TWEEDIE'S FORMULAE AND DIFFUSION GENERATIVE MODELS BEYOND GAUSSIAN

WENPIN TANG, NIZAR TOUZI, ZIKUN ZHANG, AND XUN YU ZHOU

Abstract. Diffusion models have achieved remarkable success in generating samples from unknown data distributions. Most popular stochastic differential equation–based diffusion models perturb the target distribution by adding Gaussian noise, transforming it into a simple prior, and then use denoising score matching, a consequence of Tweedie's formula, to learn the score function and generate clean samples from noise. However, non-Gaussian diffusion models with state-dependent diffusion coefficient have been largely underexplored, as have the corresponding Tweedie's formulae. In this work, we extend Tweedie's formula to important non-Gaussian processes, including geometric Brownian motion (GBM), squared Bessel (BESQ) processes, and Cox-Ingersoll-Ross (CIR) processes, thereby yielding the corresponding denoising score-matching objectives. We then apply the derived formulae to image and financial time series generation using GBM- and CIR-based diffusion models, and to empirical Bayes estimation under the BESQ setting. The reported experimental results demonstrate the potential of non-Gaussian models.

Key words: Bessel processes, denoising score matching, diffusion models, empirical Bayes, financial time series, geometric Brownian motion, Tweedie's formula.

1. Introduction

Diffusion models are a family of generative models that genuinely create samples from unknown target distributions [31, 65, 66]. They underpin the recent success in text-to-image creators such as DALL·E 2 [53], Stable Diffusion [58], and Flux [5], in text-to-video generators such as Sora [49], Make-A-Video [64], Seedance [23], and Veo [24], and in diffusion large language models such as Mercury [38], LLaDA [47], Dream [74], and WeDLM [43]. Recently, diffusion models have also been applied to other fields, including operations research [44, 75] and finance [1, 22, 27] for tabular data generation.

The idea of diffusion generative models relies on a forward–backward procedure:

- *Forward process*: starting from a training sample of the target distribution $X_0 \sim p_{\text{data}}(\cdot)$, the model gradually adds noise to transform the signal into noise $X_0 \to \cdots \to X_T \sim p_{\text{noise}}(\cdot)$.
- *Backward process*: start with the noise $X_T \sim p_{\text{noise}}(\cdot)$, and reverse the forward process to recover the signal from noise $X_T \to \cdots \to X_0 \sim p_{\text{data}}(\cdot)$.

The backward process is also termed as diffusion generative sampling or inference. In this paper, we adopt the continuous-time formulation, where the forward process $\{X_t\}_{0\le t\le T}$ is governed by a stochastic differential equation (SDE):

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \quad X_0 \sim p_{\text{data}}(\cdot).$$

Date: June 14, 2026.

As will be detailed in Section 2, the key to diffusion generative sampling hinges on the *score function* [1]:

$$\nabla \log p(t,x) := \nabla \log \left(\frac{\mathrm{d}}{\mathrm{d}x} \mathbb{P}(X_t \in \mathrm{d}x) \right).$$

Learning or estimating this score function is often referred to as *score-matching*.

For most existing diffusion generative models, e.g., variance exploding (VE) and variance preserving (VP) models [66], the diffusion coefficient $\sigma(t,x) = \sigma(t)I$ is only time-dependent, and the drift parameter $b(t,x) = f(t)x$ is linear in the state (or space) variable.[2] In this case, the process $\{X_t\}_{0\le t\le T}$ is Gaussian, and there is a systematic way to learn the score function $\nabla \log p(t,x)$ via *Tweedie's formula* [15, 56].

Now we briefly explain Tweedie's formula, which first appeared in correspondence between Herbert Robbins and Maurice Tweedie, and was rediscovered in various contexts [45, 46, 50, 52].[3] It was later popularized by Bradley Efron in the context of empirical Bayes estimation to tackle selection bias in genome analysis [15, 17]; see [33] for comprehensive discussions of empirical Bayes and Tweedie's formula. Let

$$U \sim \nu(\cdot) \quad \text{and} \quad V \,|\, U \sim \mathcal{N}(U, \sigma^2 I), \tag{1.1}$$

where $\mathcal{N}(\mu, \Sigma)$ denotes a Gaussian random vector with mean μ and covariance matrix Σ. Denote by $p_V(\cdot)$ the (marginal) distribution of V. Then Tweedie's formula yields

$$\mathbb{E}(U \,|\, V = v) = v + \sigma^2 \nabla \log p_V(v). \tag{1.2}$$

Notably, the equation (1.2) does not involve the "prior" $\nu(\cdot)$, and can be used to learn the score function of Gaussian diffusion models. To illustrate, consider the VE model $\mathrm{d}X_t = \sigma(t)\,\mathrm{d}W_t$, $X_0 \sim p_{\mathrm{data}}(\cdot)$. Specializing to $U = X_0$ and $V = X_t$ (so $\nu(\cdot)$ and σ^2 in (1.1) are $p_{\mathrm{data}}(\cdot)$ and $\int_0^t \sigma^2(s)\,\mathrm{d}s$, respectively) yields

$$\nabla \log p_{\mathrm{VE}}(t, X_t) = \frac{\mathbb{E}(X_0 \,|\, X_t) - X_t}{\int_0^t \sigma^2(s)\,\mathrm{d}s} \quad a.s., \tag{1.3}$$

under suitable integrability conditions on $X_0 \sim p_{\mathrm{data}}(\cdot)$. Next, we sample (X_0, X_t) according to $X_0 \sim p_{\mathrm{data}}(\cdot)$, $X_t \,|\, X_0 \sim \mathcal{N}\left(X_0, (\int_0^t \sigma^2(s)\,\mathrm{d}s) I\right)$, and regress X_0 over X_t to learn the score function because $\mathbb{E}(X_0 \,|\, X_t)$ is the L^2 projection of X_0 over X_t (see [60] for related discussions). More precisely, letting $\{s_\theta^{\mathrm{VE}}(\cdot,\cdot)\}_\theta$ be a parametrized family approximating the score function, we aim to solve the problem:

$$\min_\theta \mathbb{E}_{t\sim\mathcal{U}[0,T],\,(X_0,X_t)} \left| s_\theta^{\mathrm{VE}}(t, X_t) + \frac{X_t - X_0}{\int_0^t \sigma^2(s)\,\mathrm{d}s} \right|^2,$$

where $\mathcal{U}[0,T]$ denotes the uniform distribution over $[0,T]$, and (X_0, X_t) is sampled as described. The optimization problem is known as *denoising score-matching*, which a priori chooses the L^2 loss to achieve $\mathbb{E}_{t\sim\mathcal{U}[0,T],\,X_t} |s_{\theta_*}^{\mathrm{VE}}(t, X_t) - \nabla \log p_{\mathrm{VE}}(t, X_t)|^2 \approx 0$ if $\{s_\theta^{\mathrm{VE}}(\cdot,\cdot)\}_\theta$

[1] In other words, the score function is the logarithmic derivative of the probability density of X_t.

[2] For the VE model, $b(t,x) = 0$ and $\sigma(t,x) = \sigma(t)I$, with $\sigma(t) = a\sqrt{t}$ or ab^t for some $a, b > 0$. For the VP model, $b(t,x) = -\alpha(t)x$ and $\sigma(t,x) = \sqrt{2\alpha(t)}I$, with $\alpha(t) = at + b$ for some $a, b > 0$. In both cases, while the state X_t is typically high-dimensional, noises are independently added to each component of X_t.

[3] The formula is also known as Eddington–Tweedie formula [33] or Masreliez's theorem [45].

is sufficiently rich.[4] It is worth noting that Tweedie's formula suggests that the L^2 loss is a natural choice for denoising score matching, since the conditional expectation corresponds to the optimal L^2 regression estimator. Tweedie's formula and the corresponding denoising score-matching for other diffusion models with $\sigma(t,x) = \sigma(t)I$ and $b(t,x) = f(t)x$ can be derived similarly. In fact, the score function of any such model can be obtained from that of the VE model with $\sigma(t) = \sqrt{2t}$ by a space–time reparametrization [36]; see also [70, Section 5.1] and [71, Section 4.2].

Contributions. As discussed earlier, Tweedie's formula (1.2) is formulated for Gaussian distributions, and denoising score matching for existing diffusion models predominantly relies on their Gaussian structure, which enables the direct application of Tweedie's formula. However, using non-Gaussian models, such as (time-dependent) geometric Brownian motion (GBM) and Bessel-type processes, may be advantageous in certain generative AI tasks, which calls for a study on Tweedie's formula and the corresponding denoising score matching beyond Gaussian, especially for those with state-dependent diffusion coefficients. The main contribution of this paper is to carry out this study. Taking the one-dimensional setting for instance, the key to generalize Tweedie's formula is the following simple observation (see Proposition 2.3):

$$\sigma^2(t,x)\nabla \log p(t,x) + 2\sigma(t,x)\partial_x\sigma(t,x) = b(t,x) + \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathbb{E}\left(X_{t-\varepsilon} - X_t \,|\, X_t = x\right). \tag{1.4}$$

The identity (1.4) enables us to derive Tweedie's formula for non-Gaussian diffusion models with state-dependent diffusion coefficients and, consequently, the corresponding denoising score-matching objectives. Table 1 summarizes explicit Tweedie's formulae for several non-Gaussian models we derive in this paper which, to our best knowledge, is novel.[5] Notably, we emphasize that the flexible choice of time-dependent drift and diffusion coefficients is crucial to the empirical success of the corresponding diffusion models. In particular, specializing to $t = 1$ provides further extensions to Efron's generalization of Tweedie's formula to the exponential family with linear sufficient statistics and in the canonical form [15].[6] As for applications, we use non-Gaussian models (with score matching via Tweedie's formula) to perform image generation, financial time series generation, and empirical Bayes estimation, which have been largely underexplored in prior work.

---

[4] Denoising score-matching requires to solve:

$$\min_\theta \mathbb{E}_{t\sim\mathcal{U}[0,T],\,(X_0,X_t)}\left|s_\theta(t,X_t) - \nabla\log p(t,X_t\,|\,X_0)\right|^2,$$

which can be shown to be equivalent to the problem $\min_\theta \mathbb{E}_{t\sim\mathcal{U}[0,T],\,X_t}\left|s_\theta(t,X_t) - \nabla\log p(t,X_t)\right|^2$ [70, 73]. While Tweedie's formula provides a theoretical justification for denoising score-matching, the reverse is not true: denoising score-matching does not inherently imply Tweedie's formula. Another consequence is the sample complexity of denoising score-matching: the classical result [67, 68] shows that if $\nabla\log p(t,x)$ is p-times differentiable, then $\mathbb{E}_{t\sim\mathcal{U}[0,T],\,X_t}|s_{\theta_*}(t,X_t) - \nabla\log p(t,X_t)|^2 \lesssim m^{-\frac{2p}{2p+d}}$, where m is the number of samples. Refer to [10, 19, 48] for sharper results on diffusion score estimation via low-dimensional adaptation.

[5] As mentioned earlier, as noises are added component-wise to a multi-dimensional diffusion process, we only need those formulae in a scalar space.

[6] The exponential family is of the form $p(V\,|\,U) = h(V)\exp(T(V)\eta(U) - A(U))$, where $T(V)$ is the sufficient statistics. It is said to be in the canonical form if $\eta(U) = U$. Efron [15] generalizes Tweedie's formula to the exponential family with $T(V) \propto V$ and $\eta(U) = U$, which however does not include GBM and Bessel processes.

TABLE 1. Tweedie's Formulae for Gaussian and Non-Gaussian Models

Process	$b(t,x)$	$\sigma(t,x)$	Score function $\nabla \log p(t,x)$
VE	0	$\sigma(t)$	$\Sigma^{-2}(t)(\mathbb{E}(X_0 \,\vert\, X_t = x) - x)$
VP	$-\alpha(t)x$	$\sqrt{2\alpha(t)}$	$(1-e^{-2A(t)})^{-1}(e^{-A(t)}\,\mathbb{E}(X_0 \,\vert\, X_t = x) - x)$
GBM	$\mu(t)x$	$\sigma(t)x$	$\left(\frac{U(t)}{\Sigma^2(t)} - \frac{3}{2}\right)\frac{1}{x} - \frac{1}{\Sigma^2(t)}\frac{\log x}{x} + \frac{1}{\Sigma^2(t)}\frac{1}{x}\mathbb{E}(\log X_0 \,\vert\, X_t = x)$
BESQ	$2(\nu+1)$	$2\sqrt{x}$	$\frac{\nu}{x} - \frac{1}{2t} + \frac{1}{2t\sqrt{x}}\mathbb{E}(\sqrt{X_0} I_{\nu+1}(\frac{\sqrt{xX_0}}{t})/I_\nu(\frac{\sqrt{xX_0}}{t}) \,\vert\, X_t = x) =: s_\nu^{\mathrm{BESQ}}(t,x)$
CIR	$\alpha(t)(\mu(t)-x)$	$\sigma(t)\sqrt{x}$	$e^{A(t)} s^{\mathrm{BESQ}}_{\frac{2\alpha(t)\mu(t)}{\sigma^2(t)}-1(\equiv\nu)}\left(\frac{1}{4}\int_0^t \sigma^2(s)e^{A(s)}\,\mathrm{d}s, e^{A(t)}x\right)$
CEV	$\mu(t)x$	$\sigma(t)x^\beta$	$\frac{-2\beta+1}{x} - \frac{2(\beta-1)e^{2(\beta-1)U(t)}}{x^{2\beta-1}} s^{\mathrm{BESQ}}_{\frac{1}{2(\beta-1)}}\left((\beta-1)^2\int_0^t \sigma^2(s)e^{2(\beta-1)U(s)}\,\mathrm{d}s, e^{2(\beta-1)U(t)}x^{-2(\beta-1)}\right)$
BES(3)	$1/x$	1	$\frac{1}{x} + \frac{1}{t}\mathbb{E}((X_0 - x)\coth(\frac{xX_0}{t}) \,\vert\, X_t = x)$

Here, for all $t \geq 0$, $\Sigma^2(t) := \int_0^t \sigma^2(s)\,\mathrm{d}s$, $A(t) := \int_0^t \alpha(s)\,\mathrm{d}s$, and $U(t) := \int_0^t \mu(s)\,\mathrm{d}s$.
$I_\nu(\cdot)$ denotes the modified Bessel function of the first kind of order ν.
VE and VP are Gaussian models; GBM, BESQ, CIR, CEV and BES(3) are (positive) non-Gaussian models.

Related Works. SDE/score-based continuous diffusion models are formally introduced in [66]. Their empirical success across various applications relies on the careful design of the drift and diffusion coefficients in the forward SDE. State-of-the-art VE and VP models [36, 66] are Gaussian-based: they progressively add Gaussian noise to the original data distribution, use state-*independent* diffusion coefficients for simplicity and numerical stability, and generate the reverse sampling process from a Gaussian noise distribution.

Non-Gaussian SDE-based diffusion models are often associated with modeling discrete or categorical data on the probability simplex. [55] proposes a multi-dimensional CIR process as the forward noising process for simplex diffusions, which leads to a Dirichlet prior after normalizing the limiting multivariate Gamma distribution. Though the authors note that the CIR process is well suited for simplex diffusions over categorical data due to its positivity and the existence of a limiting distribution and comment on numerical simulations, they do not actually conduct numerical experiments. [21] applies an additive logistic transformation to the Ornstein–Uhlenbeck process to construct a positive forward SDE for simplex diffusions, derives the corresponding score-matching objective, and demonstrates their approach using MNIST images with pixels discretized into three categories. [3] proposes a Dirichlet diffusion score model for discrete and categorical data by constructing a multivariate diffusion process on the probability simplex. The forward process converges to a Dirichlet distribution and is built from independent univariate Jacobi diffusion processes via a stick-breaking construction. Inspired by Dale's law, [61] considers GBM with constant coefficients as a multiplicative forward process to model non-negative data and proposes a new multiplicative score-matching loss to train the model, showing the promise and applicability of the non-Gaussian models to datasets and domains where multiplicative noise is preferred. Their multiplicative score-matching loss for GBM differs from ours in that it is obtained by directly multiplying the state variable with the classical denoising score-matching loss while we focus on deriving such objectives for a broader class of non-Gaussian models in a systematic way. [39] puts forward a diffusion-based generative framework for financial time series that incorporates GBM into the forward process in the price space. Although the authors observe that, under a balance between the drift and diffusion coefficients, the model reduces to additive Gaussian noise

injection (i.e, a VE formulation) in the log-price space, they do not derive the corresponding score matching objectives for general GBM-based diffusion models.

Our primary goal is to derive Tweedie's formulae as denoising score matching objectives for various (important) diffusion models beyond Gaussian. That is, we express the score function $\nabla \log p(t,x)$ in terms of the conditional expectation $\mathbb{E}(g(t,x,X_0)\,|\,X_t=x)$ for some explicit function g. Upon the completion of this paper, we noted a recent preprint [72] that calculated $\mathbb{E}(X_0\,|\,X_t=x)$ for additive models beyond Gaussian via some transformations (such as differentiation and integration) of the density function $p(t,x)$. This type of "Tweedie calculus" (see also [26, 62] for special cases) is different from our results because it only applies to additive models whereas GBM and Bessel processes are not additive, and it goes "the other way around" by seeking an expression involving the density function for the conditional expectation rather than seeking a conditional expectation representation for the score function.

Organization of the Paper. The remainder of the paper is organized as follows. Section 2 provides background on diffusion models, where a proof of (1.4) is given. In Section 3, we derive Tweedie's formulae for various non-Gaussian diffusion models, including GBM and (squared) Bessel processes. Numerical experiments are reported in Section 4. We conclude in Section 5. Additional results on GBM-based Bayes estimation and experiments specifics are placed in the appendix.

2. Diffusion models and Tweedie's formulae

This section provides background on diffusion models, and presents a simple yet general approach to derive Tweedie's formulae in the context of diffusion models. We follow closely the presentation of [70].

Consider a forward SDE:

$$\mathrm{d}X_t = b(t,X_t)\,\mathrm{d}t + \sigma(t,X_t)\,\mathrm{d}W_t, \quad X_0 \sim p_{\mathrm{data}}(\cdot), \tag{2.1}$$

where $\{W_t\}_{t\geq 0}$ is n-dimensional Brownian motion, and $b:\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}^d$ and $\sigma:\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}^{d\times n}$ are drift and diffusion coefficients, respectively. Some conditions on $b(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ are required to ensure that (2.1) is well-posed (i.e. having existence and uniqueness of solution); see standard textbooks e.g., [35, 69] for details about the well-posedness of SDEs.

For ease of presentation, we assume that X_t has a (suitably smooth) probability density function $p(t,\cdot)$. The following theorem gives the time-reversal of the SDE (2.1), which lays the foundation of diffusion generative models.

Theorem 2.1. [2, 29] *Denote by $a(t,x) := \sigma(t,x)\sigma(t,x)^\top$. Under suitable conditions on $b(\cdot,\cdot)$, $\sigma(\cdot,\cdot)$ and $\{p(t,\cdot)\}_{0\leq t\leq T}$, we define*

$$\overline{\sigma}(t,x) = \sigma(T-t,x), \quad \overline{b}(t,x) = -b(T-t,x) + \frac{\nabla\cdot(p(T-t,x)a(T-t,x))}{p(T-t,x)},$$

as well as the process $\{Y_t\}_{0\leq t\leq T}$ by

$$\mathrm{d}Y_t = \overline{b}(t,Y_t)\,\mathrm{d}t + \overline{\sigma}(t,Y_t)\,\mathrm{d}B_t, \quad Y_0\sim p(T,\cdot),$$

where $\{B_t\}_{0\leq t\leq T}$ is a copy of Brownian motion. Then $\{Y_t\}_{0\leq t\leq T}$ and $\{X_{T-t}\}_{0\leq t\leq T}$ have the same marginal distribution, i.e., Y is the time reversal of X in law.

As mentioned earlier, the high-level idea of diffusion models is to recreate samples of the hidden target distribution from *noise*. However, the initialization $Y_0 \sim p(T,\cdot)$ depends on the *unknown* $p_{\text{data}}(\cdot)$ in each sample generation. One way to resolve this issue is to replace the initialization $Y_0 \sim p(T,\cdot)$ with some noise $p_{\text{noise}}(\cdot)$:

$$\mathrm{d}Y_t = \bar{b}(t,Y_t)\,\mathrm{d}t + \bar{\sigma}(t,Y_t)\,\mathrm{d}B_t, \quad Y_0 \sim p_{\text{noise}}(\cdot). \tag{2.2}$$

The choice of $p_{\text{noise}}(\cdot)$ is model-specific. For instance, $p_{\text{noise}}(\cdot)$ is taken as $\mathcal{N}\left(0, (\int_0^T \sigma^2(s)\,\mathrm{d}s) I\right)$ for the VE model, and $\mathcal{N}(0,I)$ for the VP model. It is expected that the closer the distributions $p(T,\cdot)$ and $p_{\text{noise}}(\cdot)$ are, the closer the distribution of Y_T sampled from (2.2) is to $p_{\text{data}}(\cdot)$; see [4, 6, 7, 41, 42] and [70, Section 6] along with references therein for the convergence theory of diffusion models.

Since $b(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ are chosen in advance, all but the term $\nabla \log p(T-t, Y_t)$ in (2.2) are available. So in order to sample the backward process (2.2), we need to learn or estimate the score function $\nabla \log p(t,x)$, known as *score-matching*, via a parameterized family $\{s_\theta(t,x)\}_\theta$. There are several existing score-matching methods, among which the most widely used one is denoising score matching [32, 73]. As explained in the introduction, denoising score-matching is essentially equivalent to Tweedie's formula for Gaussian models. However, Tweedie's formulae for non-Gaussian processes are absent, leaving out important examples such as GBM and Bessel processes. The goal of this paper is to develop a systematic approach to generalize Tweedie's formula to include a wider class of diffusion models for potential applications, premised upon the following result.

Proposition 2.2. *Under suitable conditions on $b(\cdot,\cdot)$, $\sigma(\cdot,\cdot)$ and $\{p(t,\cdot)\}_{0\le t\le T}$, we have for almost every x,*

$$a(t,x)\nabla \log p(t,x) + \nabla \cdot a(t,x) = b(t,x) + \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\mathbb{E}\left(X_{t-\varepsilon} - X_t \,|\, X_t = x\right). \tag{2.3}$$

Proof. By Theorem 2.1, we have:

$$Y_{T-t+\varepsilon} - Y_{T-t} = \int_{T-t}^{T-t+\varepsilon} \left(-b(T-s,Y_s) + a(T-s,Y_s)\nabla \log p(s,Y_s) + \nabla \cdot a(T-s,Y_s)\right)\mathrm{d}s + \int_{T-t}^{T-t+\varepsilon} \sigma(T-s,Y_s)\,\mathrm{d}B_s.$$

By the Lebesgue differentiation theorem, we get for almost every x,

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\mathbb{E}\left(Y_{T-t+\varepsilon} - Y_{T-t} \,|\, Y_{T-t} = x\right) = -b(t,x) + a(t,x)\nabla \log p(t,x) + \nabla \cdot a(t,x). \tag{2.4}$$

Identifying the left side of (2.4) with $\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\mathbb{E}\left(X_{t-\varepsilon} - X_t \,|\, X_t = x\right)$ yields the desired result. □

The identity (2.3) provides a systematic way to derive Tweedie's formulae for general diffusion models via the computation of $\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\mathbb{E}\left(X_{t-\varepsilon} - X_t \,|\, X_t = x\right)$:

$$\nabla \log p(t,x) = a(t,x)^{-1}\left(b(t,x) + \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\mathbb{E}\left(X_{t-\varepsilon} - X_t \,|\, X_t = x\right) - \nabla \cdot a(t,x)\right),$$

when $a(t,x)$ is invertible. As shown in Section 3, this quantity has a closed-form expression for many important models. Also note that the "almost everywhere" identity (2.3) is suitable for denoising score matching, which is defined as an L^2 loss.

3. Tweedie's Formulae

In this section we derive Tweedie's formulae for various diffusion processes. Although the SDE (2.1) is formulated in $\mathbb{R}^d$ with an n-dimensional Brownian motion, as discussed earlier diffusion generative models (including the VE and VP) always noises/denoise independently to each component of the vector state. Therefore, throughout this section we assume that $d = n = 1$.

3.1. Variance Exploding/Preserving Processes.

We start with deriving the formulae for the processes in the Gaussian family most commonly used by the current generative diffusion models.

3.1.1. *Variance Exploding Processes.* Consider the VE process:

$$\mathrm{d}X_t = \sigma(t)\,\mathrm{d}W_t, \quad X_0 \sim p_{\mathrm{data}}(\cdot),$$

where $\sigma(\cdot)$ is positive, continuous and bounded away from 0. Then we have

$$X_t = X_0 + \sqrt{\int_0^t \sigma^2(s)\,\mathrm{d}s}\cdot Z, \quad Z \sim \mathcal{N}(0,1). \tag{3.1}$$

Our goal is to compute the conditional expectation $\mathbb{E}(X_{t-\varepsilon} \,|\, X_0 = z, X_t = x)$ for $0 < \varepsilon \le t$.

Let $p_\Sigma(x) := \frac{1}{\sqrt{2\pi\Sigma}} \exp\left(-\frac{x^2}{2\Sigma}\right)$ be the probability density of a Gaussian random variable with mean zero and variance $\Sigma > 0$. Set

$$\Sigma_1 := \int_0^{t-\varepsilon} \sigma^2(s)\,\mathrm{d}s, \quad \Sigma_2 := \int_{t-\varepsilon}^t \sigma^2(s)\,\mathrm{d}s.$$

It is known (see [20]) that the probability density of $(X_{t-\varepsilon} \,|\, X_0 = z, X_t = x)$ is given by

$$\begin{aligned}\mathrm{d}\mathbb{P}(X_{t-\varepsilon} \in \mathrm{d}y \,|\, X_0 = z, X_t = x)/\mathrm{d}y &= \frac{p_{\Sigma_1}(z-y)p_{\Sigma_2}(x-y)}{p_{\Sigma_1+\Sigma_2}(x-z)} \\ &\propto \exp\left(-\frac{\Sigma_1+\Sigma_2}{2\Sigma_1\Sigma_2}\left(y - \frac{x\Sigma_1 + z\Sigma_2}{\Sigma_1+\Sigma_2}\right)^2\right).\end{aligned}$$

That is, $(X_{t-\varepsilon} \,|\, X_0 = z, X_t = x)$ is a Gaussian random variable with mean $\frac{x\Sigma_1+z\Sigma_2}{\Sigma_1+\Sigma_2}$ and variance $\frac{\Sigma_1\Sigma_2}{\Sigma_1+\Sigma_2}$. Thus,

$$\mathbb{E}(X_{t-\varepsilon} \,|\, X_0 = z, X_t = x) = \frac{x\Sigma_1 + z\Sigma_2}{\Sigma_1+\Sigma_2},$$

and

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathbb{E}(X_{t-\varepsilon} - X_t \,|\, X_0 = z, X_t = x) = (z-x)\sigma^2(t)\left(\int_0^t \sigma^2(s)\,\mathrm{d}s\right)^{-1}.$$

This recovers the standard Tweedie's formula:

$$\nabla \log p_\sigma^{\mathrm{VE}}(t,x) = \left(\int_0^t \sigma^2(s)\,\mathrm{d}s\right)^{-1}\left(\mathbb{E}(X_0 \,|\, X_t = x) - x\right). \tag{3.2}$$

3.1.2. *Variance Preserving Processes.* For the VP process

$$\mathrm{d}X_t = -\alpha(t)X_t\,\mathrm{d}t + \sqrt{2\alpha(t)}\,\mathrm{d}W_t, \quad X_0 \sim p_{\mathrm{data}}(\cdot),$$

where $\alpha(\cdot)$ is a increasing positive function, we know that

$$X_t = e^{-\int_0^t \alpha(s)\,\mathrm{d}s} \cdot X_0 + \sqrt{1 - e^{-2\int_0^t \alpha(s)\,\mathrm{d}s}} \cdot Z, \quad Z \sim \mathcal{N}(0,1).$$

Hence, by (3.1) and (3.2), we get Tweedie's formula for the VP process:

$$\nabla \log p_\alpha^{\mathrm{VP}}(t,x) = (1 - e^{-2\int_0^t \alpha(s)\,\mathrm{d}s})^{-1}(e^{-\int_0^t \alpha(s)\,\mathrm{d}s}\,\mathbb{E}(X_0 \,|\, X_t = x) - x). \tag{3.3}$$

3.2. **Geometric Brownian Motion.** We consider $b(t,x) = \mu(t)x$ and $\sigma(t,x) = \sigma(t)x$ with $\sigma(t) \geq 0$. The forward process is the time-dependent geometric Brownian motion:

$$\mathrm{d}X_t = \mu(t)X_t\,\mathrm{d}t + \sigma(t)X_t\,\mathrm{d}W_t, \quad X_0 \sim p_{\mathrm{data}}(\cdot),$$

which has a closed form solution

$$X_t = X_0 \exp\left(\int_0^t \left(\mu(s) - \frac{1}{2}\sigma^2(s)\right)\,\mathrm{d}s + \int_0^t \sigma(s)\,\mathrm{d}W_s\right), \quad t \geq 0.$$

It is easy to see that

$$\mathbb{E}(X_{t-\varepsilon} \,|\, X_0 = z, X_t = x) = z e^{\int_0^{t-\epsilon}(\mu(s) - \frac{1}{2}\sigma^2(s))\,\mathrm{d}s} \cdot \mathbb{E}\left(e^{M_{t-\epsilon}} \,|\, M_t = \kappa(t)\right),$$

where

$$M_t := \int_0^t \sigma(s)\,\mathrm{d}W_s, \quad \text{and} \quad \kappa(t) := \log\left(\frac{x}{z}\right) - \int_0^t \left(\mu(s) - \frac{1}{2}\sigma^2(s)\right)\,\mathrm{d}s, \quad \forall t \geq 0.$$

By the Dubins–Schwarz theorem [13], there exists a copy of Brownian motion, denoted by $\{B_t\}_{t\geq 0}$, such that $M_t = B_{\tau_t}$, where $\tau_t := \langle M\rangle_t = \int_0^t \sigma^2(s)\,\mathrm{d}s$. Then we have

$$\mathbb{E}\left(e^{M_{t-\epsilon}} \,|\, M_t = \kappa(t)\right) = \mathbb{E}\left(e^{B_{\tau_{t-\epsilon}}} \,|\, B_{\tau_t} = \kappa(t)\right) = \exp\left(\frac{\int_0^{t-\epsilon}\sigma^2(s)\,\mathrm{d}s}{\int_0^t \sigma^2(s)\,\mathrm{d}s}\left(\kappa(t) + \frac{1}{2}\int_{t-\epsilon}^t \sigma^2(s)\,\mathrm{d}s\right)\right),$$

where we have used the fact that $(B_{\tau_{t-\epsilon}} \,|\, B_{\tau_t} = \kappa(t)) \sim \mathcal{N}(\kappa(t)\tau_{t-\epsilon}/\tau_t, \tau_{t-\epsilon}(\tau_t - \tau_{t-\epsilon})/\tau_t)$ in the second equality. Consequently, we have

$$\begin{aligned}\mathbb{E}(X_{t-\varepsilon} \mid X_0 = z, X_t = x) = x \exp\Bigg(\Bigg(&\frac{\int_0^{t-\epsilon}(\mu(s) - \frac{1}{2}\sigma^2(s))\,\mathrm{d}s}{\int_0^t(\mu(s) - \frac{1}{2}\sigma^2(s))\,\mathrm{d}s} - \frac{1}{\int_0^t \sigma^2(s)\,\mathrm{d}s}\log\frac{x}{z} \\ &+ \frac{\int_0^{t-\epsilon}\sigma^2(s)\,\mathrm{d}s}{\int_0^t \sigma^2(s)\,\mathrm{d}s}\left(\frac{-\int_0^t(\mu(s) - \frac{1}{2}\sigma^2(s))\,\mathrm{d}s}{\int_{t-\epsilon}^t \sigma^2(s)\,\mathrm{d}s} + \frac{1}{2}\right)\Bigg)\int_{t-\epsilon}^t \sigma^2(s)\,\mathrm{d}s\Bigg)\end{aligned}$$

and

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathbb{E}(X_{t-\varepsilon} - X_t \,|\, X_0 = z, X_t = x) = x\left(-\frac{\sigma^2(t)}{\int_0^t \sigma^2(s)\,\mathrm{d}s}\log\frac{x}{z} - \mu(t) + \frac{\sigma^2(t)}{2} + \sigma^2(t)\frac{\int_0^t \mu(s)\,\mathrm{d}s}{\int_0^t \sigma^2(s)\,\mathrm{d}s}\right).$$

We obtain Tweedie's formula for time-dependent GBM:

$$\nabla \log p_{\mu,\sigma}^{\mathrm{GBM}}(t,x) = \left(\frac{\int_0^t \mu(s)\,\mathrm{d}s}{\int_0^t \sigma^2(s)\,\mathrm{d}s} - \frac{3}{2}\right)\frac{1}{x} - \frac{1}{\int_0^t \sigma^2(s)\,\mathrm{d}s}\frac{\log x}{x} + \frac{1}{\int_0^t \sigma^2(s)\,\mathrm{d}s}\frac{1}{x}\mathbb{E}(\log X_0 \,|\, X_t = x). \tag{3.4}$$

3.3. **Squared Bessel Processes.** Here we take $b(t,x) = 2(\nu+1)$ with $\nu > 0$, and $\sigma(t,x) = 2\sqrt{x}$. The forward process is governed by

$$\mathrm{d}X_t = 2(\nu+1)\,\mathrm{d}t + 2\sqrt{X_t}\,\mathrm{d}W_t, \quad X_0 \sim p_{\mathrm{data}}(\cdot), \tag{3.5}$$

which is transient. It follows from [51, (2.i)] (see also [54, Chapter XI]) that the transition density of the process $\{X_t\}_{t\geq 0}$ is

$$q(t,x,y) := \mathrm{d}\mathbb{P}(X_{t_0+t} \in \mathrm{d}y \,|\, X_{t_0} = x)/\mathrm{d}y = \frac{1}{2t}\left(\frac{y}{x}\right)^{\frac{\nu}{2}} \exp\left(-\frac{x+y}{2t}\right) I_\nu\left(\frac{\sqrt{xy}}{t}\right),$$

where $I_\nu(\cdot)$ is the modified Bessel function of the first kind of order ν. So the probability density of $(X_{t-\varepsilon} \,|\, X_0 = z, X_t = x)$ is given by

$$\begin{aligned}\mathrm{d}\mathbb{P}(X_{t-\varepsilon} \in \mathrm{d}y \,|\, X_0 = z, X_t = x)/\mathrm{d}y &= \frac{q(t-\varepsilon,z,y)q(\varepsilon,y,x)}{q(t,z,x)} \\ &= \frac{\exp\left(-\frac{ty}{2(t-\varepsilon)\varepsilon}\right) I_\nu\left(\frac{\sqrt{zy}}{t-\varepsilon}\right) I_\nu\left(\frac{\sqrt{xy}}{\varepsilon}\right)}{\int_0^\infty \exp\left(-\frac{ty}{2(t-\varepsilon)\varepsilon}\right) I_\nu\left(\frac{\sqrt{zy}}{t-\varepsilon}\right) I_\nu\left(\frac{\sqrt{xy}}{\varepsilon}\right) \mathrm{d}y}.\end{aligned}$$

By letting

$$F(a,b,c;\nu) := \int_0^\infty e^{-cu} I_\nu(a\sqrt{u}) I_\nu(b\sqrt{u})\,\mathrm{d}u \quad \text{and} \quad G(a,b,c;\nu) := -\frac{\partial F}{\partial c}, \tag{3.6}$$

we have

$$\mathbb{E}(X_{t-\varepsilon} \,|\, X_0 = z, X_t = x) = \frac{G\left(\frac{\sqrt{z}}{t-\varepsilon}, \frac{\sqrt{x}}{\varepsilon}, \frac{t}{2(t-\varepsilon)\varepsilon}; \nu\right)}{F\left(\frac{\sqrt{z}}{t-\varepsilon}, \frac{\sqrt{x}}{\varepsilon}, \frac{t}{2(t-\varepsilon)\varepsilon}; \nu\right)}. \tag{3.7}$$

According to [12, (10.43.28)] and by a change of variable $\sqrt{u} \to u$, we get

$$F(a,b,c;\nu) = \frac{1}{c}\exp\left(\frac{a^2+b^2}{4c}\right) I_\nu\left(\frac{ab}{2c}\right). \tag{3.8}$$

Thus,

$$\begin{aligned}G(a,b,c;\nu) &= \frac{1}{c^2}\exp\left(\frac{a^2+b^2}{4c}\right)\left(\left(1+\frac{a^2+b^2}{4c}\right) I_\nu\left(\frac{ab}{2c}\right) + \frac{ab}{2c} I'_\nu\left(\frac{ab}{2c}\right)\right) \\ &= \frac{1}{c^2}\exp\left(\frac{a^2+b^2}{4c}\right)\left(\left(1+\frac{a^2+b^2}{4c}+\nu\right) I_\nu\left(\frac{ab}{2c}\right) + \frac{ab}{2c} I_{\nu+1}\left(\frac{ab}{2c}\right)\right),\end{aligned} \tag{3.9}$$

where we use the fact that $I'_\nu(u) = \frac{\nu}{u} I_\nu(u) + I_{\nu+1}(u)$ (see [12, (10.29.2)]) in the last equation. Combining (3.7), (3.8) and (3.9) yields

$$\begin{aligned}\mathbb{E}(X_{t-\varepsilon} \,|\, X_0 = z, X_t = x) &= \frac{2(t-\varepsilon)\varepsilon}{t}\left(1+\nu+\frac{\varepsilon^2 z + (t-\varepsilon)^2 x}{2(t-\varepsilon)\varepsilon t} + \frac{\sqrt{xz}}{t}\frac{I_{\nu+1}\left(\frac{\sqrt{xz}}{t}\right)}{I_\nu\left(\frac{\sqrt{xz}}{t}\right)}\right) \\ &= x + 2\varepsilon\left(1+\nu-\frac{x}{t}+\frac{\sqrt{xz}}{t}\frac{I_{\nu+1}\left(\frac{\sqrt{xz}}{t}\right)}{I_\nu\left(\frac{\sqrt{xz}}{t}\right)}\right) + \mathcal{O}(\varepsilon^2),\end{aligned}$$

which implies that

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathbb{E}(X_{t-\varepsilon}-X_t\,|\,X_0=z,X_t=x)=2\left(1+\nu-\frac{x}{t}+\frac{\sqrt{xz}}{t}\frac{I_{\nu+1}\left(\frac{\sqrt{xz}}{t}\right)}{I_\nu\left(\frac{\sqrt{xz}}{t}\right)}\right).$$

We obtain Tweedie's formula for squared Bessel process:

$$\begin{aligned}s_\nu^{\mathrm{BESQ}}(t,x):&=\nabla\log p_\nu^{\mathrm{BESQ}}(t,x)\\&=\frac{\nu}{x}-\frac{1}{2t}+\frac{1}{2t\sqrt{x}}\mathbb{E}\left(\frac{\sqrt{X_0}I_{\nu+1}\left(\frac{\sqrt{xX_0}}{t}\right)}{I_\nu\left(\frac{\sqrt{xX_0}}{t}\right)}\,\middle|\,X_t=x\right).\end{aligned}\tag{3.10}$$

More generally, we take $b(t,x)=\mu$ and $\sigma(t,x)=\sigma\sqrt{x}$, with $\sigma>0$ and $\mu>\frac{\sigma^2}{2}$. The forward process evolves as

$$\mathrm{d}X_t=\mu\,\mathrm{d}t+\sigma\sqrt{X_t}\,\mathrm{d}W_t,\quad X_0\sim p_{\mathrm{data}}(\cdot).$$

It is easy to see that $\{\frac{4}{\sigma^2}X_t\}_{t\geq 0}$ is a squared Bessel process with index $\nu=\frac{2\mu}{\sigma^2}-1$. Hence, Tweedie's formula is

$$\nabla\log p_{\mu,\sigma}^{\mathrm{BESQ}}(t,x)=\frac{\sigma^2}{4}s_{\frac{2\mu}{\sigma^2}-1}^{\mathrm{BESQ}}\left(t,\frac{\sigma^2x}{4}\right),$$

where $s_\nu^{\mathrm{BESQ}}(\cdot,\cdot)$ is defined by (3.10).

3.4. **CIR Processes.** We now consider the forward process of the form

$$\mathrm{d}X_t=\alpha(t)(\mu(t)-X_t)\,\mathrm{d}t+\sigma(t)\sqrt{X_t}\,\mathrm{d}W_t,\quad X_0\sim p_{\mathrm{data}}(\cdot).\tag{3.11}$$

This process is known as the CIR process [9], which has been used to model interest rate [57, 63], as well as continuous-state branching processes with immigration as the scaling limit of the Galton–Watson counterparts [37, 40].

3.4.1. *Time-Independent Case.* We first consider (3.11) where $\alpha(t),\mu(t)$ and $\sigma(t)$ are time independent; i.e. $b(t,x)=\alpha(\mu-x)$ and $\sigma(t,x)=\sigma\sqrt{x}$, with constants $\alpha>0$, $\sigma>0$, and $\mu>\frac{\sigma^2}{2\alpha}$. It follows from [9, Equation (18)] (see also [34, Proposition 6.3.1.1]) that $\{X_t\}_{t\geq 0}$ is a space-time-changed squared Bessel process:

$$X_t\overset{d}{=}e^{-\alpha t}Z_{\frac{\sigma^2}{4\alpha}(e^{\alpha t}-1)},$$

where $\{Z_t\}_{t\geq 0}$ is squared Bessel process with index $\nu=\frac{2\alpha\mu}{\sigma^2}-1$. As a result, Tweedie's formula for CIR processes is

$$\nabla\log p_{\alpha,\mu,\sigma}^{\mathrm{CIR}}(t,x)=e^{\alpha t}s_{\frac{2\alpha\mu}{\sigma^2}-1}^{\mathrm{BESQ}}\left(\frac{\sigma^2}{4\alpha}(e^{\alpha t}-1),e^{\alpha t}x\right),\tag{3.12}$$

where $s_\nu^{\mathrm{BESQ}}(\cdot,\cdot)$ is defined by (3.10).

3.4.2. *Time-Dependent Case.* We now turn to the general time-dependent CIR (3.11). We choose $\alpha(t) > 0$, $\mu(t) > 0$, and $\sigma(t) > 0$ such that $\frac{2\alpha(t)\mu(t)}{\sigma^2(t)} - 1 \equiv \nu$. By [63, Corollary 3.1], we have:

$$X_t \overset{d}{=} e^{-\int_0^t \alpha(s)\,\mathrm{d}s} Z_{\frac{1}{4}\int_0^t \sigma^2(s) e^{\int_0^s \alpha(u)\,\mathrm{d}u}\,\mathrm{d}s},$$

where $\{Z_t\}_{t\geq 0}$ is squared Bessel process with index ν.[7] Tweedie's formula is then given by

$$\nabla \log p^{\mathrm{CIR}}_{\alpha,\mu,\sigma}(t,x) = e^{\int_0^t \alpha(s)\,\mathrm{d}s} s^{\mathrm{BESQ}}_{\nu}\left(\frac{1}{4}\int_0^t \sigma^2(s) e^{\int_0^s \alpha(u)\,\mathrm{d}u}\,\mathrm{d}s, e^{\int_0^t \alpha(s)\,\mathrm{d}s} x\right), \tag{3.13}$$

where $s^{\mathrm{BESQ}}_{\nu}(\cdot,\cdot)$ is defined by (3.10).

3.5. **(Time-Dependent) CEV Processes.** We consider $b(t,x) = \mu(t)x$ and $\sigma(t,x) = \sigma(t)x^{\beta}$, with $\mu(t) > 0$, $\sigma(t) > 0$ and $\beta > 1$. The forward process is[8]

$$\mathrm{d}X_t = \mu(t)X_t\,\mathrm{d}t + \sigma(t)X_t^{\beta}\,\mathrm{d}W_t, \quad X_0 \sim p_{\mathrm{data}}(\cdot). \tag{3.14}$$

It follows from [34, Lemma 6.4.3.1] that $\{X_t\}_{t\geq 0}$ is a time-change of a power of squared Bessel process

$$X_t \overset{d}{=} e^{\int_0^t \mu(s)\,\mathrm{d}s} Z^{-\frac{1}{2(\beta-1)}}_{(\beta-1)^2\int_0^t \sigma^2(s) e^{2(\beta-1)\int_0^s \mu(u)\,\mathrm{d}u}\,\mathrm{d}s},$$

where $\{Z_t\}_{t\geq 0}$ is squared Bessel with index $\frac{1}{2(\beta-1)}$.[9] As a result, Tweedie's formula for (time-dependent) CEV processes is given by

$$\begin{aligned}\nabla \log p^{\mathrm{CEV}}_{\mu,\sigma,\beta}(t,x) &= \frac{-2\beta+1}{x} \\ &\quad - \frac{2(\beta-1)e^{2(\beta-1)\int_0^t \mu(s)\,\mathrm{d}s}}{x^{2\beta-1}} s^{\mathrm{BESQ}}_{\frac{1}{2(\beta-1)}}\left((\beta-1)^2\int_0^t \sigma^2(s) e^{2(\beta-1)\int_0^s \mu(u)\,\mathrm{d}u}\,\mathrm{d}s, e^{2(\beta-1)\int_0^t \mu(s)\,\mathrm{d}s} x^{-2(\beta-1)}\right),\end{aligned}$$

where $s^{\mathrm{BESQ}}_{\nu}(\cdot,\cdot)$ is defined by (3.10).

3.6. **Bessel Processes.** Here we take $b(t,x) = \frac{2\nu+1}{2x}$ and $\sigma(t,x) = 1$, with $\nu > 0$. The forward process is governed by

$$\mathrm{d}X_t = \frac{2\nu+1}{2X_t}\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0 \sim p_{\mathrm{data}}(\cdot). \tag{3.15}$$

By [51, (2.i')], the transition density of the process $\{X_t\}_{t\geq 0}$ is

$$\tilde{q}(t,x,y) := \mathrm{d}\mathbb{P}(X_{t_0+t} \in \mathrm{d}y \,|\, X_{t_0} = x)/\mathrm{d}y = \frac{1}{t}\left(\frac{y}{x}\right)^{\nu} y \exp\left(-\frac{x^2+y^2}{2t}\right) I_{\nu}\left(\frac{xy}{t}\right).$$

[7] As shown in [63], the time-dependent CIR process, after space-time scaling, is squared Bessel process with a "time-dependent" index $\nu(t) := \frac{2\alpha(t)\mu(t)}{\sigma(t)} - 1$, being characterized by the Laplace transform. For our generation purpose, we set this index to be constant/time-independent in order to apply the result from the squared Bessel process.

[8] When $\mu(t) \equiv \mu$ and $\sigma(t) \equiv \sigma$ are time-independent, this process is known as the constant elasticity of variance (CEV) process [8, 11, 18].

[9] In contrast with the CIR processes, the index does not depend on the time-dependent coefficients $\mu(t)$ and $\sigma(t)$.

So the probability density of $(X_{t-\varepsilon} \,|\, X_0 = z, X_t = x)$ is

$$\mathrm{d}\mathbb{P}(X_{t-\varepsilon} \in \mathrm{d}y \,|\, X_0 = z, X_t = x)/\mathrm{d}y = \frac{y \exp\left(-\frac{ty^2}{2(t-\varepsilon)\varepsilon}\right) I_\nu\left(\frac{zy}{t-\varepsilon}\right) I_\nu\left(\frac{xy}{\varepsilon}\right)}{\int_0^\infty y \exp\left(-\frac{ty^2}{2(t-\varepsilon)\varepsilon}\right) I_\nu\left(\frac{zy}{t-\varepsilon}\right) I_\nu\left(\frac{xy}{\varepsilon}\right) \mathrm{d}y}.$$

Note from (3.6) that $F(a,b,c;\nu) = 2\int_0^\infty u e^{-cu^2} I_\nu(au) I_\nu(bu)\,\mathrm{d}u$. Furthermore, let

$$H(a,b,c;\nu) := \int_0^\infty u^2 e^{-cu^2} I_\nu(au) I_\nu(bu)\,\mathrm{d}u.$$

Then we have

$$\mathbb{E}(X_{t-\varepsilon} \,|\, X_0 = z, X_t = x) = \frac{2H\left(\frac{z}{t-\varepsilon}, \frac{x}{\varepsilon}, \frac{t}{2(t-\varepsilon)\varepsilon}; \nu\right)}{F\left(\frac{z}{t-\varepsilon}, \frac{x}{\varepsilon}, \frac{t}{2(t-\varepsilon)\varepsilon}; \nu\right)}. \tag{3.16}$$

Recall the expression of $F(a,b,c)$ from (3.8). However, for general ν, there is no closed form expression for $H(a,b,c)$.

Next, let us consider the special case $\nu = \frac{1}{2}$, which corresponds to the three dimensional Bessel process. Noting $I_{\frac{1}{2}}(u) = \sqrt{\frac{2}{\pi}}\frac{\sinh u}{\sqrt{u}}$, we have

$$\begin{aligned}
&H\left(a,b,c;\frac{1}{2}\right) \\
&= \frac{2}{\pi\sqrt{ab}} \int_0^\infty u e^{-cu^2} \sinh(au)\sinh(bu)\,\mathrm{d}u \\
&= \frac{1}{\pi\sqrt{ab}} \left(\int_0^\infty u e^{-cu^2} \cosh((a+b)u)\,\mathrm{d}u - \int_0^\infty u e^{-cu^2} \cosh((a-b)u)\,\mathrm{d}u \right) \\
&= \frac{1}{\pi\sqrt{ab}} \left(\frac{a+b}{2c} \int_0^\infty e^{-cu^2} \sinh((a+b)u)\,\mathrm{d}u - \frac{a-b}{2c} \int_0^\infty e^{-cu^2} \sinh((a-b)u)\,\mathrm{d}u \right) \\
&= \frac{(a+b)\exp\left(\frac{(a+b)^2}{4c}\right) \operatorname{erf}\left(\frac{a+b}{2\sqrt{c}}\right) - (a-b)\exp\left(\frac{(a-b)^2}{4c}\right) \operatorname{erf}\left(\frac{a-b}{2\sqrt{c}}\right)}{4\sqrt{\pi abc^3}},
\end{aligned} \tag{3.17}$$

where we use the formula [25, (3.546.1)] in the last equation, with $\operatorname{erf}(u) := \frac{2}{\sqrt{\pi}}\int_0^u e^{-z^2}\,\mathrm{d}z$ the error function of the standard normal. Combining (3.16), (3.8) and (3.17) yields:

$$\begin{aligned}
\mathbb{E}(X_{t-\varepsilon} \,|\, X_0 = z, X_t = x) &= \frac{\varepsilon z + (t-\varepsilon)x}{2t} \frac{e^{\frac{xz}{t}}}{\sinh\left(\frac{xz}{t}\right)} \operatorname{erf}\left(\frac{\varepsilon z + (t-\varepsilon)x}{\sqrt{2t(t-\varepsilon)\varepsilon}}\right) \\
&\quad - \frac{\varepsilon z - (t-\varepsilon)x}{2t} \frac{e^{-\frac{xz}{t}}}{\sinh\left(\frac{xz}{t}\right)} \operatorname{erf}\left(\frac{\varepsilon z - (t-\varepsilon)x}{\sqrt{2t(t-\varepsilon)\varepsilon}}\right) \\
&= x + \frac{\varepsilon(z-x)}{t} \coth\left(\frac{xz}{t}\right) + \mathcal{O}(\varepsilon^2),
\end{aligned}$$

which implies

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \mathbb{E}(X_{t-\varepsilon} - X_t \,|\, X_0 = z, X_t = x) = \frac{z-x}{t} \coth\left(\frac{xz}{t}\right).$$

This leads to Tweedie's formula for three-dimensional Bessel process:

$$s^{\mathrm{BES}}_{\frac{1}{2}}(t,x) := \nabla \log p^{\mathrm{BES}}_{\frac{1}{2}}(t,x) = \frac{1}{x} + \frac{1}{t}\mathbb{E}\left((X_0 - x)\coth\left(\frac{xX_0}{t}\right) \middle| X_t = x \right). \tag{3.18}$$

A slightly more general case than the above is to choose $b(t,x) = \frac{\sigma^2}{x}$ and $\sigma(t,x) = \sigma$, with $\sigma > 0$. The forward process is

$$\mathrm{d}X_t = \frac{\sigma^2}{X_t}\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_0 \sim p_{\mathrm{data}}(\cdot).$$

It is easy to see that $\{\frac{1}{\sigma}X_t\}_{t\geq 0}$ is a three-dimensional Bessel process. As a result, Tweedie's formula is

$$\nabla \log p^{\mathrm{BES}}_{\frac{1}{2},\sigma}(t,x) = \sigma s^{\mathrm{BES}}_{\frac{1}{2}}(t,\sigma x),$$

where $s^{\mathrm{BES}}_{\frac{1}{2}}(\cdot,\cdot)$ is defined by (3.18).

4. Numerical experiments

4.1. **Diffusion Models.** This section provides numerical experiments using non-Gaussian models, where we focus on GBM and CIR processes, to formulate the forward process of SDE-based diffusion models. The experimental details are deferred to Appendix B. For the sake of comparison, we also presents the results generated by Gaussian models. We first derive denoising score matching objectives for these models right from the Tweedie's formula established.

Variance Exploding Processes. From (3.2) we know

$$\mathbb{E}(X_0 \mid X_t) = X_t + \Sigma^2(t)\cdot \nabla \log p^{\mathrm{VE}}_{\sigma}(t, X_t),$$

where we have denoted $\Sigma^2(t) := \int_0^t \sigma^2(s)\,\mathrm{d}s$. Thus, the denoising score matching objective is

$$\begin{aligned} &\min_{\theta} \mathbb{E}\left| X_t + \Sigma^2(t)\cdot s^{\mathrm{VE}}_{\theta}(t,X_t) - X_0 \right|^2 \\ =\ &\min_{\theta} \mathbb{E}_{t\sim\mathcal{U}[0,T],\, X_0\sim p_{\mathrm{data}}(\cdot),\, Z\sim\mathcal{N}(0,1)} \Sigma^2(t)|\Sigma(t)\cdot s^{\mathrm{VE}}_{\theta}(t, X_0 + \Sigma(t)Z) + Z|^2, \end{aligned}$$

where the equality is because $X_t = X_0 + \Sigma(t)Z$ with $Z\sim\mathcal{N}(0,1)$. One can alternatively adopt the reparameterization $\epsilon^{\mathrm{VE}}_{\theta}(t,x) := -\Sigma(t)\cdot s^{\mathrm{VE}}_{\theta}(t,x)$ to obtain a noise-prediction objective

$$\min_{\theta} \mathbb{E}_{t\sim\mathcal{U}[0,T],\, X_0\sim p_{\mathrm{data}}(\cdot),\, Z\sim\mathcal{N}(0,1)} \Sigma^2(t)|\epsilon^{\mathrm{VE}}_{\theta}(t, X_0 + \Sigma(t)Z) - Z|^2.$$

Variance Preserving Processes. By (3.3), we have

$$\mathbb{E}(X_0 \mid X_t) = \frac{1}{A(t)}(X_t + \Sigma^2(t)\cdot \nabla \log p^{\mathrm{VP}}_{\alpha}(t, X_t)),$$

where $A(t) := \exp(-\int_0^t \alpha(s)\,\mathrm{d}s)$ and $\Sigma^2(t) := 1 - A^2(t) = 1 - \exp(-2\int_0^t \alpha(s)\,\mathrm{d}s)$. Hence, the corresponding denoising score matching objective is given by

$$\begin{aligned}
&\min_\theta \mathbb{E}\left|\frac{1}{A(t)}(X_t + \Sigma^2(t)\cdot s_\theta^{\mathrm{VP}}(t, X_t)) - X_0\right|^2 \\
=\ &\min_\theta \mathbb{E}_{t\sim\mathcal{U}[0,T],\, X_0\sim p_{\mathrm{data}}(\cdot),\, Z\sim\mathcal{N}(0,1)} \frac{\Sigma^2(t)}{A^2(t)}|\Sigma(t)\cdot s_\theta^{\mathrm{VP}}(t, A(t)X_0 + \Sigma(t)Z) + Z|^2 \\
=\ &\min_\theta \mathbb{E}_{t\sim\mathcal{U}[0,T],\, X_0\sim p_{\mathrm{data}}(\cdot),\, Z\sim\mathcal{N}(0,1)} \frac{1}{\lambda_t}|\epsilon_\theta^{\mathrm{VP}}(t, A(t)X_0 + \Sigma(t)Z) - Z|^2,
\end{aligned}$$

where the first equality is because $X_t = A(t)X_0 + \Sigma(t)Z,\ Z\sim\mathcal{N}(0,1)$, and in the second inequality we let $\epsilon_\theta^{\mathrm{VP}}(t,x) := -\Sigma(t)\cdot s_\theta^{\mathrm{VP}}(t,x)$ be the noise-prediction reparameterization and $\lambda_t := A^2(t)/\Sigma^2(t)$ the signal-to-noise ratio.

Geometric Brownian Motion. It follows from (3.4) that

$$\mathbb{E}(\log X_0 \,|\, X_t) = \Sigma^2(t) X_t \nabla \log p_{\mu,\sigma}^{\mathrm{GBM}}(t, X_t) - \Sigma^2(t)\left(\frac{U(t)}{\Sigma^2(t)} - \frac{3}{2}\right) + \log X_t,$$

where $U(t) := \int_0^t \mu(s)\,\mathrm{d}s$ and $\Sigma^2(t) := \int_0^t \sigma^2(s)\,\mathrm{d}s$. Replacing the true score $\nabla \log p_{\mu,\sigma}^{\mathrm{GBM}}(t,x)$ with parameterized family $\{s_\theta^{\mathrm{GBM}}(t,x)\}_\theta$, we get the denoising score matching objective:

$$\begin{aligned}
&\min_\theta \mathbb{E}\left|\Sigma^2(t) X_t s_\theta^{\mathrm{GBM}}(t, X_t) - \Sigma^2(t)\left(\frac{U(t)}{\Sigma^2(t)} - \frac{3}{2}\right) + \log X_t - \log X_0\right|^2 \\
=\ &\min_\theta \mathbb{E}_{t\sim\mathcal{U}[0,T],\, X_0\sim p_{\mathrm{data}}(\cdot),\, Z\sim\mathcal{N}(0,1)} \Sigma^2(t)|\Sigma(t)(1 + X_t s_\theta^{\mathrm{GBM}}(t, X_t)) + Z|^2,
\end{aligned}$$

where $X_t = X_0 \exp(U(t) - \frac{1}{2}\Sigma^2(t) + \Sigma(t)Z)$ with $Z\sim\mathcal{N}(0,1)$.

We need to be mindful that naïvely applying the above score matching objective for training leads to an excessively large initial loss due to the multiplicative exponential Gaussian noise in the forward process. Moreover, even when the score network is well trained, sampling from the reverse-time dynamics may suffer from numerical blow-up because of the state-dependent diffusion coefficient. To improve both the training and sampling stability, we choose a noise-prediction reparameterization:

$$\epsilon_\theta^{\mathrm{GBM}}(t,x) := -\Sigma(t)(1 + x s_\theta^{\mathrm{GBM}}(t,x)),$$

under which the score-matching objective can be equivalently rewritten as:

$$\min_\theta \mathbb{E}_{t\sim\mathcal{U}[0,T]}\left(\Sigma^2(t)\ \mathbb{E}_{X_0\sim p_{\mathrm{data}}(\cdot),\, Z\sim\mathcal{N}(0,1)}\left|\epsilon_\theta^{\mathrm{GBM}}\left(t, X_0\exp\left(U(t) - \frac{1}{2}\Sigma^2(t) + \Sigma(t)Z\right)\right) - Z\right|^2\right). \tag{4.1}$$

Then the Euler–Maruyama sampler for the backward generation is given by

$$\begin{aligned}
y_{t+\Delta t} &= y_t + ((2\sigma^2(T-t) - \mu(T-t))y_t + \sigma^2(T-t)y_t^2 s_\theta^{\mathrm{GBM}}(T-t, y_t))\Delta t + \sigma(T-t)y_t\sqrt{\Delta t}z_t \\
&= y_t\left(1 + \left(\sigma^2(T-t) - \mu(T-t) - \frac{\sigma^2(T-t)}{\sigma_{T-t}}\epsilon_\theta^{\mathrm{GBM}}(T-t, y_t)\right)\Delta t + \sigma(T-t)\sqrt{\Delta t}z_t\right),
\end{aligned}$$

where $z_t\sim\mathcal{N}(0,1)$ and $y_0\sim p_{\mathrm{noise}}^{\mathrm{GBM}}(\cdot) = \mathrm{LogNormal}(U(T) - \frac{1}{2}\Sigma^2(T), \Sigma^2(T))$.

CIR Processes. We set $\sigma(t) = \sqrt{2\alpha(t)}$ and $\mu(t) \equiv \mu$, so that the stationary distribution is $p_{\text{noise}}^{\text{CIR}}(\cdot) = \mathcal{G}(\mu, 1)$, a Gamma distribution. Let $A(t) := \int_0^t \alpha(s)\,\mathrm{d}s$. The conditional distribution $X_t \,|\, X_0$ is $\frac{1-e^{-A(t)}}{2}K$, where $K \sim \chi^2(2\mu, 2X_0\frac{1}{e^{A(t)}-1})$; see, e.g., [55]. Here, $\chi^2(k, m)$ is the non-central chi-squared distribution with k degrees of freedom and non-centrality parameter m. The denoising score-matching objective can be derived from (3.12) as follows:

$$\min_\theta \mathbb{E}\ \lambda(t) \left| \left(e^{-\frac{1}{2}A(t)}(e^{A(t)}-1) \left(s_\theta^{\text{CIR}}(t, X_t) + \frac{1-\mu}{X_t} \right) + e^{\frac{1}{2}A(t)} \right) \sqrt{X_t} - \frac{\sqrt{X_0} I_\mu \left(\frac{2\sqrt{X_t X_0}}{e^{-\frac{1}{2}A(t)}(e^{A(t)}-1)} \right)}{I_{\mu-1} \left(\frac{2\sqrt{X_t X_0}}{e^{-\frac{1}{2}A(t)}(e^{A(t)}-1)} \right)} \right|^2, \tag{4.2}$$

where $\lambda(t)$ is the weighting function and the expectation is taken over $t \sim \mathcal{U}[0, T]$, $X_0 \sim p_{\text{data}}(\cdot)$, $K \sim \chi^2(2\mu, 2X_0\frac{1}{e^{A(t)}-1})$, and $X_t = \frac{1-e^{-A(t)}}{2}K$. The Euler–Maruyama sampler for the backward generation is given by

$$y_{t+\Delta t} = y_t + \alpha(T-t)(2y_t s_\theta^{\text{CIR}}(T-t, y_t) + 2 - \mu + y_t)\Delta t + \sqrt{2\alpha(T-t)y_t \Delta t} z_t, \tag{4.3}$$

where $z_t \sim \mathcal{N}(0, 1)$ and $y_0 \sim p_{\text{noise}}^{\text{CIR}}(\cdot)$. In our experiments below, we will set $\mu = 1$, $\lambda(t) = e^{-A(t)}$, and adopt the reparameterization $\epsilon_\theta^{\text{CIR}}(t, x) := (1 - e^{-A(t)}) \cdot s_\theta^{\text{CIR}}(t, x) + 1$ to improve training efficiency and stability.

4.1.1. *Image Generation.* We present experimental results on the MNIST dataset. A key property of GBM and CIR processes is that, starting from a non-negative initial condition, the process remains non-negative for all time. Motivated by this property, we consider simplex diffusion models for categorical data [21, 55] which perform the diffusion process in the space of probability simplex, leveraging the discrete nature of the data.

Following [21], we create a discrete version of the MNIST dataset by mapping each pixel value in $\{0, 1, \ldots, 255\}$ to three categories: dark (0–85), medium (86–170), and bright (171–255). Each pixel is then represented as a 3-dimensional real vector. Specifically, we encode a dark/medium/bright pixel as $(a+b, a, a)$, $(a, a+b, a)$, or $(a, a, a+b)$ in $\mathbb{R}_+^3$, respectively, for some $a, b \geq 0$, thereby constructing the training dataset consisting of "dimension-reduced" MNIST images (see Figure 1a). For generated samples, the pixel category is determined by the index of the largest component of the resulting 3-dimensional vector: if the maximum occurs in the first, second, or third dimension, the pixel is classified as dark, medium, or bright, respectively.

Since both GBM and VE exhibit exploding variance and do not admit limiting distributions, whereas both CIR and VP have limiting distributions, we conduct a fair comparison by grouping models with similar properties. Specifically, we compare GBM with VE and CIR with VP, respectively. More details of the experiment can be found in Appendix B.

The visualization results are presented in Figure 1. The GBM-based samples exhibit a noisier background with moderate stroke clarity, where digits are recognizable but show some irregularity in stroke width and curvature. By contrast, the VE-based samples demonstrate cleaner digit boundaries and more uniform stroke thickness across the grid, suggesting that the VE SDE provides a more stable diffusion trajectory that better preserves the structural regularity of handwritten numerals. The sampling error for GBM-based models mainly stems

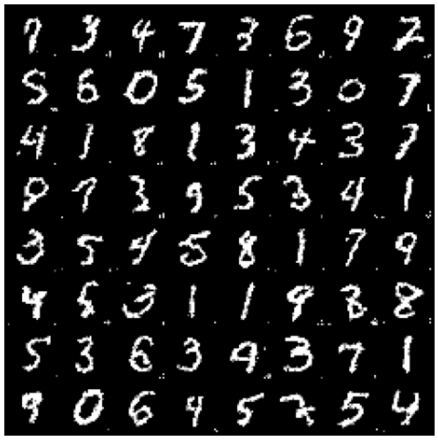
FIGURE 1. Real and Generated MNIST Images

(A) Preprocessed Dataset

(B) GBM-Based Samples

(C) VE-Based Samples

(D) CIR-Based Samples



(E) VP-Based Samples

TABLE 2. Relative FID Scores of Different Diffusion Models

Model	VE	GBM	VP	CIR
Relative FID ↓	1.00	1.28	1.24	0.80

from numerical instability in the SDE sampler during reverse-time sampling, as well as inefficient mixing of the forward process. First, in contrast to the additive Gaussian noise in VE models, the multiplicative exponential Gaussian noise in the GBM forward process can induce large fluctuations in the sampling trajectories, leading to numerical instability. In addition, the state-dependent diffusion coefficient $\sigma(t)x$ introduces a multiplicative dependence on the previous state, y_t, in each iteration, which would further accumulate numerical errors. Nevertheless, these adverse effects are partially alleviated by assigning pixel categories according to the index of the maximum component. Second, the absence of a stationary distribution of GBM leads to a mismatch between the true terminal distribution of the forward process, $p(T, \cdot)$, and the initialization distribution used for reverse-time sampling, $p_{\text{noise}}^{\text{GBM}}(\cdot)$, thereby introducing additional sampling error.

The CIR-based samples demonstrate a clear visual advantage over their VP-based counterparts, exhibiting sharper stroke definition, more consistent digit morphology, and a notably cleaner background across the generated grid. The VP-based samples suffer from higher background noise levels and less precise stroke boundaries, with several digits showing signs of blurring or incomplete formation. These observations suggest the effectiveness of CIR-based models. Notably, as pointed out by [55], using a CIR process as the forward process for simplex diffusion models seems more natural than using VE or VP SDEs, since the normalization of the CIR limiting Gamma distribution is a Dirichlet distribution, which serves as the conjugate prior of the categorical distribution and plays a role somewhat analogous to that of the Gaussian distribution in continuous diffusion models. We note that the Euler–Maruyama sampler (4.3) for the CIR process may suffer from numerical instability in the final few steps, due to the time-dependent diffusion coefficient $\sqrt{2\alpha(t)x}$, in contrast to the VP formulation. Moreover, the VP framework is generally preferred in large-scale experiments because of the simplicity of its score-matching objective, as well as the flexibility in designing loss weighting functions and reverse-time samplers. Improving the training and sampling efficiency and the stability of CIR-based models remains an important direction for future work.

Table 2 reports the relative FID [30] of the four diffusion models, computed based on 2000 generated samples for each model. Since our primary interest lies in the effectiveness of these non-Gaussian models, we focus on relative sample quality rather than absolute performance. Accordingly, we treat the VE model as the baseline by normalizing its FID score as 1.00 and report the scores of the VP-, GBM-, and CIR-based models relative to this baseline. Lower relative FID values indicate better performance. As shown, the CIR-based model achieves the best performance in this context of simplex diffusion models and the FID results are consistent with the visual quality presented in Figure 1. We note that further improvements in overall sample quality would require more sophisticated network architectures and more refined training and sampling strategies for different forward SDEs, which are beyond the scope of this work.

4.1.2. *Financial Time Series for Portfolio Management.* For financial time series, we consider $N = 4$ stocks: AAPL, AMZN, JPM, and TSLA. Let r_t^i denote the log return of stock i at time t, $i \in [N]$. For a consecutive time window $\{t+1, \dots, t+L\}$ of length L, we construct a data point

$$(r_{t+1}^1, \cdots, r_{t+L}^1; r_{t+1}^2, \cdots, r_{t+L}^2; \cdots; r_{t+1}^N, \cdots, r_{t+L}^N) \in \mathbb{R}^{NL}.$$

We set $L = 64$ and let t range over all trading days since January 1, 2010. This yields a dataset with 3,587 data points.

We present results generated by GBM- and VP-based diffusion models for their better empirical performances and relative simplicity [10]. To apply GBM as the forward process, we exponentiate the log returns to obtain positive price ratios:

$$e^{r_t^i} = \frac{p_t^i}{p_{t-1}^i} > 0,$$

where $p_t^i > 0$ is the price of stock i at time t. The resulting data points take the form:

$$(e^{r_{t+1}^1}, \cdots, e^{r_{t+L}^1}; e^{r_{t+1}^2}, \cdots, e^{r_{t+L}^2}; \cdots; e^{r_{t+1}^N}, \cdots, e^{r_{t+L}^N}) \in \mathbb{R}_+^{NL}.$$

Following [27, 28], we evaluate 64-day cumulative log-returns under three portfolio strategies: the equal-weight portfolio, the Markowitz global minimum variance portfolio (GMVP), and the risk-parity portfolio. After converting the generated price ratios to log-returns by taking logarithm, we compare the distribution of the generated samples with that of the real data in terms of mean, standard deviation, quantiles, and overall distributional shape.

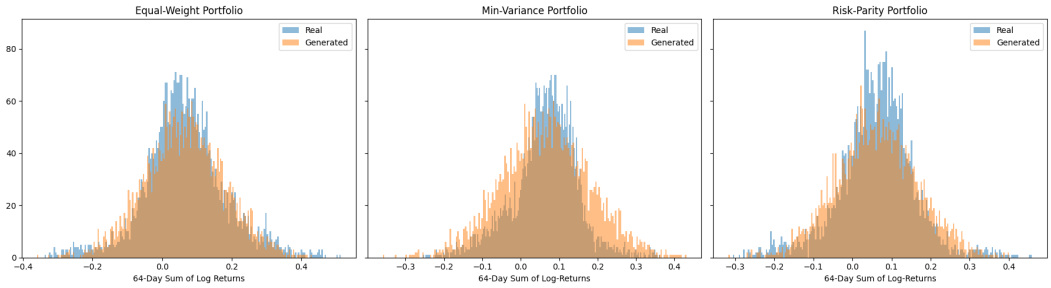
TABLE 3. Real and Generated 64-Day Log-Return Statistics of GBM- and VP-Based Models

	Equal-Weight			GMVP			Risk-Parity		
Statistics	Real	GBM	VP	Real	GBM	VP	Real	GBM	VP
Mean	6.56%	6.52%	6.24%	6.29%	6.51%	6.28%	6.13%	6.14%	6.00%
Median	6.01%	6.54%	6.35%	6.96%	6.55%	6.40%	6.58%	6.15%	6.03%
Std Dev	11.65%	11.41%	8.99%	8.45%	11.41%	8.99%	9.73%	10.35%	7.96%
1% Quantile	-25.20%	-19.55%	-16.83%	-16.91%	-19.55%	-16.79%	-21.02%	-17.99%	-14.06%
5% Quantile	-11.74%	-12.09%	-8.73%	-9.26%	-12.09%	-8.68%	-10.61%	-10.74%	-7.55%
10% Quantile	-6.31%	-7.99%	-5.03%	-4.92%	-7.99%	-4.99%	-5.21%	-7.20%	-3.95%
25% Quantile	-0.08%	-1.05%	0.61%	2.05%	-1.05%	0.66%	1.07%	-0.67%	1.08%

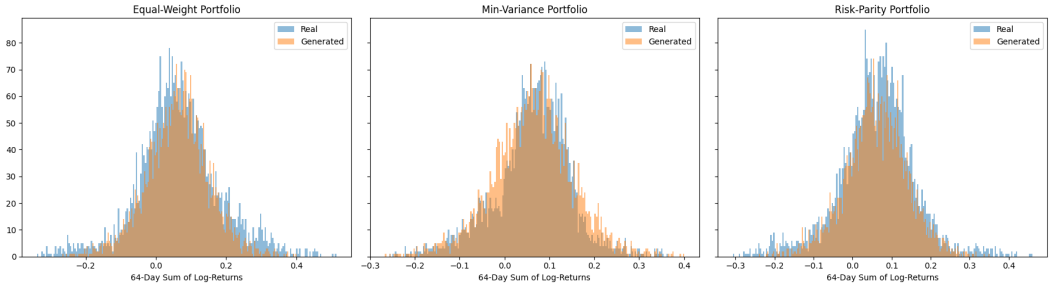
The statistical properties of samples generated by time-dependent GBM- and VP-based models are summarized in Table 3 and Figure 2 presents the histograms of the real and generated 64-day cumulative log-returns under the three portfolio constructions. We observe that the GBM-based model generates log-returns that align more closely with the real data

[10] We also experimented with VE- and CIR-based models. For VE-based models, although the score network can be well-trained, the explosive behavior of the VE process and the high sensitivity of financial data leads to a shifted generated distribution. For CIR-based models, the score matching objective involves modified Bessel functions and multiple exponential terms, leading to high computational complexity and numerical instability, which makes the resulting sampling process unsuitable for this financial setting.

FIGURE 2. Real and Generated 64-Day Sum Log-Returns of Three Portfolios



(A) GBM-Based Samples



(B) VP-Based Samples

in terms of mean and volatility than the VP-based model, particularly for the Equal-Weight and Risk-Parity portfolios. Both models exhibit a modest mismatch with the real data in the tail regions, which is largely attributable to the high sensitivity of financial time series to numerical errors inherent in the simulation pipeline.

Overall, our results, if still preliminary, show that the CIR-based simplex diffusion model excels in MNIST image generation while the GBM-based model outperforms in financial data generation. These findings highlight the potentials and effectiveness of non-Gaussian processes for formulating diffusion models in specific settings and tasks.

4.2. **Empirical Bayes Estimation.** Beyond its importance for diffusion models, Tweedie's formula is also a cornerstone of empirical Bayes methods [15]. Here, we present an application to empirical Bayes estimation based on the BESQ version, which goes beyond Efron's generalized Tweedie's formula. In Appendix A, we also present an empirical Bayes estimation based directly on the GBM, and compare it with Brownian motion (BM) models by taking logarithm.

Suppose there are some large number N of possibly correlated noncentral chi-squared variates $z_i > 0$ have been observed, each with its own unknown noncentrality parameter u_i,

$$z_i \sim \chi^2(3, u_i), \quad i = 1, 2, \cdots, N.$$

We aim to estimate the corresponding u_i values through an empirical Bayes approach. Suppose that u has been sampled from a prior distribution $g(\cdot)$, and then $z \,|\, u \sim \chi^2(3, u)$:

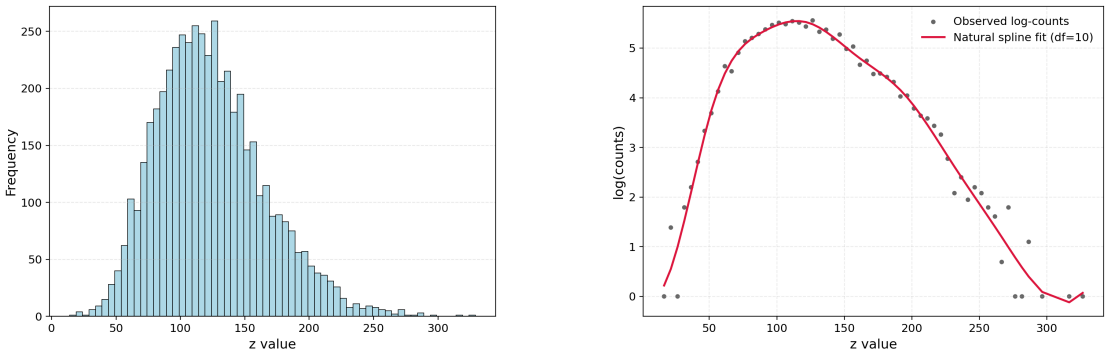
$$u_1, \cdots, u_N \sim g(\cdot), \quad \text{and } z_i \sim \chi^2(3, u_i), \quad i = 1, \cdots, N.$$

Since $X_t \sim t \cdot \chi^2(2(\nu+1), X_0/t)$ for all $t > 0$ in BESQ (3.5), by Tweedie's formula (3.10) (in which $\nu = \frac{1}{2}$, $t = 1$, $X_0 = u$, and $X_1 = z$ are specified), we obtain the posterior expectation of $f(u,z)$ given z as $\mathbb{E}(f(u,z) \,|\, z) = (2s(z)+1)\sqrt{z}$, where $f(u,z) := \sqrt{u}\coth\sqrt{uz}$ and $s(\cdot)$ denotes the score function of the marginal density $p(\cdot)$ of z, i.e., $s(z) = \frac{\mathrm{d}}{\mathrm{d}z}\log p(z)$. Thus, the corresponding empirical Bayes formula is

$$f(\hat{u}_i, z_i) = (2\hat{s}(z_i)+1)\sqrt{z_i} \quad \text{for } i = 1, \cdots, N, \tag{4.4}$$

where $\hat{s}(\cdot)$ is an estimator of the true score $s(\cdot)$, which can be obtained by leveraging the observations $z_1, \cdots, z_N$ through Lindsey's method [14, 15, 16] and $\hat{u}_i$ is an estimate of u_i. Note that $f(u,z)$ is strictly increasing w.r.t. u for any fixed $z > 0$; so there exists a unique $\hat{u}_i$ satisfying (4.4) for a given z_i.

FIGURE 3. Gamma Example



(A) Histogram of 5000 z_i's

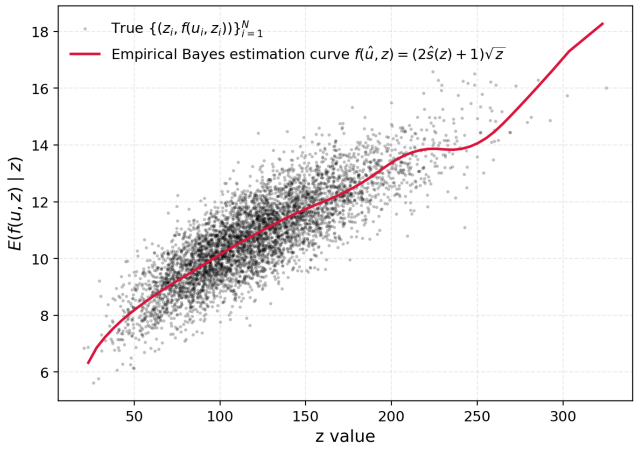
(B) Natural Spline Fit with 10 Degrees of Freedom

We randomly sample $N = 5000$ values of u_i from Gamma distribution $\Gamma(12, 10)$. Figure 3a shows the frequency of 5000 generated z_i's, where there are 63 bins. Denote the center of the k-th bin by x_k and the corresponding bar height by y_k for $k = 1, \ldots, 63$. Figure 3b shows $\log y_k$ against x_k, with bins satisfying $y_k = 0$ excluded, and a natural spline with 10 degrees of freedom is fitted to the points. The derivative of this spline provides a score estimator $\hat{s}(\cdot)$. Figure 4 presents empirical Bayes estimation curve $f(\hat{u}, z) = (2\hat{s}(z)+1)\sqrt{z}$ for the Gamma example data, along with the actual $\{(z_i, f(u_i, z_i))\}_{i=1}^N$ plotted. One can see that the points are closely centered around the estimated curve, even for large and small values of z_i, indicating the effectiveness of the empirical Bayes approach for observations under noncentral chi-squared noise.

5. CONCLUSION

We derive analytically Tweedie's formulae for several important non-Gaussian processes, including geometric Brownian motion, squared Bessel processes and CIR processes, which directly yields denoising score matching objectives for diffusion models based on the corresponding processes. Empirically, we employ Tweedie's formulas derived from GBM and CIR processes for MNIST image generation and financial time series modeling, demonstrating the

FIGURE 4. Empirical Bayes Estimation Curves for BESQ



effectiveness of the resulting denoising score-matching methods. We also apply Tweedie's formula under the BESQ framework to estimate the noncentrality parameter from the noncentral chi-squared noise for empirical Bayes estimation. To sum, we extend original Tweedie's formula to non-Gaussian processes and showcase the promise of non-Gaussian diffusion models for practical applications, opening the gate to their adoption in task-specific domains.

Acknowledgment. Tang is supported by NSF CAREER Award DMS-2538791, the Tang Family Assistant Professorship and a Columbia-CityU/HK collaborative project that is supported by InnoHK Initiative, The Government of the HKSAR and the AIFT Lab. Touzi is partially supported by NSF grant DMS-2508581. Zhang is supported by Tang Fellowship. Zhou acknowledges financial supports through the Nie Center for Intelligent Asset Management at Columbia.

APPENDIX A. EMPIRICAL BAYES ESTIMATION BASED ON GBM

Suppose some large number N of possibly correlated log normal variates $z_i > 0$ have been observed, each with its own unobserved parameter u_i:

$$z_i \sim \text{LogNormal}(u_i, \sigma^2), \quad i = 1, 2, \cdots, N.$$

We estimate the corresponding u_i values through an empirical Bayes approach. Suppose that u has been sampled from a prior distribution $g(\cdot)$, and $z \,|\, u \sim \text{LogNormal}(u, \sigma^2)$ observed with σ^2 known:

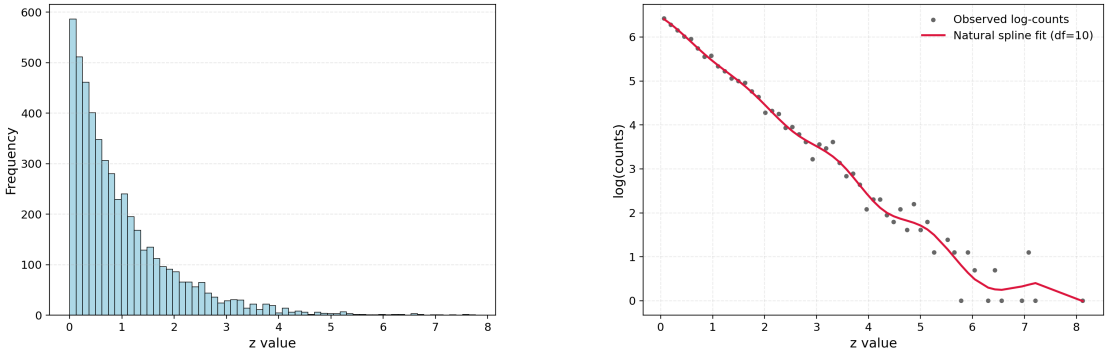
$$u_1, \cdots, u_N \sim g(\cdot), \quad \text{and } z_i \sim \exp(u_i + \sigma\mathcal{N}(0,1)), \quad i = 1, \cdots, N.$$

By Tweedie's formula (3.4) (in which $\mu = \sigma^2/2$, $t = 1$, $X_0 = u$, and $X_1 = z$ are specified), we obtain the posterior expectation of u given z as $\mathbb{E}(u \,|\, z) = \sigma^2 z \cdot s(z) + \sigma^2 + \log z$, where $s(\cdot)$ denotes the score function of the marginal density $p(\cdot)$ of z, i.e., $s(z) = \frac{\mathrm{d}}{\mathrm{d}z} \log p(z)$. Thus, the corresponding empirical Bayes formula is

$$\hat{u}_i = \sigma^2 z_i \cdot \hat{s}(z_i) + \sigma^2 + \log z_i \quad \text{for } i = 1, \cdots, N,$$

where $\hat{s}(\cdot)$ is an estimator to the true score $s(\cdot)$.

FIGURE 5. Exponential Example



(A) Histogram of 5000 z_i's

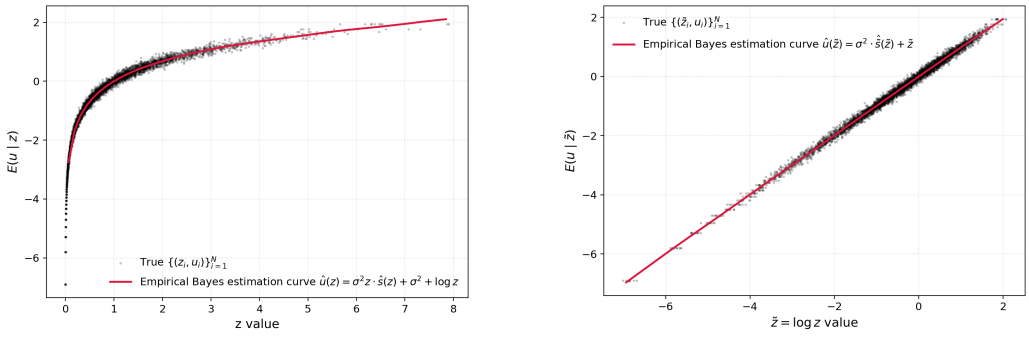
(B) Natural Spline Fit with 10 Degrees of Freedom ($\sigma = 0.1$)

Following the experimental setting in [15], we set $N = 5000$ u_i values as 10 repetitions each of

$$u_i = \log\log\frac{500}{i - 0.5}, \quad i = 1, \cdots, 500.$$

The empirical distribution of e^{u_i} closely matches an exponential distribution with rate 1. Then z_i is generated via LogNormal(u_i, σ^2) for each i. Here, we take $\sigma = 0.1$. Figure 5a shows the frequency of 5000 generated z_i's, where there are 63 bins. Denote the center of the k-th bin by x_k and the corresponding bar height by y_k for $k = 1, \ldots, 63$. Figure 5b shows $\log y_k$ against x_k, with bins satisfying $y_k = 0$ excluded, and a natural spline with 10 degrees of freedom is fitted to the points. The derivative of this spline provides a score estimator $\hat{s}(\cdot)$. Figure 6a presents empirical Bayes estimation curve $\hat{u}(z) = \sigma^2 z \cdot \hat{s}(z) + \sigma^2 + \log z$ for the exponential example data with $\sigma = 0.1$, along with the actual $\{(z_i, u_i)\}_{i=1}^N$ plotted. One can see that the points are closely centered around the estimated curve, even for large values of z_i, indicating the effectiveness of the empirical Bayes approach for observations under log-normal noise.

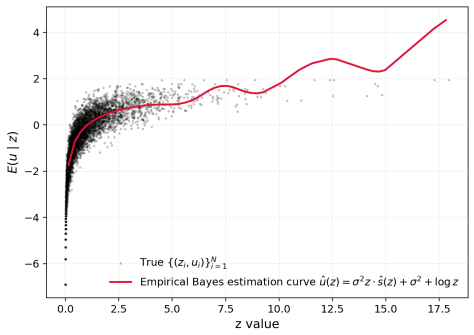
FIGURE 6. Empirical Bayes Estimation Curves ($\sigma = 0.1$)



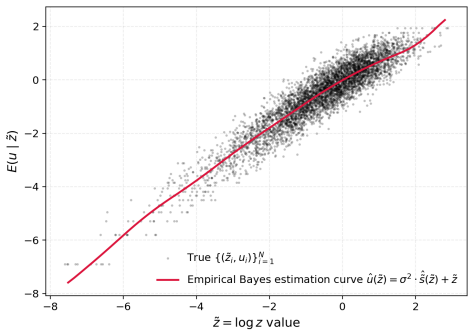
(A) GBM model in z-space

(B) BM model in $(\log z)$-space

FIGURE 7. Empirical Bayes Estimation Curves ($\sigma = 0.5$)

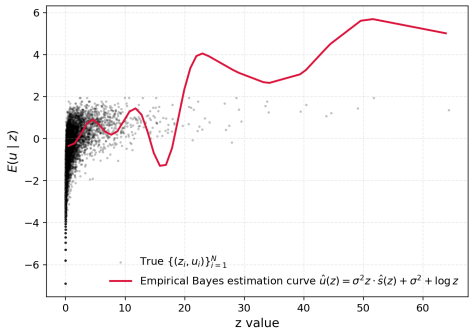


(A) GBM model in z-space

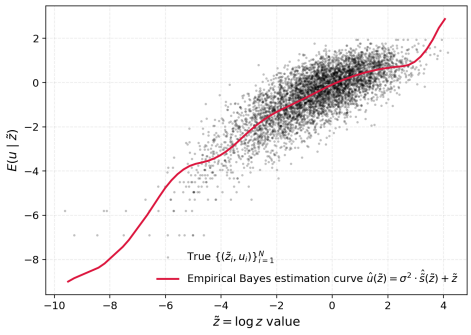


(B) BM model in $(\log z)$-space

FIGURE 8. Empirical Bayes Estimation Curves ($\sigma = 1.0$)



(A) GBM model in z-space



(B) BM model in $(\log z)$-space

We also consider directly applying Tweedie's formula in the $(\log z)$-space to estimate u_i. Let $\tilde{z}_i = \log z_i$ for all $i \in [N]$. Then, equivalently, we observe:

$$\tilde{z}_i \sim \mathcal{N}(u_i, \sigma^2), \quad i = 1, \cdots, N,$$

and aim to estimate u_i. Therefore, we can apply Tweedie's formula for the Gaussian case in the $\tilde{z}$-space to obtain an estimate of u_i:

$$\tilde{u}_i = \sigma^2 \cdot \hat{\tilde{s}}(\tilde{z}_i) + \tilde{z}_i \quad \text{for } i = 1, \cdots, N,$$

where $\tilde{s}(\cdot)$ is the score function of the marginal of $\tilde{z}$ and $\hat{\tilde{s}}(\cdot)$ is an estimator to $\tilde{s}(\cdot)$. Estimating $\tilde{s}(\cdot)$ from $(\tilde{z}_i)_{i=1}^N$ is similar to the procedure described earlier for estimating $s(\cdot)$ from $(z_i)_{i=1}^N$. The estimation results of both models are presented in Figures 6, 7, and 8 for $\sigma = 0.1, 0.5$, and 1.0, respectively. We observe that when σ is larger, the z values can take more extremely small or large values due to the exponential Gaussian noise multiplied on u, which makes fitting a smooth spline directly in the z-space difficult and can lead to substantial estimation errors, particularly for large and small z. By contrast, in the $(\log z)$-space, Gaussian noise is added to u and thus the points are more compactly distributed, making the spline easier to fit. Therefore, for relatively large values of σ (e.g., $\sigma \geq 0.3$), it is preferable to use Tweedie's formula based on the BM model rather than the GBM model for empirical Bayes estimation.

On the other hand, for small σ (e.g., $0 < \sigma \leq 0.3$), both BM and GBM models perform adequately. When σ is very large (e.g., $\sigma \geq 2$), both models yield poor estimates for both large and small z values due to the high noise variance.

Appendix B. Experimental Details

This section provides implementation details of the experiments reported in Section 4.1. Our score networks are parameterized using a U-Net architecture [59] with approximately 1 million parameters, adapted from the implementation available at `https://colab.research.google.com/drive/120kYYBOVa1i0TD85RjlEkFjaWDxSFUx3?usp=sharing`. For training, we employ the Adam optimizer with a learning rate of 1×10^{-4}, along with an exponential moving average with a decay rate of 0.9999.

For the MNIST image generation in Section 4.1.1, we choose the model parameters:

- $\sigma(t) = 0.01 + 1.99t^{3/2}$, $\mu(t) = \sigma(t)^2/2$, $T = 1$, $a = 1.5$, and $b = 0.5$ for GBM (Figure 1b);
- $\sigma(t) = 25^t$, $T = 1$, $a = 0$, and $b = 1$ for VE (Figure 1c);
- $\alpha(t) = 0.05 + 4.95t$, $\mu(t) \equiv 1$, $T = 1$, $a = 0.5$, and $b = 1.5$ for the CIR process (Figure 1d);
- $\alpha(t) = 0.05 + 9.95t$, $T = 1$, $a = 0$, and $b = 1$ for VP (Figure 1e).

Samples are generated by the Euler–Maruyama scheme with 1000 uniform denoising steps. For the financial data generation in Section 4.1.2, we choose:

- $\sigma(t) = 0.001 + 1.999t$, $\mu(t) = \sigma(t)^2/2 - 0.25$, and $T = 1$ for GBM (Figure 2a);
- $\alpha(t) = 0.05 + 1.575t$ and $T = 1$ for VP (Figure 2b).

Samples are generated by Euler–Maruyama with 500 uniform denoising steps.

References

[1] A. Aghapour, E. Bayraktar, and F. Yuan. Solving dynamic portfolio selection problems via score-based diffusion models. 2025. arXiv:2507.09916.

[2] B. D. O. Anderson. Reverse-time diffusion equation models. *Stochastic Process. Appl.*, 12(3):313–326, 1982.

[3] P. Avdeyev, C. Shi, Y. Tan, K. Dudnyk, and J. Zhou. Dirichlet diffusion score model for biological sequence generation. In *ICML*, pages 1276–1301, 2023.

[4] J. Benton, V. D. Bortoli, A. Doucet, and G. Deligiannidis. Nearly d-linear convergence bounds for diffusion models via stochastic localization. In *ICLR*, 2024.

[5] Black Forest Labs. Flux.2: Frontier visual intelligence. `https://bfl.ai/blog/flux-2`, 2025.

[6] H. Chen, H. Lee, and J. Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *ICML*, pages 4735–4763, 2023.

[7] S. Chen, S. Chewi, J. Li, Y. Li, A. Salim, and A. Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *ICLR*, 2023.

[8] J. C. Cox. The constant elasticity of variance option pricing model. *J. Portf. Manag.*, page 15, 1996.

[9] J. C. Cox, J. E. Ingersoll, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.

[10] D. Dai, J. Fan, Y. Gu, and D. Mukherjee. Cindes: Classification induced neural density estimator and simulator. 2025. arXiv:2510.00367.

[11] F. Delbaen and H. Shirakawa. A note on option pricing for the constant elasticity of variance model. *Asia-Pac. Financ. Mark.*, 9:85–99, 2002.

[12] *Digital Library of Mathematical Functions*. `http://dlmf.nist.gov/`, Release 1.2.4 of 2025-03-15.

[13] L. E. Dubins and G. Schwarz. On continuous martingales. *Proc. Natl. Acad. Sci.*, 53(5):913–916, 1965.

[14] B. Efron. Microarrays, empirical Bayes and the two-groups model. *Stat. Sci.*, 2008.

[15] B. Efron. Tweedie's formula and selection bias. *J. Amer. Statist. Assoc.*, 106(496):1602–1614, 2011.

[16] B. Efron. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*. Cambridge University Press, 2012.

[17] B. Efron and N. R. Zhang. False discovery rates and copy number variation. *Biometrika*, 98(2):251–271, 2011.

[18] D. C. Emanuel and J. D. MacBeth. Further results on the constant elasticity of variance call option pricing model. *J. Financ. Quant. Anal.*, 17(4):533–554, 1982.

[19] J. Fan, Y. Gu, and X. Li. Optimal estimation of a factorizable density using diffusion models with relu neural networks. 2025. arXiv:2510.03994.

[20] P. Fitzsimmons, J. Pitman, and M. Yor. Markovian bridges: construction, Palm interpretation, and splicing. In *Seminar on Stochastic Processes, 1992 (Seattle, WA, 1992)*, volume 33 of *Progr. Probab.*, pages 101–134. Birkhäuser Boston, Boston, MA, 1993.

[21] G. Floto, T. Jonsson, M. Nica, S. Sanner, and E. Z. Zhu. Diffusion on the probability simplex. 2023. arXiv:2309.02530.

[22] X. Gao, J. Zha, and X. Y. Zhou. Data-driven generative simulation of SDEs using diffusion models. 2025. arXiv:2509.08731.

[23] Y. Gao, H. Guo, T. Hoang, W. Huang, L. Jiang, F. Kong, H. Li, J. Li, L. Li, and X. Li. Seedance 1.0: Exploring the boundaries of video generation models. 2025. arXiv:2506.09113.

[24] Google. State-of-the-art video and image generation with Veo 2 and Imagen 3. `https://blog.google/technology/google-labs/video-image-generation-update-december-2024/`, 2024.

[25] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, eighth edition, 2015.

[26] J. Gu and R. Koenker. Unobserved heterogeneity in income dynamics: An empirical Bayes perspective. *J. Bus. Econom. Statist.*, 35(1):1–16, 2017.

[27] Z. Guo, J. Li, W. Tang, and D. D. Yao. Diffusion generative models meet compressed sensing, with applications to imaging and finance. 2025. arXiv:2509.03898.

[28] Z. Guo, W. Tang, and R. Xu. Conditional diffusion guidance under hard constraint: a stochastic analysis approach. 2026. arXiv:2602.05533.

[29] U. G. Haussmann and E. Pardoux. Time reversal of diffusions. *Ann. Probab.*, 14(4):1188–1205, 1986.

[30] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Neurips*, volume 30, pages 6629–6640, 2017.

[31] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Neurips*, volume 33, pages 6840–6851, 2020.

[32] A. Hyvärinen. Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.*, 6:695–709, 2005.

[33] N. Ignatiadis and B. Sen. *Empirical Bayes: From Herbert Robbins to modern theory and applications*. 2025. Lecture notes available at `https://nignatiadis.github.io/assets/lecture_notes/Empirical-Bayes.pdf`.

[34] M. Jeanblanc, M. Yor, and M. Chesney. *Mathematical methods for financial markets*. Springer Finance. Springer-Verlag London, Ltd., London, 2009.

[35] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[36] T. Karras, M. Aittala, T. Aila, and S. Laine. Elucidating the design space of diffusion-based generative models. In *Neurips*, volume 35, pages 26565–26577, 2022.

[37] K. Kawazu and S. Watanabe. Branching processes with immigration and related limit theorems. *Teor. Verojatnost. i Primenen.*, 16:34–51, 1971.

[38] S. Khanna, S. Kharbanda, S. Li, H. Varma, E. Wang, S. Birnbaum, Z. Luo, Y. Miraoui, A. Palrecha, and S. Ermon. Mercury: Ultra-fast language models based on diffusion. 2025. arXiv:2506.17298.

[39] G. Kim, S.-Y. Choi, and Y. Kim. A diffusion-based generative model for financial time series via geometric Brownian motion. 2025. arXiv:2507.19003.

[40] J. Lamperti. Continuous state branching processes. *Bull. Amer. Math. Soc.*, 73:382–386, 1967.

[41] H. Lee, J. Lu, and Y. Tan. Convergence for score-based generative modeling with polynomial complexity. In *Neurips*, volume 35, pages 22870–22882, 2022.

[42] G. Li, Y. Wei, Y. Chen, and Y. Chi. Towards faster non-asymptotic convergence for diffusion-based generative models. In *ICLR*, 2024.
[43] A. Liu, M. He, S. Zeng, S. Zhang, L. Zhang, C. Wu, W. Jia, Y. Liu, X. Zhou, and J. Zhou. WeDLM: Reconciling diffusion language models with standard causal attention for fast inference. 2025. arXiv:2512.22737.
[44] H. Liu, T. Zhu, N. Jia, J. He, and Z. Zheng. Learning to simulate from heavy-tailed distribution via diffusion model. 2024. SSRN 4975931.
[45] C. Masreliez. Approximate non-gaussian filtering with linear state and observation relations. *IEEE Trans. Autom. Control*, 20(1):107–110, 1975.
[46] K. Miyasawa. An empirical Bayes estimator of the mean of a normal population. *Bull. Inst. Internat. Statist*, 38(181-188):1–2, 1961.
[47] S. Nie, F. Zhu, Z. You, X. Zhang, J. Ou, J. Hu, J. Zhou, Y. Lin, J.-R. Wen, and C. Li. Large language diffusion models. 2025. arXiv:2502.09992.
[48] K. Oko, S. Akiyama, and T. Suzuki. Diffusion models are minimax optimal distribution estimators. In *ICML*, pages 26517–26582, 2023.
[49] OpenAI. Sora: Creating video from text. `https://openai.com/sora`, 2024.
[50] L. R. Pericchi and A. F. M. Smith. Exact and approximate posterior moments for a normal location parameter. *J. Roy. Statist. Soc. Ser. B*, 54(3):793–804, 1992.
[51] J. Pitman and M. Yor. A decomposition of Bessel bridges. *Z. Wahrsch. Verw. Gebiete*, 59(4):425–457, 1982.
[52] N. G. Polson. A representation of the posterior mean for a location model. *Biometrika*, 78(2):426–430, 1991.
[53] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen. Hierarchical text-conditional image generation with clip latents. arXiv:2204.06125.
[54] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, third edition, 1999.
[55] P. H. Richemond, S. Dieleman, and A. Doucet. Categorical SDEs with simplex diffusion. 2022. arXiv:2210.14784.
[56] H. Robbins. An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I*, pages 157–163. Univ. California Press, Berkeley-Los Angeles, Calif., 1956.
[57] L. Rogers. Which model for term-structure of interest rates should one use? *IMA*, 65:93, 1995.
[58] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pages 10684–10695, 2022.
[59] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241, 2015.
[60] S. Saremi and A. Hyvärinen. Neural empirical Bayes. *J. Mach. Learn. Res.*, 20:Paper No. 181, 23, 2019.
[61] N. Shetty, M. Prasath, and C. S. Seelamantula. Dale meets langevin: A multiplicative denoising diffusion model. 2025. arXiv:2510.02730.
[62] J. Shi, J. Feng, and W. Song. Estimation in linear regression with laplace measurement error using tweedie-type formula. *J. Syst. Sci. Complex.*, 32(4):1211–1230, 2019.
[63] H. Shirakawa. Squared Bessel processes and their applications to the square root interest rate model. *Asia-Pac. Financ. Mark.*, 9(3):169–190, 2002.
[64] U. Singer, A. Polyak, T. Hayes, X. Yin, J. An, S. Zhang, Q. Hu, H. Yang, O. Ashual, and O. Gafni. Make-a-video: Text-to-video generation without text-video data. In *ICLR*, 2023.
[65] Y. Song and S. Ermon. Generative modeling by estimating gradients of the data distribution. In *Neurips*, volume 32, page 11918–11930, 2019.
[66] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021.
[67] C. J. Stone. Optimal rates of convergence for nonparametric estimators. *Ann. Statist.*, 8(6):1348–1360, 1980.
[68] C. J. Stone. Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, 10(4):1040–1053, 1982.
[69] D. W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes*, volume 233 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, 1979.

[70] W. Tang and H. Zhao. Score-based diffusion models via stochastic differential equations. *Statistic Surveys*, 19:28–64, 2025.
[71] W. Tang and H. Zhao. Contractive diffusion probabilistic models. 2026. To appear in SIAM J. Imaging Sci.
[72] S. Torres. Tweedie calculus. 2026. arXiv:2604.14486.
[73] P. Vincent. A connection between score matching and denoising autoencoders. *Neural Comput.*, 23(7):1661–1674, 2011.
[74] J. Ye, Z. Xie, L. Zheng, J. Gao, Z. Wu, X. Jiang, Z. Li, and L. Kong. Dream 7b: Diffusion large language models. 2025. arXiv:2508.15487.
[75] Z. Zhao, C. Yeh, L. Kong, and K. Wang. Diffusion-DFL: decision-focused diffusion models for stochastic optimization. In *ICLR*, 2026.

Department of Industrial Engineering and Operations Research, Columbia University.
Email address: wt2319@columbia.edu

Department of Finance and Risk Engineering, New York University.
Email address: nt2635@nyu.edu

Department of Industrial Engineering and Operations Research, Columbia University.
Email address: zz3367@columbia.edu

Department of Industrial Engineering and Operations Research, Columbia University.
Email address: xz2574@columbia.edu