, and let $\varepsilon_{\text{sample}}^{\text{synt}}$ denote the errors between the synthetic estimators and the sample ones. The sign of the correlation between $\varepsilon_{\text{true}}^{\text{sample}}$ and $\varepsilon_{\text{sample}}^{\text{synt}}$ directly impacts the quality of the estimations of the true values based on synthetic data.

As already mentioned, one of the reasons the QBM estimators are much closer to the real value than the RBM ones, is that the RBM shows the tendency to learn the random fluctuations in the sample. This translates in a positive correlation between $\varepsilon_{\text{true}}^{\text{sample}}$ and $\varepsilon_{\text{sample}}^{\text{synt}}$. The QBM estimators, on the other hand, manage to neglect the fluctuations more often, generating a *negative* correlation between the two error terms.

As shown in the last column of Table 1, for both the first and third moment we obtain a large negative correlation of the error terms between sample and true data, and synthetic and sample data. To understand why this happens, it is sufficient to realize that whenever a sample presents, for example, a positive large outlier, the error term $\varepsilon_{\text{true}}^{\text{sample}}$ will be positive. However, if the generative model does not learn to replicate this outlier, the error $\varepsilon_{\text{sample}}^{\text{synt}}$ between the synthetic estimation and the sample one will typically be negative. As shown in the fifth column of Table 1, the RBM estimations exhibit positive correlations in the error terms across all distribution moments. This suggests that when an outlier appears in the sample, it is likely to be replicated in the synthetic data as well.

5 Real Data: Risk Assessment of Young Firms

The risk assessment of financial investments forms an important part of investors' capital allocation decision and the financial stability regulatory framework. Value-at-Risk (VaR) has

[5] As a robustness check, we repeated the same experiment after introducing a random shift in the center of the distribution to remove the advantage conferred by symmetry. The results, presented in Appendix A, confirm the findings reported here.

become a cornerstone of both financial risk management and regulatory frameworks. It provides a probabilistic estimate of the potential loss in value of a portfolio over a defined time horizon at a given confidence level, offering a standardized approach to assessing market risk (Jorion 2007). Regulatory bodies such as the Basel Committee on Banking Supervision have incorporated VaR, and later Expected Shortfall (ES) (Artzner et al. 1999), into capital adequacy requirements, requiring banks to hold capital against potential losses implied by their internal risk models (Basel Committee on Banking Supervision 2011).

Estimating these types of measures accurately becomes particularly challenging in the presence of small samples, which are common when working with low-frequency data, short-lived securities, or emerging asset classes. Small samples exacerbate the inherent trade-off between bias and variance in risk estimation and limit the ability to reliably infer tail behavior, to which VaR and ES are especially sensitive (Danielsson and Vries 1997). Parametric approaches, such as those assuming normality or GARCH-based dynamics, may suffer from model misspecification, while non-parametric methods like historical simulation become unreliable due to insufficient observations in the tail (Pritsker 2006).

To address this, we leverage the QBM to generate synthetic data that augments the existing limited observations. Specifically, we train the QBM using the first year of returns data from the stock exchange and use it to generate additional synthetic returns. By augmenting the original dataset with these QBM-synthesized observations, we aim to improve the accuracy of various risk measures.

5.1 Stock Return Data

The stock market data used in this study is obtained from the Centre for Research in Security Prices (CRSP). The CRSP database provides individual stock data spanning December 31, 1925, to December 31, 2015, and includes information from the NYSE, AMEX, NASDAQ, and NYSE Arca exchanges. The prices adjusted for stock splits and dividend payouts are used to then calculate the returns. We use these returns to calculate various risk metrics from the risk management literature. We exclude stocks with fewer than 60 months of data. Stocks with an average price below $5 are also excluded as is customary in the finance literature. Furthermore, due to computational limitations we have restricted our sample to the 400 stocks with the lowest unique security identifier (PERMNO).[6]

5.2 Risk Assessment

The risk measures we employ characterize the return distribution in terms of both overall uncertainty and exposure to undesirable outcomes. In financial risk management, variance and kurtosis are commonly used to quantify uncertainty, while downside risk is typically assessed using skewness, VaR, ES, and the tail index for the left tail of the return distribution. For formal definitions of these risk measures, see Appendix C; for further discussion, refer to Danielsson (2011).

[6]We disregard six stocks with more than 10 zero return days in the first 200 days (stale prices).

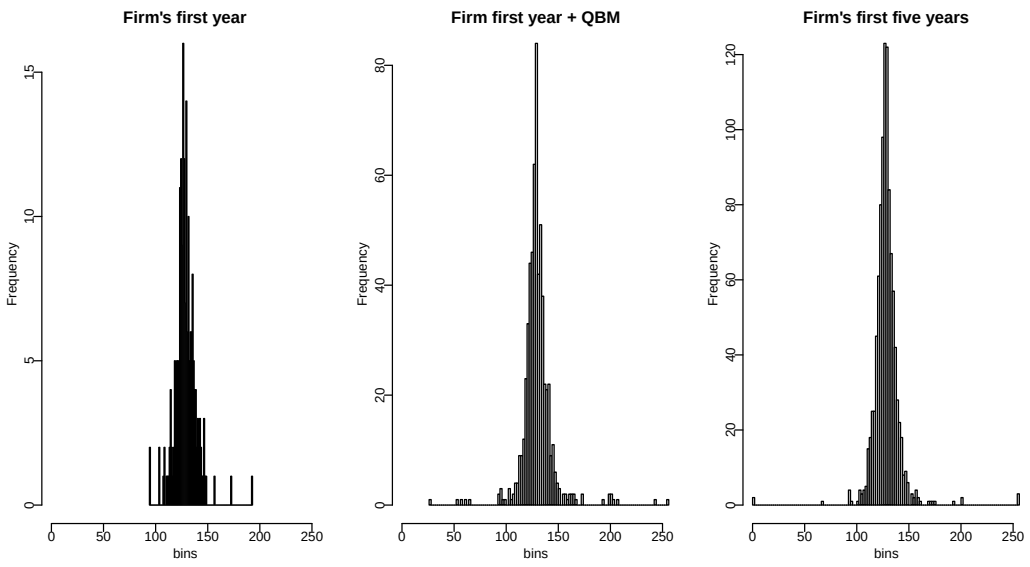

Figure 3: Augmenting 1 year firm data with QBM data

This figure shows an example of augmenting a firms first year of data with QBM synthesized data. The left panel shows the histogram of the first year of data. The centre panel shows the first year firm data augmented with 400 QBM synthesized data. The panel to the right shows the five year sample data. We have discretized the data by creating equally sized bins of size $(0.5+0.5)/(2^8)$, as the support is limited between -50% and 50% returns.

These measures form the basis of our empirical framework, which evaluates whether QBM-generated data improve the ability to explain a firm's longer sample (five year) risk profile. We focus on the five-year horizon because it incorporates more data than short-term estimates, yielding more stable and reliable risk estimate. At the same time, it remains short enough to avoid capturing structural shifts in the firm's fundamentals that could occur over longer horizons. A graphical representation can be seen in Figure 3. To assess the predictive contribution of QBM, we compare its performance against a RBM. This comparison is particularly relevant from a quantum computing perspective, as it illustrates the potential for quantum generative models to improve risk estimation over classical models, especially in data-constrained environments where traditional methods may underperform.

Figure 4 provides a graphical representation of our analysis for the VaR measured at 5%. The upper panel shows a scatter plot of the VaR of the firms extracted from the first year, VaR_i^{1y}, versus the VaR from the five year sample for the same firm, VaR_i^{1-5y}. The lower panel shows the scatter plot of the same 5 year sample on the y-axis, against the difference between the measure extracted from the 1 year sample augmented with the QBM data and the VaR with the original one year sample, $\Delta\text{VaR}_i^{1y,QBM}$. The difference captures the added value of the QBM augmented data. The positive relationship in the lower panel shows that the $\Delta\text{VaR}_i^{1y,QBM}$ helps to better predict the larger sample risk measure.

To further analyze this relationship we setup the following regression framework:

$$\text{Risk}_i^{1-5y} = c + \beta_1 \text{Risk}_i^{1y} + \beta_2 \Delta\text{Risk}_i^{1y,QBM} + \varepsilon_i. \tag{14}$$

Here, Risk_i^{1-5y} denotes a given risk measure for firm i, calculated over the first five years of its returns. The term Risk_i^{1y} reflects the same measure based only on the first year of data. Furthermore, $\Delta\text{Risk}_i^{1y,QBM}$ captures the difference between the risk measure estimated from

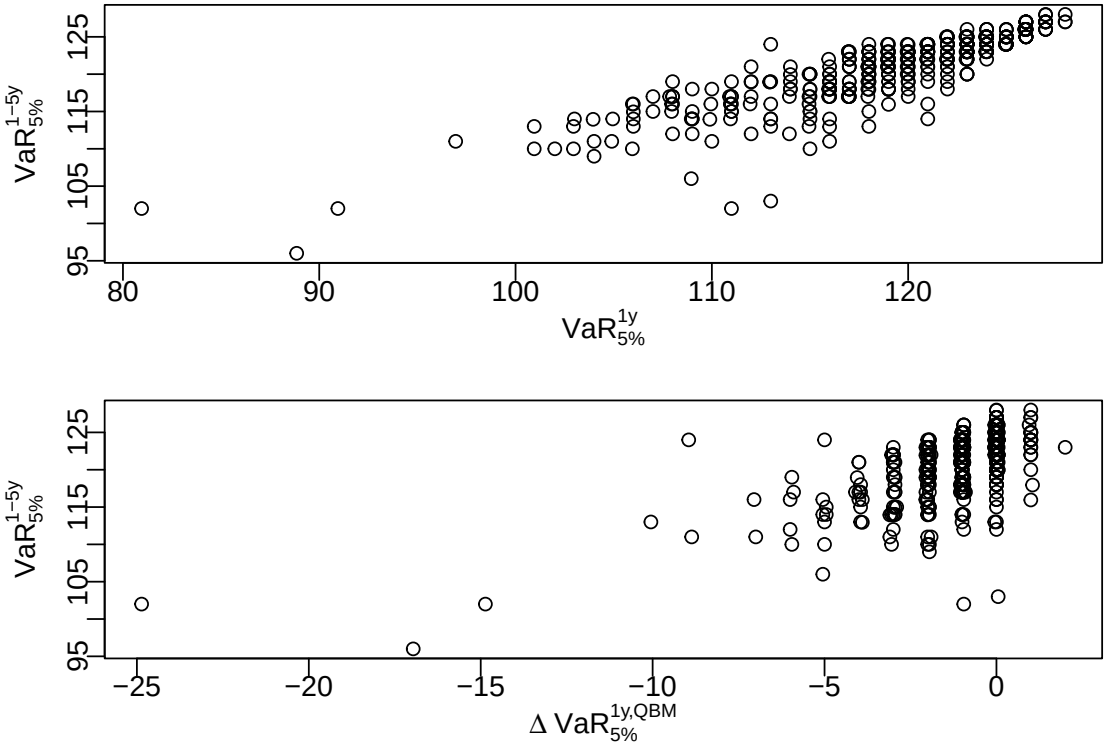

Figure 4: Value-at-Risk prediction - Sample and added value augmentation

The upper panel of this figure shows the scatter plot of the VaR_i^{1y} measurement from the first year sample period against the VaR_i^{1-5y} measures for 394 US stocks. The lower panel shows the scatter plot of $\Delta\text{VaR}_i^{1y,QBM}$ against VaR_i^{1-5y}. We use the CRSP database to obtain time series return data for these firms. To train the QBM model we first need to discretize the data, by creating 256 equally sized buckets between returns between -50% and +50%. Returns beyond this range are winsorized to the lowest or highest bin. We disregard stocks with more than 10 zero return days in the first 200 days (stale prices). For the augmentation we add 400 synthesized observations to the original 200 sample observations.

QBM-augmented data and that from the original one-year sample. Our primary interest lies in the estimate for the coefficient β_2, where a positive significant value indicates whether the additional information provided by QBM-generated observations improves the prediction of long sample risk.

Table 2 summarizes the results of the regression analysis, showing that the predictability of risk measures varies slightly across different metrics.[7] The first row highlights that using risk measures extracted from the one year ordinary sample significantly predicts their five-year counterparts in the expected positive direction. In the second row, we report the coefficient estimates of our variables of interest, the information added by the QBM synthesized data. The two uncertainty measures, standard deviation and kurtosis, benefit from the augmentation of QBM-synthesized data, as indicated by their positive and significant coefficients. The slightly lower significance level for the standard deviation is intuitive; since the standard deviation is

[7]Tables 4 and 5 in Appendix D present the results of augmenting the data with 200 and 600 synthesized observations, respectively. The results are broadly similar, with the exception of the tail index estimate. For the sample with 200 synthesized observations, the estimate is in the expected direction but statistically insignificant. In contrast, with 600 synthesized observations, the estimate becomes highly significant, highlighting the need for a larger synthesized sample to reliably estimate this data-intensive metric.

Table 2: Prediction large sample risk with small sample risk measures

	QBM					
	st. dev.	$VaR_{5\%}$	$ES_{5\%}$	skew	kurtosis	tail index
Risk^{1y}	0.86***	0.72***	0.81***	0.67***	2.86***	0.60***
	(0.02)	(0.03)	(0.03)	(0.08)	(0.28)	(0.03)
$\Delta\text{Risk}^{1y,\text{QBM}}$	0.13**	0.38***	0.07**	−0.04	0.34***	0.05**
	(0.05)	(0.06)	(0.03)	(0.03)	(0.09)	(0.02)
Constant	−0.89	34.48***	24.84***	0.06	−16.74***	0.01**
	(0.55)	(3.52)	(3.07)	(0.07)	(5.43)	(0.002)
R^2	0.77	0.77	0.73	0.17	0.21	0.51
	RBM					
Risk^{1y}	0.82***	0.82***	0.78***	0.69***	2.70***	0.60***
	(0.02)	(0.03)	(0.03)	(0.10)	(0.75)	(0.03)
$\Delta\text{Risk}^{1y,\text{RBM}}$	−0.02	−0.01*	−0.02**	0.0004	0.28	0.01
	(0.01)	(0.005)	(0.01)	(0.09)	(0.63)	(0.01)
Constant	1.01**	21.81***	24.28***	0.11	−1.65	0.01***
	(0.39)	(3.00)	(3.02)	(0.07)	(7.30)	(0.002)
R^2	0.77	0.74	0.73	0.17	0.18	0.50
Observations	394	394	391	394	394	394

This table displays the results of the regression analysis of Equation (14) to asses the increased accuracy of risk measures for young firms (evaluated with data from approximately one year after the initial public offering). We use the CRSP database to obtain time series return data for these firms. The independent variables for each column are the risk metrics calculated for the first year of a firms listing (Risk^{1y}). The second risk measure ($\Delta\text{Risk}^{1y,\text{QBM}}$) is the difference between $Risk^{1y}$ and the risk measure extracted from the sample data augmented with the QBM synthesized data. The dependent variable is the same metric on an extended sample of the first 5 years after a firms listing. The second panel displays the results of a similar excercise, with the difference being that the synthesized data comes from the RBM. For the analysis we discretize the data by creating 256 equally sized buckets between returns between -50% and +50%. We disregard stocks with more than 10 zero return days in the first 200 days (stale prices). We add 400 synthesized observations to the original 200 sample observations for the augmented sample. The stars that indicate the significance level are *p<0.1; **p<0.05; ***p<0.01.

less sensitive to tail observations, it is expected to be well estimated even in relatively small samples. In contrast, kurtosis, which captures the fourth moment of the distribution, is highly sensitive to tail observations and is therefore expected to benefit the most from additional data.

The measures focused on characterizing left-tail behavior, excluding skewness, are positive and statistically significant in predicting the large-sample risk measures.[8] While the short-sample estimates themselves are informative of the longer-sample counterparts, the same predictive property does not hold uniformly for the QBM-augmented estimates. However, metrics that rely exclusively on left-tail data do improve with QBM-based augmentation. This is expected, as these measures are particularly sensitive to rare observations and thus benefit the most from an expanded sample.

The variation in significance across the 5% VaR, ES, and the left-tail index can be partly explained by the limitations of Boltzmann Machines. Extending the distribution beyond the range observed in the original data, without imposing strong parametric assumptions, is inherently challenging. Among these three measures, the 5% VaR is the least reliant on data beyond the observed range, making it more robust to synthetic augmentation. In contrast, both ES and the tail index depend more heavily on extrapolated tail behavior, which leads to lower significance levels compared to the 5% VaR.

In the lower panel, we contrast the QBM results with those from the RBM. The RBM struggles to add significant information to the one-year sample, with most coefficients being insignificant. While the VaR and ES are significant, they have the wrong sign. This aligns with the simulation results, where the QBM consistently outperforms its classical counterpart.

6 Conclusions

This paper explores the potential of Quantum Boltzmann Machines (QBMs) as generative models for improving risk estimation in data-scarce environments. Compared to classical Restricted Boltzmann Machines (RBMs), QBMs offer a more flexible framework that can better approximate complex, heavy-tailed distributions when observations are limited. Our results suggest that even in constrained settings, QBMs can add useful information to small samples, enhancing the estimation of risk measures that are typically difficult to assess with limited data.

We apply this framework to a financial setting, focusing on risk assessment for newly listed firms using return data from the CRSP database. By augmenting short return samples with QBM-generated data, we show that estimates of long-term risk measures, particularly those sensitive to tail behavior, such as value-at-risk, expected shortfall, and kurtosis, improve meaningfully. In contrast, RBM-augmented samples do not provide the same level of predictive enhancement. These findings suggest that QBMs can offer practical benefits in financial applications where robust risk estimation is needed despite limited historical data.

It is important to note that the current generation of quantum hardware still imposes limits on the training and scalability of QBMs. Nonetheless, as quantum computing technology contin-

[8] As an alternative, skewness is measured using the robust coefficient of asymmetry proposed by Hinkley (1975) and applied in Conrad et al. (2013). The results remain insignificant when asymmetry is measured using the 25% and 75% quantiles.

ues to evolve, the relative advantage of QBMs over classical models is expected to grow. This makes QBMs a potential tool for financial modeling, particularly in applications where data scarcity is a fundamental constraint. By leveraging quantum mechanical properties to better capture distributional complexity, QBMs represent a promising step in the direction of a more robust and accurate financial risk assessment.

Acknowledgement

We would like to thank Stenio Fernandes, Maryam Haghighi, Miguel Molico and Cameron Perot for their helpful comments. We also thank the seminar participants at the Bank of Canada and Quantum Computing Applications in Economics and Finance Conference 2025.

References

Ackley, D. H., G. E. Hinton, and T. J. Sejnowski (1985). "A learning algorithm for Boltzmann machines". In: *Cognitive science* 9.1, pp. 147–169.

Amin, M. H., E. Andriyash, J. Rolfe, B. Kulchytskyy, and R. Melko (May 2018). "Quantum Boltzmann Machine". In: *Physical Review X* 8.2.

Anschuetz, E. R. and Y. Cao (2019). "Realizing quantum boltzmann machines through eigenstate thermalization". In: *arXiv preprint arXiv:1903.01359*.

Apolloni, B., C. Carvalho, and D. De Falco (1989). "Quantum stochastic optimization". In: *Stochastic Processes and their Applications* 33.2, pp. 233–244.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). "Coherent measures of risk". In: *Mathematical Finance* 9.3, pp. 203–228.

Basel Committee on Banking Supervision (2011). *Basel III: A global regulatory framework for more resilient banks and banking systems*. https://www.bis.org/publ/bcbs189.htm.

Bhasin, N. K., S. Kadyan, K. Santosh, R. HP, R. Changala, and B. K. Bala (2024). "Enhancing Quantum Machine Learning Algorithms for Optimized Financial Portfolio Management". In: *2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*. IEEE, pp. 1–7.

Boothby, K., P. Bunyk, J. Raymond, and A. Roy (2019). *Next-Generation Topology of D-Wave Quantum Processors*. Tech. rep. https://www.dwavequantum.com/media/jwwj5z3z/14-1026a-c_next-generation-topology-of-dw-quantum-processors.pdf.

Boothby, K., A. D. King, and J. Raymond (2021). *Zephyr Topology of D-Wave Quantum Processors*. Tech. rep. https://www.dwavequantum.com/media/2uznec4s/14-1056a-a_zephyr_topology_of_d-wave_quantum_processors.pdf.

Buonaiuto, G., F. Gargiulo, G. De Pietro, et al. (2023). "Best practices for portfolio optimization by quantum computing, experimented on real quantum devices". In: *Scientific Reports* 13, p. 19434.

Chu, Y., X. Zhao, Y. Zou, W. Xu, and Y. Z. J Han (2018). "A Decoding Scheme for Incomplete Motor Imagery EEG With Deep Belief Network". In: *Frontiers in neuroscience* 12, p. 680.

Conrad, J., R. F. Dittmar, and E. Ghysels (2013). "Ex ante skewness and expected stock returns". In: *The Journal of Finance* 68.1, pp. 85–124.

Coyle, B., M. T. Henderson, J. C. J. Le, N. Kumar, M. Paini, and E. Kashefi (2021). "Quantum versus classical generative modelling in finance". In: *Quantum Science and Technology* 6 (2), p. 024013.

Dallaire-Demers, P. and N. Killoran (2018). "Quantum generative adversarial networks". In: *Physical Review A* 98 (1).

Danielsson, J. (2011). *Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab*. John Wiley & Sons.

Danielsson, J. and C. G. de Vries (1997). "Tail index and quantile estimation with very high frequency data". In: *Journal of Empirical Finance* 4.2–3, pp. 241–257.

Danielsson, J. and C. G. de Vries (2005). "Using a bootstrap method to choose the sample fraction in tail index estimation". In: *Journal of Multivariate Analysis* 94.2, pp. 295–314.

Decelle, A., B. Seoane, and L. Rosset (2023). "Unsupervised hierarchical clustering using the learning dynamics of restricted Boltzmann machines". In: *Physical Review E* 108.1, p. 014110.

Dirac, P. A. M. (1930). *The principles of quantum mechanics*. Oxford university press.

Du, Z., J. C. Escanciano, and X. Zhu (2018). "Forecasting value-at-risk: The role of conditional skewness and kurtosis". In: *Journal of Financial Econometrics* 16.3, pp. 424–458.

Duffie, D. and J. Pan (1997). "An overview of value at risk". In: *Journal of derivatives* 4.3, pp. 7–49.

Embrechts, P., C. Kluppelberg, and T. Mikosch (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, NJ: Princeton University Press.

Fahlman, S. E., G. E. Hinton, and T. J. Sejnowski (1983). "Massively Parallel Architecture for AI: NETL, Thistle, and Boltzmann Machines". In: *Proc. Natl Conf. AI*.

Falco, D. de, B. Apolloni, and N. Cesa-Bianchi (1988). "A numerical implementation of quantum annealing". In: *Stochastic Processes, Physics and Geometry* 324.

Gabaix, X. (2016). "Power laws in economics: An introduction". In: *Journal of Economic Perspectives* 30.1, pp. 185–206.

Ganguly, S. (2023). "Implementing quantum generative adversarial network (qgan) and qcbm in finance". In: *arXiv preprint arXiv:2308.08448*.

Gao, J., M. J. Kim, and I. Tsiakas (2022). "Bias-corrected expected shortfall estimation and backtesting". In: *Journal of Financial Econometrics* 20.2, pp. 313–341.

Golden, S. (Feb. 1965). "Lower Bounds for the Helmholtz Function". In: *Phys. Rev.* 137 (4B), B1127–B1128.

Herman, D., C. Googin, X. Liu, Y. Sun, A. Galda, I. Safro, M. Pistoia, and Y. Alexeev (July 2023). "Quantum computing for finance". In: *Nature Reviews Physics* 5.8, pp. 450–465.

Hill, B. M. (1975). "A simple general approach to inference about the tail of a distribution". In: *The Annals of Statistics* 3.5, pp. 1163–1174.

Hinkley, D. V. (1975). "On power transformations to symmetry". In: *Biometrika* 62.1, pp. 101–111.

Hinton, G. E. (2002). "Training products of experts by minimizing contrastive divergence". In: *Neural computation* 14.8, pp. 1771–1800.

Hinton, G. E. and R. R. Salakhutdinov (2006). "Reducing the dimensionality of data with neural networks". In: *science* 313.5786, pp. 504–507.

Hoga, Y. (2022). "Limit theory for forecasts of extreme distortion risk measures and expectiles". In: *Journal of Financial Econometrics* 20.1, pp. 18–44.

Hopfield, J. J. (1982). "Neural networks and physical systems with emergent collective computational abilities." In: *Proceedings of the national academy of sciences* 79.8, pp. 2554–2558.

Jorion, P. (2007). *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill.

Kadowaki, T. and H. Nishimori (1998). "Quantum annealing in the transverse Ising model". In: *Physical Review E* 58.5, p. 5355.

Kardar, M. (2007). *Statistical physics of particles*. Cambridge University Press.

Liu, J.-G. and L. Wang (2018). "Differentiable learning of quantum circuit born machines". In: *Physical Review A* 98.6, p. 062324.

Martins, L. F. and J. F. Ziegel (2021). "Estimating and evaluating conditional expected shortfall with reduced sample sizes". In: *Journal of Business & Economic Statistics* 39.1, pp. 260–273.

McNeil, A. J. and R. Frey (2000). "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach". In: *Journal of Empirical Finance* 7.3-4, pp. 271–300.

Orlandi, F., E. Barbierato, and A. Gatti (2024). "Enhancing Financial Time Series Prediction with Quantum-Enhanced Synthetic Data Generation: A Case Study on the S&P 500 Using a Quantum Wasserstein Generative Adversarial Network Approach with a Gradient Penalty". In: *Electronics* 13.11, p. 2158.

Patton, A. J., J. F. Ziegel, and R. Chen (2019). "Dynamic semiparametric models for expected shortfall (and Value-at-Risk)". In: *Journal of Econometrics* 211.2, pp. 388–413.

Pritsker, M. (2006). "The hidden dangers of historical simulation". In: *Journal of Banking & Finance* 30.2, pp. 561–582.

Ramzan, F., C. Sartori, S. Consoli, and D. Reforgiato Recupero (2024). "Generative adversarial networks for synthetic data generation in finance: Evaluating statistical similarities and quality assessment". In: *AI* 5.2, pp. 667–685.

Salakhutdinov, R., A. Mnih, and G. Hinton (2007). "Restricted Boltzmann machines for collaborative filtering". In: *Proceedings of the 24th international conference on Machine learning*, pp. 791–798.

Skavysh, V., S. Priazhkina, D. Guala, and T. R. Bromley (Aug. 2023). "Quantum monte carlo for economics: Stress testing and macroeconomic deep learning". In: *Journal of Economic Dynamics and Control* 153, p. 104680.

Smolensky, P. et al. (1986). "Information processing in dynamical systems: Foundations of harmony theory". In: *Department of Computer Science, University of Colorado, Boulder*.

Song, H., T. Song, Q. He, Y. Liu, and D. L. Zhou (2019). "Geometry and symmetry in the quantum boltzmann machine". In: *Physical Review A* 99 (4).

Stamatopoulos, N., D. J. Egger, Y. Sun, C. Zoufal, R. Iten, N. Shen, and S. Woerner (July 2020). "Option Pricing using Quantum Computers". In: *Quantum* 4, p. 291.

Stamatopoulos, N. and W. J. Zeng (Apr. 2024). "Derivative Pricing using Quantum Signal Processing". In: *Quantum* 8, p. 1322.

Stein, J., D. Schuman, M. Benkard, T. Holger, W. Sajko, M. Kölle, J. Nüßlein, L. Sünkel, O. Salomon, and C. Linnhoff-Popien (2024). "Exploring Unsupervised Anomaly Detection with Quantum Boltzmann Machines in Fraud Detection". In: *Proceedings of the 16th International Conference on Agents and Artificial Intelligence*. SCITEPRESS - Science and Technology Publications, pp. 177–185.

Thompson, C. J. (1965). "Inequality with Applications in Statistical Mechanics". In: *Journal of Mathematical Physics* 6, pp. 1812–1813.

Tüysüz, C., M. Demidik, L. Coopmans, E. Rinaldi, V. Croft, Y. Haddad, M. Rosenkranz, and K. Jansen (2024). "Learning to generate high-dimensional distributions with low-dimensional quantum Boltzmann machines". In: *arXiv preprint arXiv:2410.16363*.

Woerner, S. and D. J. Egger (2019). “Quantum risk analysis”. In: *npj Quantum Information* 5.1, p. 15.

Zhou, X., H. Zhao, Y. Cao, X. Fei, G. Liang, and J. Zhao (2024). “Carbon market risk estimation using quantum conditional generative adversarial network and amplitude estimation”. In: *Energy Conversion and Economics* 5.4, pp. 193–210.

Zhu, E. Y., S. Johri, D. Bacon, M. Esencan, J. Kim, M. Muir, N. Murgai, J. Nguyen, N. Pisenti, A. Schouela, K. Sosnova, and K. Wright (Nov. 2022). “Generative quantum learning of joint probability distribution functions”. In: *Physical Review Research* 4.4.

Appendix A Estimating non-symmetric distributions

As we described in Section 4, estimating a symmetric distribution on a symmetric interval can introduce spurious negative correlations between the errors made with the sample estimations of the true distributions and the synthetic estimations of the sample distributions. This in turn enhances the estimation of the moments of the true distributions with respect to the fully general setting.

For this reason, here we replicate the simulated estimation study of section 4 with a moving true distribution. We keep the estimation interval fixed at $[-10, 10]$, but add a random shift to the distribution center; in other words, in each simulation run the true distribution is given by a Student-t with 5 degrees of freedom for the variable $(x - s)$, with s a random variable uniformly distributed between $[-10, 10]$. For each simulation run, we first extract a value for s, then extract a large sample from a Student-t (5) distribution centered in $(x - s)$, and finally we keep only 100 of the sample points falling in the interval $[-10, 10]$. In this way we are able to generate a sample of true (and sample) distributions which are not centered in the domain of the estimated variables, in order to cancel the enhancing effects of symmetry in the estimation process. A drawback of this procedure is that the true values of the distribution moments change between the runs, hence the only significant measurements of the estimation performance are the error terms between QBM and RBM estimation on one side, and sample and true values on the other (as opposed to the situation for the fixed true distribution, in which it was meaningful to compare the averages and standard deviations of the estimated moments). In Table 3 we show the comparison results for the new estimation procedure.

Table 3: Comparison of RBM and QBM Moving Averages

	RBM with moving mean			QBM with moving mean		
	$\varepsilon^{\text{sample}}_{\text{true}}$	$\varepsilon^{\text{synt}}_{\text{true}}$	$\varepsilon^{\text{synt}}_{\text{sample}}$	$\varepsilon^{\text{sample}}_{\text{true}}$	$\varepsilon^{\text{synt}}_{\text{true}}$	$\varepsilon^{\text{synt}}_{\text{sample}}$
x	0.13	0.60	0.57	0.13	0.42	0.41
x^2	1.29	6.83	6.70	1.29	2.66	2.43
x^3	13.33	49.82	47.65	13.33	20.93	17.23
x^4	135.55	582.85	565.61	135.55	243.20	211.18

RBM to QBM comparison for moving true distributions. $\varepsilon^{\text{sample}}_{\text{true}}$ is the root mean squared error between the sample estimation of the moments and the true values (identical in the RBM and QBM cases). $\varepsilon^{\text{synt}}_{\text{true}}$ and $\varepsilon^{\text{synt}}_{\text{sample}}$ are, respectively, the root mean squared error between the synthetic estimations and the true values and the synthetic estimations and the sample ones. As can be seen from the table, most of the error in the estimations of the true values come from the inaccuracy of the synthetic estimation of the sample distributions.

The results of Section 4 are confirmed by this new procedure: the estimation errors with the QBM model are substantially lower than with the RBM model, and are of the same order of magnitude of the errors between the sample and true distributions. The correlations terms, while playing a smaller role in this setting, are still negative for the QBM estimation errors: they are approximately equal to -8% for the first, second and third moments, and closer to -7% for the fourth moment. The RBM estimation errors, on the other hand, present small positive correlation.

Appendix B Relevance of correlation terms

In Section 4 we described the importance of the correlation terms between the error terms $\varepsilon^{\text{sample}}_{\text{true}}$ and $\varepsilon^{\text{synt}}_{\text{sample}}$ in the performance of the estimation of the true distribution using synthetic data. In this appendix we want to derive the precise role these terms play in the estimation results. Our estimation of a statistics $\hat{\theta}$ consists of a two-step process. First, we obtain a sample, typically of small size, extracted from the true distribution. We use this sample to train a generative model, from which we extract an arbitrarily large number of new data points. Here we are interested in studying the relevant factors influencing the magnitude of the mean square error, $\text{MSE}(\hat{\theta})$, between the estimator evaluated on the synthetic data, and the true value. With a simple variance decomposition we obtain:

$$\begin{aligned}\text{MSE}(\hat{\theta}) &= \mathbb{E}[(\hat{\theta}_{\text{synt}} - \theta_{\text{true}})^2] \\ &= \mathbb{E}[(\varepsilon^{\text{synt}}_{\text{sample}})^2] + \mathbb{E}[(\varepsilon^{\text{sample}}_{\text{true}})^2] + 2\,\mathbb{E}[\varepsilon^{\text{synt}}_{\text{sample}}\,\varepsilon^{\text{sample}}_{\text{true}}]\end{aligned} \tag{15}$$

where

$$\varepsilon^{\text{synt}}_{\text{sample}} = (\hat{\theta}_{\text{synt}} - \hat{\theta}_{\text{sample}}),$$

$$\varepsilon^{\text{sample}}_{\text{true}} = (\hat{\theta}_{\text{sample}} - \theta_{\text{true}}),$$

θ_{true} is the true value of the estimator, $\hat{\theta}_{\text{sample}}$ and $\hat{\theta}_{\text{synt}}$ are the estimators evaluated with the different sets of data, and $\hat{\theta} = \hat{\theta}_{\text{synt}}$.

Whenever the covariance term in Equation (15) is null, we see that the best estimation is performed when the error between synthetic data and sample data is minimized, and the quality of the estimation is at best as good as the sample one. If, however, as we saw in Section 4 the covariance term is negative, we can obtain an estimation of the true value of θ_{true} better than the one available with sample data alone.

Appendix C Risk measures

In this section we shortly discuss the risk measures we extract from the financial data. Consider a series of return data $R_1, R_2, ..., R_n$. The sorted sample, i.e., order statistics, can be represented as

$$\max(R_1, ..., R_n) = R_{(n,n)} \geqslant R_{(n-1,n)} \geqslant ... \geqslant R_{(1,n)} = \min(R_1, ..., R_n).$$

From the order-statistics based on the returns $R_{(j,n)}$ we can define the following risk measures:

$$\text{Skewness} = \frac{E\left[(R - E(R))^3\right]}{E\left[(R - E(R))^2\right]^{3/2}} = \frac{\sum_{j=1}^{n}\left(R_{(j,n)} - \frac{1}{n}\sum_{k=1}^{n} R_{(k,n)}\right)^3}{\left(\sum_{j=1}^{n}\left(R_{(j,n)} - \frac{1}{n}\sum_{k=1}^{n} R_{(k,n)}\right)^2\right)^{3/2}}$$

and

$$\text{Kurtosis} = \frac{E\left[(R-E(R))^4\right]}{E\left[(R-E(R))^2\right]^2} = \frac{\sum_{j=1}^{n}\left(R_{(j,n)} - \frac{1}{n}\sum_{k=1}^{n} R_{(k,n)}\right)^4}{\left(\sum_{j=1}^{n}\left(R_{(j,n)} - \frac{1}{n}\sum_{k=1}^{n} R_{(k,n)}\right)^2\right)^2}.$$

In the financial risk literature and in many of the financial institutions regulatory frameworks the VaR forms an important assessment tool to gauge the level of risk. The VaR estimates the potential loss an investment portfolio may incur over a specific period, given a certain level of confidence.

$$VaR_p = -R_{([np],n)}$$

where $[np]$ is the integer part of np. Due to the need to better characterize the shape of the distribution beyond the VaR quantile, financial regulators also adopted the conditional tail expectation, or expected shortfall:

$$ES_p = -E[R|R < -VaR] = -\frac{1}{[np]}\sum_{j=1}^{[np]} R_{(j,n)}.$$

As stated in the main text, an important statistic to characterize the risk of a stock is the thickness of the tails of the distribution. The tail index α provides such a measure, as moments only exist up to α, i.e., $E \mid X_i \mid^p < \infty$ only for $p < \alpha$. A decrease in α gives a heavier tail. To estimate the tail index α, the most popular tool is the Hill estimator (Hill 1975):

$$\frac{1}{\widehat{\alpha}_k} = \frac{1}{k}\sum_{i=0}^{k-1}\left(\log\left(R^*_{n-i,n}\right) - \log\left(R^*_{n-k,n}\right)\right),$$

where k is the number of upper-order statistics used in the estimation of α. Furthermore, we are interested in the left tail and therefore $R^*_{(j,n)} = -R_{(j,n)}$.

Appendix D Tables

Table 4: Prediction large sample risk measures (**200** synthesized observations)

	QBM					
	st. dev.	$VaR_{5\%}$	$ES_{5\%}$	skew	kurtosis	tail index
Risk^{1y}	0.85^{***}	0.77^{***}	0.81^{***}	0.69^{***}	2.72^{***}	0.59^{***}
	(0.02)	(0.03)	(0.03)	(0.08)	(0.27)	(0.03)
$\Delta\text{Risk}^{1y,\text{QBM}}$	0.07^{**}	0.36^{***}	0.05^{**}	-0.01	0.24^{***}	0.02
	(0.03)	(0.07)	(0.02)	(0.02)	(0.06)	(0.02)
Constant	-0.09	29.29^{***}	23.60^{***}	0.10	-14.06^{***}	0.01^{***}
	(0.32)	(3.34)	(2.99)	(0.06)	(4.69)	(0.002)
R^2	0.77	0.76	0.72	0.17	0.21	0.51
	RBM					
Risk^{1y}	0.82^{***}	0.83^{***}	0.79^{***}	0.68^{***}	2.74^{***}	0.58^{***}
	(0.03)	(0.03)	(0.03)	(0.09)	(0.57)	(0.03)
$\Delta\text{Risk}^{1y,\text{RBM}}$	-0.02	-0.005	-0.02^{**}	-0.02	0.30	-0.01
	(0.01)	(0.004)	(0.01)	(0.07)	(0.44)	(0.01)
Constant	0.94^{**}	21.30^{***}	24.00^{***}	0.10	-2.88	0.01^{***}
	(0.38)	(2.98)	(2.99)	(0.07)	(6.74)	(0.002)
R^2	0.77	0.74	0.73	0.17	0.18	0.50

This table displays the results of the regression analysis of Equation (14) to asses the increased accuracy of risk measures for young firms (evaluated with data from approximately one year after the initial public offering). We use the CRSP database to obtain time series return data for these firms. The independent variables for each column are the risk metrics calculated for the first year of a firms listing (Risk^{1y}). The second risk measure ($\Delta\text{Risk}^{1y,\text{QBM}}$) is the difference between $Risk^{1y}$ and the risk measure extracted from the sample data augmented with the QBM synthesized data. The dependent variable is the same metric on an extended sample of the first 5 years after a firms listing. The second panel displays the results of a similar excercise, with the difference being that the synthesized data comes from the RBM. For the analysis we discretize the data by creating 256 equally sized buckets between returns between -50% and +50%. We disregard stocks with more than 10 zero return days in the first 200 days (stale prices). We add **200** synthesized observations to the original 200 sample observations for the augmented sample. The stars that indicate the significance level are *p<0.1; **p<0.05; ***p<0.01.

Table 5: Prediction large sample risk measures (**600** synthesized observations)

	QBM					
	st. dev.	$VaR_{5\%}$	$ES_{5\%}$	skew	kurtosis	tail index
Risk^{1y}	0.87^{***}	0.71^{***}	0.80^{***}	0.65^{***}	2.96^{***}	0.60^{***}
	(0.02)	(0.03)	(0.03)	(0.08)	(0.29)	(0.03)
$\Delta\text{Risk}^{1y,\text{QBM}}$	0.23^{***}	0.40^{***}	0.11^{***}	-0.07^{*}	0.41^{***}	0.07^{***}
	(0.06)	(0.06)	(0.03)	(0.04)	(0.09)	(0.02)
Constant	-1.90^{***}	36.22^{***}	26.15^{***}	0.01	-19.66^{***}	0.003
	(0.65)	(3.53)	(3.13)	(0.08)	(5.66)	(0.003)
R^2	0.77	0.77	0.73	0.17	0.21	0.51
	RBM					
Risk^{1y}	0.82^{***}	0.82^{***}	0.77^{***}	0.71^{***}	2.75^{***}	0.60^{***}
	(0.02)	(0.03)	(0.03)	(0.12)	(0.80)	(0.03)
$\Delta\text{Risk}^{1y,\text{RBM}}$	-0.02^{*}	-0.01^{*}	-0.03^{***}	0.03	0.33	0.005
	(0.01)	(0.005)	(0.01)	(0.11)	(0.69)	(0.01)
Constant	1.11^{***}	21.85^{***}	24.56^{***}	0.12	-1.87	0.01^{***}
	(0.40)	(3.00)	(3.01)	(0.07)	(7.20)	(0.002)
R^2	0.77	0.74	0.73	0.17	0.18	0.50

This table displays the results of the regression analysis of Equation (14) to asses the increased accuracy of risk measures for young firms (evaluated with data from approximately one year after the initial public offering). We use the CRSP database to obtain time series return data for these firms. The independent variables for each column are the risk metrics calculated for the first year of a firms listing (Risk^{1y}). The second risk measure ($\Delta\text{Risk}^{1y,\text{QBM}}$) is the difference between $Risk^{1y}$ and the risk measure extracted from the sample data augmented with the QBM synthesized data. The dependent variable is the same metric on an extended sample of the first 5 years after a firms listing. The second panel displays the results of a similar excercise, with the difference being that the synthesized data comes from the RBM. For the analysis we discretize the data by creating 256 equally sized buckets between returns between -50% and +50%. We disregard stocks with more than 10 zero return days in the first 200 days (stale prices). We add **600** synthesized observations to the original 200 sample observations for the augmented sample. The stars that indicate the significance level are $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.