

Embedded Variational Neural Stochastic Differential Equations for Learning Heterogeneous Dynamics

Sandeep Kumar^{*1,2}, Reema Gupta^{†1}, and S. Chakraverty^{‡1,2}

¹Department of Mathematics, National Institute of Technology, Rourkela, India

²Poverty Alleviation Research Centre, National Institute of Technology, Rourkela, India

Abstract

This study examines the challenges of modeling complex and noisy data related to socioeconomic factors over time, with a focus on data from various districts in Odisha, India. Traditional time-series models struggle to capture both trends and variations together in this type of data. To tackle this, a Variational Neural Stochastic Differential Equation (V-NSDE) model is designed that combines the expressive dynamics of Neural SDEs with the generative capabilities of Variational Autoencoders (VAEs). This model uses an encoder and a decoder. The encoder takes the initial observations and district embeddings and translates them into a Gaussian distribution, which determines the mean and log-variance of the first latent state. Then the obtained latent state initiates the Neural SDE, which utilize neural networks to determine the drift and diffusion functions that govern continuous-time latent dynamics. These governing functions depend on the time index, latent state, and district embedding, which help the model learn the unique characteristics specific to each district. After that, using a probabilistic decoder, the observations are reconstructed from the latent trajectory. The decoder outputs a mean and log-variance for each time step, which follows the Gaussian likelihood. The Evidence Lower Bound (ELBO) training loss improves by adding a KL-divergence regularization term to the negative log-likelihood (nll). The obtained results demonstrate the effective learning of V-NSDE in recognizing complex patterns over time, yielding realistic outcomes that include clear trends and random fluctuations across different areas.

Keywords: Latent Neural SDE, Variational , Socioeconomic, Dynamical system, KL divergence.

1 Introduction

Socioeconomic development over time is a complex, dynamic process. The structural trends, short-term policies, and uncertain shocks can often impact this process. In this context, socioeconomic indicators that affect household poverty are continually evolving and are becoming increasingly diverse across regions. Due to some geographical, administrative, infrastructural, and historical factors, these indicators pose nonlinear trajectories, fluctuations, and disparities. In socioeconomically diverse states like Odisha, India, capturing such dynamics in a principled and uncertainty-aware manner remains a central challenge [1].

*sandeepksamota@gmail.com

†reema270898@gmail.com

‡Corresponding author:sne_chak@yahoo.com

Traditional approaches to model such indicators mostly rely on discrete-time econometric and statistical methods such as regression, panel data, and autoregressive techniques [2, 3, 4]. For example, Dang and Lanjouw [5] analyzed poverty dynamics in India by constructing the panel dataset from various surveys. Cheng et al. [6] employed a system dynamics model, a simulation model used to understand complex phenomena. They used it to explore the relationship between environment, disasters and poverty. Łatuszyńska and Fate [7] presented a hybrid model combining system dynamics and agent-based simulation to study the impact of public intervention on poverty. More recently, Kumar et al. [8] constructed a wealth index using principal component analysis for a district in Odisha, utilizing primary data collected, and compared the outcomes with a similar index from the national family health survey (NFHS) surveys.

These methods impose strong parametric assumptions, depend on linear or weak nonlinear relationships, and operate at fixed observation intervals. Moreover, Indian household data obtained from surveys like the district-level household and facility survey (DLHS) and NFHS are often found to be sparse, irregularly sampled, and noisy. It limits the effectiveness of traditional time-series models. Furthermore, these limitations become more evident when modeling multiple indicators with an interacting nature over a long time, where uncertainty and latent influences are crucial.

In this regard, stochastic differential equations (SDEs) can provide a mathematical framework for modeling such systems [9]. They can decompose the problem into two parts: deterministic and stochastic. By doing this, SDEs can model both drift trends and the stochastic nature of the socioeconomic problem. Therefore, it can be said that SDEs allow continuous-time modeling, uncertainty quantification, and theoretical interpretability. However, classical SDE models require the drift and diffusion functions to be specified in advance using some simplistic parametric forms that may be inadequate for capturing the high-dimensional and nonlinear interaction characteristics of real-world socioeconomic systems [10]. Furthermore, parameter estimation in multivariate SDEs can quickly become intractable as model complexity increases.

The developments in deep learning (DL) techniques have led to the advancement of continuous-time dynamical systems governed by neural networks [11]. Beginning with neural ordinary differential equations (Neural ODEs), which parameterize deterministic continuous-time dynamics using neural networks and enable end-to-end training through differentiable ODE solvers [12]. This model establishes a conceptual bridge between DL and dynamical systems theory. It provides a principled approach to modeling nonlinear temporal dynamics without introducing any discretization. Also, parallel developments such as neural controlled differential equations (Neural CDEs), which basically treat observed time series as control signals [13]. These drive latent continuous-time dynamics and are particularly effective for irregularly sampled data.

Building upon this foundation, neural stochastic differential equations (Neural SDEs) extend Neural DEs-based methodologies by incorporating stochasticity into the dynamics through learned diffusion terms [14]. Neural SDEs can effectively model the randomness, shocks, and variability present in socioeconomic dynamics [15]. They utilize the expressive power of NNs with the mathematical rigor of stochastic calculus, enabling data-driven discovery of complex continuous-time dynamics. More recent work focuses on scalable and amortized variational inference for latent SDEs, addressing computational and statistical challenges associated with learning continuous-time stochastic models from sparse and noisy observations [16, 17]. Neural SDE models have demonstrated remarkable success in modelling applications related to finance, biology, and physics, where systems are governed by stochastic processes [18]. Despite all, the application of neural SDEs in

modeling socioeconomic dynamics remains limited and underexplored.

A major challenge in learning neural SDEs from real data is the intractability of the likelihood function. Variational inference offers a solution to this problem by approximating the intractable posterior distribution over latent trajectories with a tractable variational family. Variational neural stochastic differential equations (V-NSDEs) enable efficient learning of latent continuous-time processes and provide uncertainty-aware predictions [19]. When combined with latent variable formulations, variational inference models allow high-dimensional observed indicators to be governed by a lower-dimensional latent stochastic process.

Here, we will utilize district-level socioeconomic data related to poverty in the state of Odisha, India. Odisha exhibits substantial inter-district heterogeneity across all socioeconomic indicators, such as education, housing quality, sanitation access, etc. Historically marginalized regions like the KBK (Kalahandi–Balangir–Koraput) have different developmental trajectories compared to coastal and other urbanized districts. Available survey measurements of these indicators are available at multi-year intervals and are subject to compounding uncertainty. These studies motivate the need to develop a modeling framework that can (i) explicitly handle irregular temporal sampling, (ii) quantify predictive uncertainty, and (iii) provide interpretable latent structure properties that variational latent SDEs naturally offer. A V-NSDE framework can allow such observations to be embedded within a continuous-time stochastic model. It can also reconstruct the latent trajectory and principal uncertainty quantification at the district level.

In addition to temporal uncertainty and nonlinear dynamics, district-level socioeconomic modeling should understand and incorporate spatial heterogeneity, institutional capacity, historical deprivation, and infrastructural endowments. To address this, district embedding vectors can be incorporated into the V-NSDE framework. These vectors can be learned during training the model and can act as a condition on the drift and diffusion functions. These embeddings act as continuous representations of district-specific characteristics, which enables the model to share statistical strength across districts while simultaneously capturing systematic inter-district differences.

In this work, we advance the V-NSDE framework for modeling socioeconomic dynamics by developing a latent continuous-time formulation. This method incorporates district-specific embeddings for multivariate district-level data. Our approach learns nonlinear drift and diffusion functions that govern the evolution of latent socioeconomic states, while explicitly accounting for observational noise and irregular sampling. The proposed model can provide a reliable tool for analyzing long-term socioeconomic processes under uncertainty.

We will validate the proposed method using district-level socioeconomic indicators from Odisha. Further, we will demonstrate that V-NSDE can effectively capture heterogeneous developmental patterns across districts. Beyond the specific case of Odisha, the proposed framework offers a general methodological foundation for continuous-time modeling of socioeconomic systems in data-constrained and uncertain environments.

The remainder of the paper is organized as follows. Section 2 presents the mathematical formulation of neural stochastic differential equations and the associated variational inference framework. Section 3 describes the proposed latent model structure and learning methodology. Section 4 details the district-level Odisha dataset and experimental setup. Section 5 discusses the empirical results and their socioeconomic implications. Finally, Section 6 concludes the paper and outlines

directions for future research.

2 Mathematical Preliminaries

This section provides the fundamental concepts required for the formulation of V-NSDE.

2.1 Probability Space and Random Variables

Definition 2.1 ([20]). If Ω denotes the sample space, $\mathcal{F}$ is a σ -algebra of measurable events, and $\mathbb{P}$ is a probability measure satisfying $\mathbb{P}(\Omega) = 1$, then triplet $(\Omega, \mathcal{F}, \mathbb{P})$ is called **probability space**.

Definition 2.2 ([20]). If $(\Omega, \mathcal{F}, \mathbb{P})$ is probability space, then the map

$$X : \Omega \rightarrow \mathbb{R}^m.$$

is called **random variable**.

The probability distribution of X can be characterized by probability mass function (pmf) in the discrete case and probability density function (pdf) in the continuous case. Now, we will define expectation of random variable X .

Definition 2.3 ([20]). The **expectation** of a random variable X is defined as

$$\mathbb{E}[X] = \begin{cases} \sum_x x \mathbb{P}(X = x), & \text{discrete case,} \\ \int_{\mathbb{R}} x p(x) dx, & \text{continuous case,} \end{cases}$$

where $p(x)$ is pdf of X .

More generally, the expectation of a function $g(X)$ is defined as

$$\mathbb{E}_{p(x)}[g(X)] = \int g(x) p(x) dx.$$

It follows linearity condition,

$$\mathbb{E}[aX + bY] = a \mathbb{E}[X] + b \mathbb{E}[Y],$$

for any random variables X, Y and constants $a, b \in \mathbb{R}$.

Definition 2.4 ([20]). The **variance** of a random variable X is defined as

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

The variance measures the dispersion around the mean.

Definition 2.5 ([20]). If X and Y are random variables, then **conditional probability** of X given $Y = y$ is defined as

$$p(x | y) = \frac{p(x, y)}{p(y)}, \quad p(y) > 0.$$

2.2 Bayes' Rule

Bayes' rule provides a fundamental relationship between prior knowledge, observed data, and posterior inference. It forms the mathematical backbone of Bayesian statistics and variational inference.

Let x be a observed data and z an unobserved (latent) random variable. Then, the joint distribution of (x, z) is factorized as

$$p(x, z) = p(x | z) p(z),$$

here $p(z)$ is the prior distribution over the latent variable and $p(x | z)$ is the likelihood function.

Bayes' rule expresses the posterior distribution $p(z | x)$ as [20]

$$p(z | x) = \frac{p(x | z) p(z)}{p(x)}$$

where the denominator

$$p(x) = \int p(x | z) p(z) dz$$

is known as the marginal likelihood or evidence, which is often intractable for complex models.

Definition 2.6 ([21]). The **KL divergence** between two probability distributions $p(x) \sim P$ and $q(x) \sim Q$ is defined as

$$D_{KL}(P||Q) = \int p(x) \log \frac{p(x)}{q(x)} dx.$$

It satisfies $KL(P||Q) \geq 0$ and is zero if and only if $P = Q$ almost everywhere. It measures similarity between the probability distributions.

2.3 Monte Carlo Approximation

Some expectations are analytically intractable. For those Monte Carlo sampling approximation are done as

$$\mathbb{E}_{p(X)}[g(X)] \approx \frac{1}{N} \sum_{i=1}^N g(x_i), \quad x_i \sim p(X).$$

Monte Carlo estimation plays a central role in training probabilistic deep learning models like our which includes VAEs.

2.4 Brownian Motion

Definition 2.7 ([22]). A **stochastic process** is a time parameterized collection of random variables

$$\{X_t\}_{t \geq 0},$$

defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2.8 ([22]). A d -dimensional stochastic process $\{W_t\}_{t \geq 0}$ is called standard **Brownian motion** (or Wiener process), if it satisfies

1. $W_0 = 0$ almost surely,

2. W_t has independent increments,
3. $W_t - W_s \sim \mathcal{N}(0, (t - s)I_d)$ for $0 \leq s < t$,
4. W_t has almost surely continuous paths.

2.5 Itô SDEs

For $X_t \in \mathbb{R}^m$, Itô SDE is defined as [22]

$$dX_t = f(X_t, t) dt + g(X_t, t) dW_t, \quad X_0 = x_0,$$

where f is the drift function, g is the diffusion function, and W_t is an m -dimensional Brownian motion. In the SDE, the drift term models deterministic dynamics and the diffusion term captures stochastic fluctuations.

The SDE can be equivalently written in integral form as

$$X_t = X_0 + \int_0^t f(X_s, s) ds + \int_0^t g(X_s, s) dW_s,$$

where the stochastic integral is interpreted in the Itô sense.

3 V-NSDE Methodology

This section presents the essential methodological components of the V-NSDE used in the current study. Firstly, the detailed forward architecture is explained which emphasizes the structure and functionality of the model. Subsequently, the loss formulation is discussed in terms of its importance and influence on model performance. Finally, the process of backward propagation is discussed, illustrating how it optimizes the learning mechanism and enhances the overall effectiveness of the model.

3.1 Forward Architecture Components

The forward part of the model consists of four components: Encoder, Latent NSDE, and Decoder. In the subsequent subsections each one is systematically described. The schematic diagram of the V-NSDE method is presented in Fig. (1).

3.1.1 District Embedding

To capture the district specific characteristics, embedding technique is utilized. It converts discrete inputs into low-dimensional numerical vectors. In this regard, each district is associated with a learnable, fixed-size embedding vector $\mathbf{e}_d \in \mathbb{R}^m$. These embeddings serve to personalize the VAE components for each district, allowing the model to capture district-specific characteristics that influence the observed socio-economic indicators and their temporal dynamics.

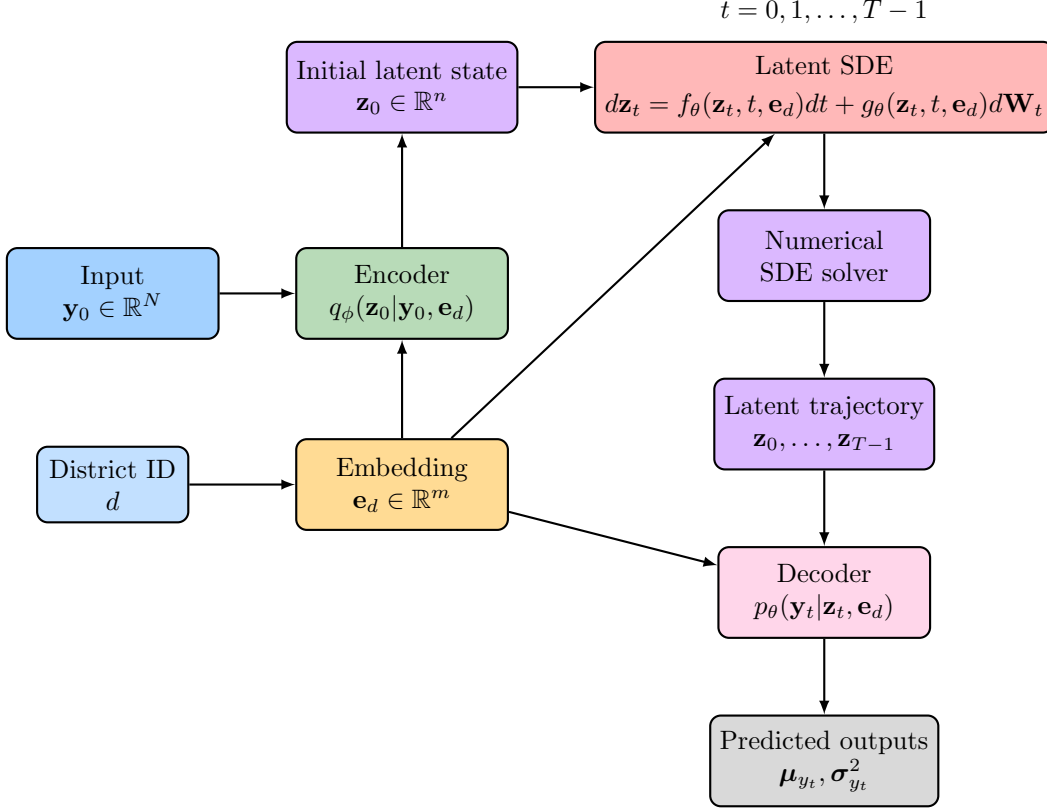


Figure 1: Block diagram of the proposed V-NSDE architecture for district d .

3.1.2 Encoder: $q_\phi(\mathbf{z}_0|\mathbf{y}_0, \mathbf{e}_d)$

The encoder is a deep neural network (DNN) whose primary role is to establish the initial probabilistic latent state $\mathbf{z}_0 \in \mathbb{R}^n$ from the very first observation of a time series $\mathbf{y}_0 \in \mathbb{R}^N$ and its corresponding static contextual information, the district embedding $\mathbf{e}_d \in \mathbb{R}^m$. As given in Eq. (1), it models the approximate posterior distribution $q_\phi(\mathbf{z}_0|\mathbf{y}_0, \mathbf{e}_d)$ as a diagonal Gaussian distribution over $\mathbf{z}_0$.

$$q_\phi(\mathbf{z}_0|\mathbf{y}_0, \mathbf{e}_d) = \mathcal{N}(\boldsymbol{\mu}_{z_0}, \text{diag}(\exp(\log \boldsymbol{\sigma}_{z_0}^2))) \quad (1)$$

The Encoder takes the concatenated initial observation and district embedding as input and output the mean $\boldsymbol{\mu}_{z_0}$ and log-variance $\log \boldsymbol{\sigma}_{z_0}^2$ as described in Eq. (2).

$$\begin{aligned} [\boldsymbol{\mu}_{z_0}, \log \boldsymbol{\sigma}_{z_0}^2] &= \text{Encoder}([\mathbf{y}_0; \mathbf{e}_d]), \\ \mathbf{z}_0 &= \boldsymbol{\mu}_{z_0} + \boldsymbol{\sigma}_{z_0} \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \end{aligned} \quad (2)$$

where $\odot$ denotes element-wise multiplication and $\boldsymbol{\epsilon}$ is a sample from a standard normal distribution for reparameterization trick.

The Encoder compresses the initial high-dimensional observation into a lower-dimensional, probabilistic latent representation that captures its essential features. The sampled $\mathbf{z}_0$ then serves as the starting point for the subsequent SDE trajectory. The probabilistic nature allows the model to capture uncertainty in the initial state.

3.1.3 Latent NSDE

The latent NSDE models the continuous-time evolution of the latent state $\mathbf{z}_t$ given its initial value $\mathbf{z}_0$ (sampled from the encoder) and the static district embedding $\mathbf{e}_d$. It captures both deterministic dynamics and stochastic fluctuations within the latent space. The temporal dynamics of the latent state $\mathbf{z}_t \in \mathbb{R}^n$ are governed by an Ito SDE defined in Eq. (3), conditioned on time t and the district embedding $\mathbf{e}_d$.

$$d\mathbf{z}_t = f_\theta(\mathbf{z}_t, t, \mathbf{e}_d)dt + g_\theta(\mathbf{z}_t, t, \mathbf{e}_d)d\mathbf{W}_t. \quad (3)$$

Here, $f_\theta : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ is the drift function and $g_\theta : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ is the diagonal diffusion function. Both f_θ and g_θ are parameterized by DNNs. $\mathbf{W}_t$ represents a standard Brownian motion. In a diagonal noise SDE, each dimension of the latent space has its own independent Wiener process, meaning $\mathbf{W}_t$ is a vector of independent random increments.

We assume that the neural drift and diffusion functions, augmented with district embeddings, satisfy global Lipschitz and linear growth conditions. Under these assumptions, the latent NSDE admits a unique strong solution for each district.

Given the initial latent state $\mathbf{z}_0$ and the parameterized drift and diffusion functions, the latent SDE is numerically integrated over the time interval $[0, T]$ using an SDE solver to get the solution given as in Eq. (4). This process yields the full latent trajectory $\{\mathbf{z}_t\}_{t=0}^T$.

$$\mathbf{z}_t = \mathbf{z}_0 + \int_0^t f_\theta(\mathbf{z}_s, s, \mathbf{e}_d)ds + \int_0^t g_\theta(\mathbf{z}_s, s, \mathbf{e}_d)d\mathbf{W}_s \quad (4)$$

The Latent SDE is the heart of the temporal modeling. It allows the model to learn complex, non-linear, and stochastic dependencies between successive latent states, ensuring that the generated latent trajectories are plausible and reflect the underlying dynamics of the real-world time series. The inclusion of $\mathbf{e}_d$ allows the dynamics to be specific to each district.

3.1.4 Decoder: $p_\theta(\mathbf{y}_t|\mathbf{z}_t, \mathbf{e}_d)$

The Decoder is also a DNN model that acts as the generative network, taking the evolved latent state $\mathbf{z}_t$ from the NSDE and the district embedding $\mathbf{e}_d$ to reconstruct the observed data $\mathbf{y}_t$. As described in Eq. (5), it models the reconstruction output by the likelihood $p_\theta(\mathbf{y}_t|\mathbf{z}_t, \mathbf{e}_d)$ as a diagonal Gaussian distribution.

$$p_\theta(\mathbf{y}_t|\mathbf{z}_t, \mathbf{e}_d) = \mathcal{N}(\mathbf{y}_t; \boldsymbol{\mu}_{y_t}, \text{diag}(\exp(\log \boldsymbol{\sigma}_{y_t}^2))) \quad (5)$$

The Decoder outputs the mean μ_{y_t} and log-variance $\log \sigma_{y_t}^2$ that takes the latent state and district embedding as input as given in Eq. (6).

$$[\boldsymbol{\mu}_{y_t}, \log \boldsymbol{\sigma}_{y_t}^2] = \text{Decoder}([\mathbf{z}_t; \mathbf{e}_d]) \quad (6)$$

For each latent state $\mathbf{z}_t$ along the trajectory, the decoder computes the parameters of a Gaussian distribution over the observed features $\mathbf{y}_t$. The inclusion of $\mathbf{e}_d$ allows the reconstruction to be contextualized by the specific district characteristics.

3.2 Backward Model Architecture Components

3.2.1 Evidence Lower Bound (ELBO) Loss Function Derivation

The primary objective when training the V-NSDE model is to maximize the marginal likelihood of the observed data, $p_\theta(\mathbf{Y})$, where $\mathbf{Y} = \{\mathbf{y}_t\}_{t=0}^T$ is the entire observed time series for a given district.

The marginal likelihood $p_\theta(\mathbf{Y})$ can be written by integrating over the latent trajectory $\mathbf{Z} = \{\mathbf{z}_t\}_{t=0}^T$:

$$p_\theta(\mathbf{Y}) = \int p_\theta(\mathbf{Y}|\mathbf{Z})p_\theta(\mathbf{Z})d\mathbf{Z}. \quad (7)$$

However, directly maximizing $p_\theta(\mathbf{Y})$ is intractable due to the complex NN model. Instead, we optimize a lower bound on the marginal likelihood, known as the Evidence Lower Bound (ELBO). The ELBO can be derived using Jensen’s inequality and properties of logarithms.

We introduce the approximate posterior $q_\phi(\mathbf{Z}|\mathbf{Y})$, which is approximated by $q_\phi(\mathbf{z}_0|\mathbf{y}_0, \mathbf{e}_d)$ and the SDE dynamics [23]. Therefore, Eq. (7) becomes

$$p_\theta(\mathbf{Y}) = \int q_\phi(\mathbf{Z}|\mathbf{Y}) \frac{p_\theta(\mathbf{Y}|\mathbf{Z})p_\theta(\mathbf{Z})}{q_\phi(\mathbf{Z}|\mathbf{Y})} d\mathbf{Z}. \quad (8)$$

Now, taking the logarithm and applying Jensen’s inequality to Eq. (8) as log is a concave function, the following inequality can be obtained:

$$\log p_\theta(\mathbf{Y}) \geq \mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{Y})} \left[\log \frac{p_\theta(\mathbf{Y}|\mathbf{Z})p_\theta(\mathbf{Z})}{q_\phi(\mathbf{Z}|\mathbf{Y})} \right]. \quad (9)$$

This lower bound in Eq. (9) is the ELBO, which further can be simplified as:

$$\text{ELBO}(\phi, \theta; \mathbf{Y}) = \mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{Y})} [\log p_\theta(\mathbf{Y}|\mathbf{Z})] - \mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{Y})} \left[\log \frac{q_\phi(\mathbf{Z}|\mathbf{Y})}{p_\theta(\mathbf{Z})} \right]. \quad (10)$$

Recognizing the second term in Eq. (10) as the KL divergence, $D_{KL}(q_\phi(\mathbf{Z}|\mathbf{Y})||p_\theta(\mathbf{Z}))$, the ELBO can be written as:

$$\text{ELBO}(\phi, \theta; \mathbf{Y}) = \mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{Y})} [\log p_\theta(\mathbf{Y}|\mathbf{Z})] - D_{KL}(q_\phi(\mathbf{Z}|\mathbf{Y})||p_\theta(\mathbf{Z})). \quad (11)$$

According to Eq. (9), maximizing ELBO is equivalent to maximize the likelihood $p_\theta(\mathbf{Y})$. Further, it is equivalent to minimizing the negative ELBO during training, given by:

$$\mathcal{L}(\phi, \theta; \mathbf{Y}) = -\text{ELBO}(\phi, \theta; \mathbf{Y}) = \underbrace{-\mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{Y})} [\log p_\theta(\mathbf{Y}|\mathbf{Z})]}_{\text{Reconstruction Loss (NLL)}} + \underbrace{\beta D_{KL}(q_\phi(\mathbf{Z}|\mathbf{Y})||p_\theta(\mathbf{Z}))}_{\text{KL Divergence Term}}. \quad (12)$$

Here, reconstruction loss term measures how well the decoder can reconstruct the observed data $\mathbf{Y}$ from the sampled latent trajectory $\mathbf{Z}$. In Gaussian likelihood setting, this typically translates to minimizing a form of Mean Squared Error or, more accurately, the Gaussian Negative Log-Likelihood (NLL). β represents beta annealing, which is a technique recommended to prevent “posterior collapse” by gradually increasing the weight of the KL divergence term during training.

The KL divergence term measures the difference between the approximate posterior distribution $q_\phi(\mathbf{Z}|\mathbf{Y})$, (the distribution over latent states inferred by the encoder and SDE dynamics) and the prior distribution $p(\mathbf{Z})$ (the assumed distribution over latent states, typically a simple one like a standard normal distribution). Minimizing this term encourages the approximate posterior to stay close to the prior, preventing the latent space from becoming too complex or irregular.

3.2.2 Weight Updates with Adjoint Method

Training a deep learning model involves adjusting its parameters (weights and biases) to minimize a loss function. This is typically achieved using gradient-based optimization techniques like Stochastic Gradient Descent (SGD) or Adam, which require computing the gradients of the loss with respect to all model parameters via backpropagation.

In the V-NSDE model, the latent trajectory $\mathbf{Z} = \{\mathbf{z}_t\}_{t=0}^T$ is generated by numerically integrating an SDE. Standard backpropagation through reverse-mode automatic differentiation (AD) would need to store the entire history of intermediate computations (activations and gradients) from the SDE solver at each time step. For SDEs simulated over many time steps, this can lead to an enormous memory footprint and computational cost, making training impractical.

In this regard, the adjoint sensitivity method provides an efficient way to compute gradients through DE solvers without storing all intermediate states. Instead of backpropagating through each step of the SDE solver, it solves a coupled backward-in-time DE, known as the adjoint equation. This adjoint equation describes how the gradients of the loss function with respect to the latent state evolve backward in time. The gradients with respect to the SDE parameters (f_θ and g_θ 's weights) can then be computed by integrating another auxiliary DE that accumulates these gradients along the backward pass. Therefore, the adjoint sensitivity method is used for memory-efficient computation of gradients through the SDE solver.

Overall Backpropagation Flow

1. **Decoder Gradients:** Using AD, gradients are first computed for the parameters (θ) of decoder based on the Gaussian NLL reconstruction loss, with respect to its inputs ($\mathbf{z}_t$ and $\mathbf{e}_d$). This provides the initial gradients $\partial L / \partial \mathbf{z}_t$ and $\partial L / \partial \mathbf{e}_d$.
2. **SDE Adjoint Gradients:** The gradients $\partial L / \partial \mathbf{z}_t$ for each time point are then passed to the adjoint SDE solver. The adjoint method efficiently computes $\partial L / \partial \mathbf{z}_0$ (gradients with respect to the initial latent state) and $\partial L / \partial \theta_f$, $\partial L / \partial \theta_g$ (gradients with respect to the drift and diffusion network parameters) by integrating the adjoint system backward in time. This is crucial for memory efficiency, especially with long time series.
3. **Encoder Gradients:** Finally, $\partial L / \partial \mathbf{z}_0$ is used to compute gradients for the parameters (ϕ) of encoder with respect to $\boldsymbol{\mu}_{z_0}$ and $\log \boldsymbol{\sigma}_{z_0}^2$ through the reparameterization trick, which then backpropagate through the encoder network to update ϕ .
4. **District Embedding Gradients:** Gradients for the district embedding $\mathbf{e}_d$ are accumulated from the encoder, SDE, and decoder, and then used to update in the embedding layer.

This end-to-end backpropagation process, facilitated by adjoint methods for the SDE part, allows the entire VAE-SDE model to be trained effectively using gradient descent.

4 Experimental Setup

In this section, we will define and solve a socioeconomic dynamical problem utilizing the described method.

4.1 Problem Description

The dataset consists of monthly observations over 168 months ($T = 168$) for six socioeconomic indicators across 30 districts in Odisha, India. These indicators include access to electricity, education attainment, housing availability, piped clean drinking water, clean cooking fuel, and improved sanitation facilities. In this regard, data was collected for year 2007 from District Level Household and Facility Survey (DLHS)-3 [24] and for year 2015 and 2020 from National Family Health Survey (NFHS)-4 and NFHS-5 [25] regarding the said socioeconomic indicators.

To make it a stochastic dynamical system, month-wise synthetic data between survey years were generated using a Brownian bridge interpolation scheme, which preserves the observed annual survey values while allowing stochastic fluctuations between them. This approach avoids unrealistic linear interpolation and reflects unobserved short-term socioeconomic variability.

A Brownian Bridge from (t_a, x_a) to (t_b, x_b) can be defined by Eq. (13) [9]:

$$x_t = x_a + \frac{t - t_a}{t_b - t_a}(x_b - x_a) + \sigma \left[W(t - t_a) - \frac{t - t_a}{t_b - t_a} W(t_b - t_a) \right], \quad (13)$$

where $W(t)$ be standard Brownian motion, $\sigma > 0$ controls the volatility.

4.2 Computational Environment Settings

This subsection will cover all the settings applied in the methodology, such as input, latent, and district embedding dimensions, probabilistic distribution for encoder and decoder, architecture details of all involved NNs, etc.

To begin with, the embedding layer consisting of 16 neurons ($m = 16$) is initialized for each district ID. There are six socioeconomic variables ($N = 6$) affecting the poverty status of households in a district, which include

1. Percentage of households that have an electricity connection,
2. Percentage of population obtained education at level 10 or more,
3. Percentage of households that own a concrete-made house,
4. Percentage of households that have a piped clean water connection in their yard,
5. Percentage of households that use clean cooking fuel (gas facility in the kitchen),
6. Percentage of households that have improved sanitation facilities (toilet with flush).

The encoder takes 22 input neurons, which include six socioeconomic indicators and 16 district embedding vectors, two hidden layers with SiLU activation, and an output layer with linear activation of 8 neurons, which includes 4-dimensional ($n = 4$) latent vectors for mean and variance.

In the latent-NSDE, the drift f_θ and diffusion g_θ networks consider 20 input neurons (latent + district embedding), two hidden layers of 64 neurons with SiLU activation, and an output layer with four neurons. Linear activation is utilized in the drift network, and tanh activation is used in the diffusion network to control the magnitude of diffusion.

Finally, the decoder takes 20 input neurons, which include four latent dimension vectors and 16 district embedding vectors, two hidden layers with SiLU activation, and an output layer with linear activation of 12 neurons, which includes 6-dimensional socioeconomic dimension vectors for mean and variance.

Prior distribution

The standard normal distribution is the most common choice for the prior distribution over the latent space, given by

$$p(\mathbf{z}_0) = \mathcal{N}(\mathbf{z}_0; \mathbf{0}, \mathbf{I}), \quad (14)$$

where $\mathbf{0}$ is a vector of zeros and $\mathbf{I}$ is the identity matrix, implying a mean of zero and unit variance for each independent latent dimension.

Negative log likelihood estimation

The NLL from Eq. (12) can be written as

$$-\mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{Y})}[\log p_\theta(\mathbf{Y}|\mathbf{Z})] = - \int \log p_\theta(\mathbf{Y}|\mathbf{Z}) q_\phi(\mathbf{Z}|\mathbf{Y}) d\mathbf{Z}. \quad (15)$$

This integral is intractable, therefore using Monte Carlo approximation we can write Eq. (15) as

$$\mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{Y})}[-\log p_\theta(\mathbf{Y}|\mathbf{Z})] \approx \frac{1}{L} \sum_{l=0}^{L-1} -\log p_\theta(\mathbf{Y}|\mathbf{Z}^l), \quad (16)$$

where $\mathbf{Z}^l \sim q_\phi(\mathbf{Z}|\mathbf{Y})$. In practice, single unbiased Monte Carlo estimator is used, as a result final NLL looks likes

$$\begin{aligned} \mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{Y})}[-\log p_\theta(\mathbf{Y}|\mathbf{Z})] &\approx -\log p_\theta(\mathbf{Y}|\mathbf{Z}) \\ &= \frac{1}{2NT} \sum_{t=0}^{T-1} \sum_{j=1}^N \left[\log(2\pi) + \log \sigma_{y_{t,j}}^2 + \frac{(y_{t,j} - \mu_{y_{t,j}})^2}{\sigma_{y_{t,j}}^2} \right] \end{aligned} \quad (17)$$

The constant term $\log(2\pi)$ is often omitted during optimization as it does not affect the gradients. Therefore, the practical NLL is simplifies to

$$\mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{Y})}[-\log p_\theta(\mathbf{Y}|\mathbf{Z})] \approx \frac{1}{2NT} \sum_{t=0}^{T-1} \sum_{j=1}^N \left[\log \sigma_{y_{t,j}}^2 + \frac{(y_{t,j} - \mu_{y_{t,j}})^2}{\sigma_{y_{t,j}}^2} \right], \quad (18)$$

where $y_{t,j}$ is the actual observed value for feature j at time t , $\mu_{y_{t,j}}$ is the predicted mean for feature j at time t by the decoder, and $\log \sigma_{y_{t,j}}^2$ is the predicted log-variance for feature j at time t by the decoder.

This loss function encourages the decoder to predict the correct mean and learn appropriate variances for uncertainty estimates in the predictions. If the model is very confident and incorrect, the NLL will be high. If it is uncertain and correct, the NLL will be lower.

4.2.1 KL divergence term calculation

In the VAE framework, the KL divergence term in the ELBO encourages the approximate posterior distribution $q_\phi(\mathbf{z}|\mathbf{x})$ to be close to a predefined prior distribution $p(\mathbf{z})$. Here in V-NSDE model, this KL divergence specifically applies to the initial latent state $\mathbf{z}_0$.

In our case, $q_\phi(\mathbf{z}_0|\mathbf{y}_0, \mathbf{e}_d)$ and $p(\mathbf{z}_0)$ follows distributions defined in Eq. (1, 14). Therefore, the closed form is given by

$$D_{KL}(q_\phi(\mathbf{z}_0|\mathbf{y}_0, \mathbf{e}_d)||p(\mathbf{z}_0)) = \frac{1}{2n} \sum_{j=1}^n [-\log \sigma_{z_0,j}^2 - 1 + \sigma_{z_0,j}^2 + \mu_{z_0,j}^2], \quad (19)$$

where $\sigma_{z_0,j}^2$ denotes the j -th element of $\exp(\log \sigma_{z_0}^2)$.

In our specific V-NSDE model, the KL divergence term is calculated only for the initial latent state $\mathbf{z}_0$ because the subsequent evolution of $\mathbf{z}_t$ is governed by the SDE dynamics, which defines the continuous-time prior over the latent trajectory $p(\mathbf{Z})$. Therefore, the KL divergence term specifically focuses on $D_{KL}(q_\phi(\mathbf{z}_0|\mathbf{y}_0, \mathbf{e}_d)||p_\theta(\mathbf{z}_0))$. The reconstruction loss, however, spans the entire time series, comparing the predicted $\mathbf{y}_t$ distributions with the observed $\mathbf{y}_t$ for all t . The total ELBO loss considered in the model is given by Eqs. (18)-(19) as

$$\begin{aligned} \mathcal{L}(\phi, \theta; \mathbf{Y}) = & \frac{1}{2NT} \sum_{t=0}^{T-1} \sum_{j=1}^N \left[\log \sigma_{y_t,j}^2 + \frac{(y_{t,j} - \mu_{y_t,j})^2}{\sigma_{y_t,j}^2} \right] \\ & + \beta \cdot \frac{1}{2n} \sum_{j=1}^n [-\log \sigma_{z_0,j}^2 - 1 + \sigma_{z_0,j}^2 + \mu_{z_0,j}^2]. \end{aligned} \quad (20)$$

In the backpropagation as described in Section 3.2, for the optimization propose we have considered Adam optimizer with default learning rate of 0.001 [26]. The model was trained for 1000 epochs till the ELBO loss got stabilized.

5 Results and Discussion

This section presents the results obtained from the trained V-NSDE model applied to the stochastic socioeconomic indicator data of Odisha. The model demonstrates stable and effective training behavior.

5.1 Model Training and Convergence

The V-NSDE model exhibits stable and well-behaved training dynamics. The ELBO loss decreases smoothly within 1000 training epochs, indicating effective optimization and numerical stability. At the end of training, the final ELBO value -10.746209 is obtained, demonstrating a satisfactory balance between data reconstruction accuracy and latent space regularization.

The ELBO objective is primarily dominated by the Gaussian NLL term that reflects accurate reconstruction of the observed socioeconomic trajectories. The KL divergence term increases gradually during training due to the application of β -annealing and later gets stable. Which explains that it successfully prevented the posterior collapse and encouraged meaningful latent representations.

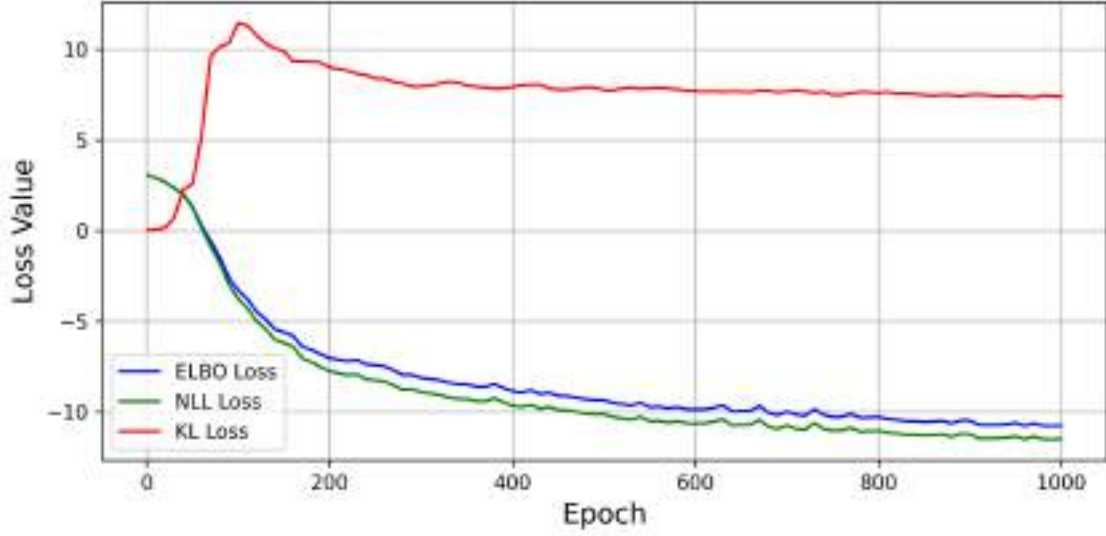


Figure 2: Training losses over epochs

As shown in Fig. (2), the overall convergence behavior confirms that the V-NSDE model is well-trained and suitable for capturing the stochastic dynamics present in district-level socioeconomic data.

5.2 District-wise Quantitative Model Fit Using NLL

To quantitatively assess model performance across districts, as depicted in Table 1, the Gaussian NLL was computed separately for each of the 30 districts of Odisha. For each district, the trained V-NSDE model predicted the mean and log-variance of the time series. The NLL is evaluated by comparing the actual normalized observations against these probabilistic predictions obtained from the V-NSDE. These results highlight the heterogeneous predictive performance of the model across districts. These results also justify the necessity of district-specific latent embedding representations within a single V-NSDE modeling framework.

5.3 Mean Trajectory Prediction and Temporal Trend Preservation

Across all considered socioeconomic variables (education, housing, sanitation, drinking water access, clean cooking fuel, and electricity), the predicted mean trajectories show strong visual agreement with the observed district-level data. The V-NSDE model consistently preserves long-term temporal trends, accurately reflecting increasing or decreasing patterns over the study period. For visualization of such, the figures of three KBK (Koraput, Balangir, and Kalahandi) region districts are shown in Figures 3–5.

The V-NSDE model successfully captures key turning points in the trajectories, which may be due to some reasons such as policy interventions, survey periods, or structural shifts. This demonstrates the ability of the stochastic latent dynamics to adapt to both smooth evolutions and volatility in the data trend.

A major advantage of the proposed V-NSDE framework lies in its principled uncertainty quantification. The model outputs mean μ and log-variance $\log \sigma^2$ as shown by Eq. (6). Therefore, we have also predicted $\pm 2\sigma$ confidence intervals around the mean trajectories which effectively

Table 1: NLL performance across districts on trained model

District	NLL	District	NLL
Rayagada	-12.599152	Debagarh	-11.589890
Puri	-12.499263	Nayagarh	-11.562165
Nuapada	-12.492243	Sonapur	-11.537984
Bhadrak	-12.302701	Ganjam	-11.523965
Cuttack	-12.268328	Balangir	-11.468771
Malkangiri	-12.147593	Baleshwar	-11.392648
Khandamal	-12.119408	Kendrapara	-11.343279
Kendujhar	-11.975780	Kalahandi	-11.311354
Nabarangapur	-11.891778	Jagatsinghpur	-11.011625
Sundargarh	-11.889838	Gajapati	-10.979132
Sambalpur	-11.875984	Baudh	-10.938087
Koraput	-11.737288	Dhenkanal	-10.704522
Bargarh	-11.737288	Jajapur	-10.615693
Khordha	-11.681852	Angul	-10.251976
Mayurbhanj	-11.663747	Jharsuguda	-10.069868
Total ELBO = Mean NLL + $\beta \times$ Mean KL divergence			
-10.746209 = -11.487712 + 0.1 $\times$ 7.415025			

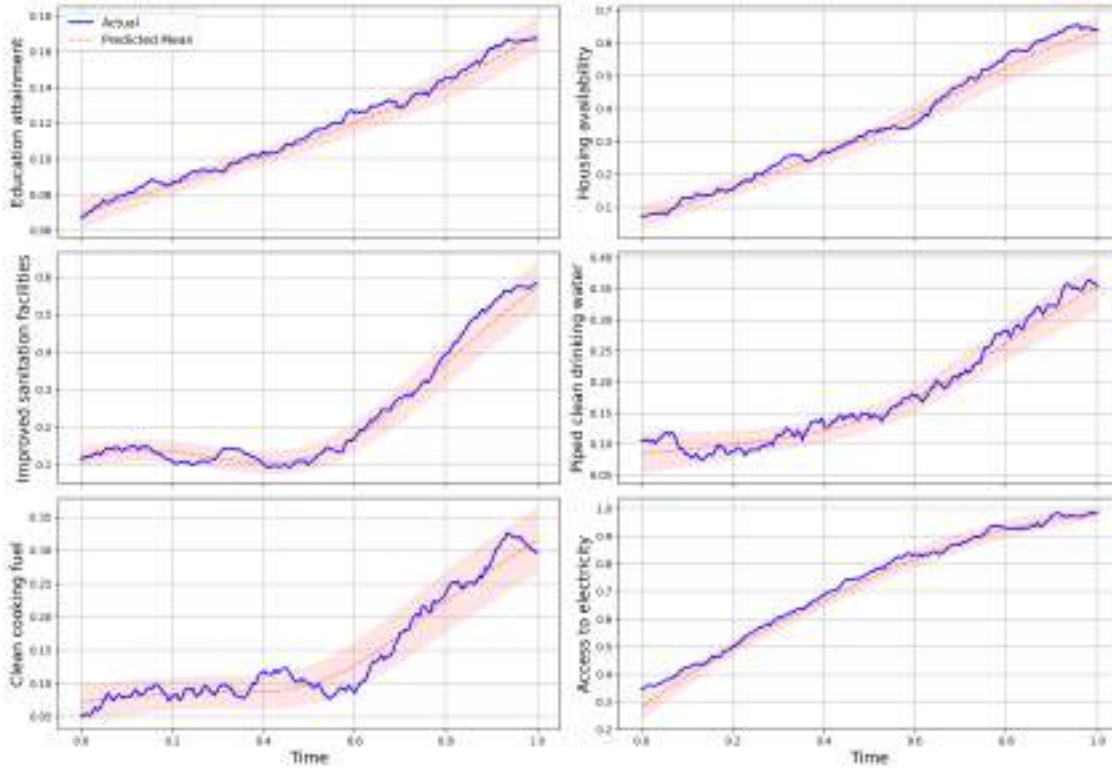


Figure 3: Likelihood predictions for Balangir District

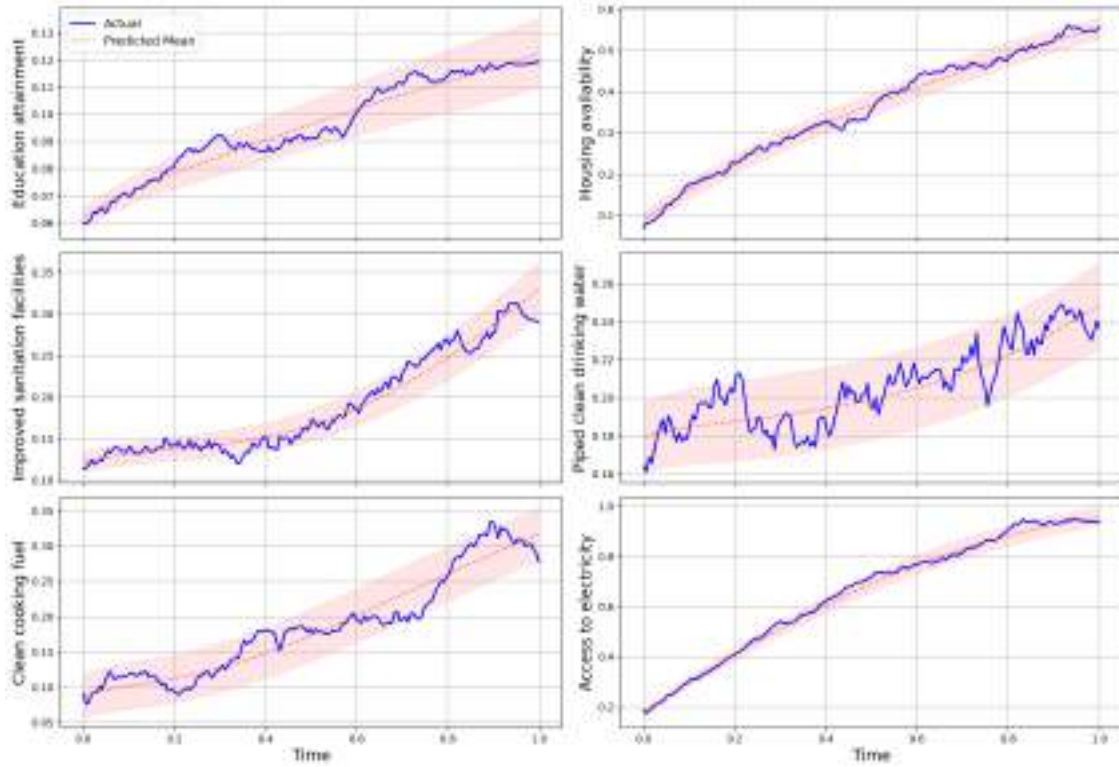


Figure 4: Likelihood predictions for Koraput District

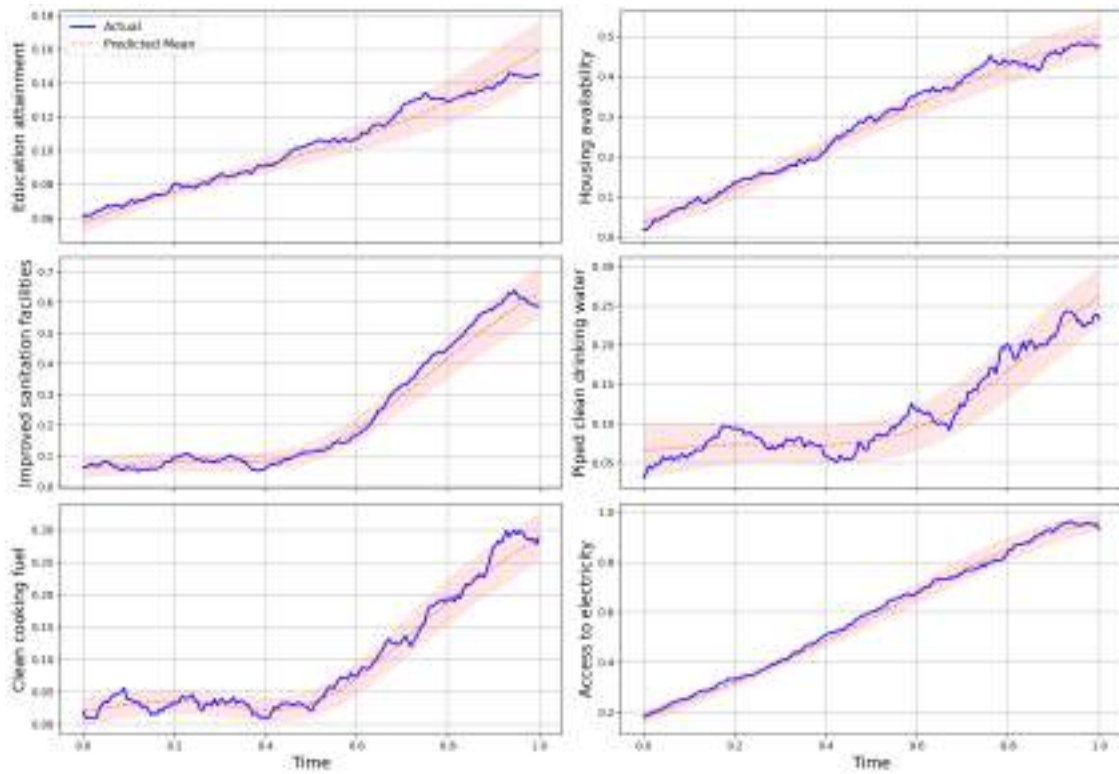


Figure 5: Likelihood predictions for Kalahandi District

captures the inherent stochasticity and data uncertainty present in the socioeconomic indicators.

Interestingly, almost all observed data points lie within the predicted uncertainty bands, with only rare instances of observations falling outside these intervals. Also, wider predictive intervals are consistently observed during data-sparse periods and in phases exhibiting higher volatility. This pattern aligns with real-world expectations for irregularly sampled and noisy socioeconomic data. This behavior confirms that the model avoids overconfident predictions and provides realistic probabilistic forecasts.

The trained model also reveals clear district-level heterogeneity in both drift behavior and volatility structure. Different districts show distinct temporal evolution patterns which reflect variations in socioeconomic development pathways. For example, backward and less-developed KBK region districts, display higher predictive uncertainty and increased volatility in drinking water, cooking fuel, and sanitation facilities indicator as depicted in Figures 3, 4, and 5.

These qualitative yet consistent patterns across districts show that the V-NSDE is able to effectively learn district-specific characteristics through latent embeddings. It allows a single unified model to represent diverse regional dynamics without requiring separate district-wise models.

The empirical analysis yields several important insights:

1. Education-related indicators exhibit relatively smoother temporal dynamics with lower predictive uncertainty across districts.
2. Housing and sanitation indicators show higher volatility, reflecting sensitivity to infrastructural and policy-driven changes.
3. Backward districts consistently display wider predictive intervals, indicating greater uncertainty and structural variability.
4. The district-wise NLL analysis provides a clear quantitative measure of model fit, complementing the qualitative trajectory and uncertainty assessments.

The results demonstrate that the proposed V-NSDE framework effectively captures mean dynamics, uncertainty structure, and district-level heterogeneity in socioeconomic time series. This makes the model a robust and interpretable tool for modeling complex and irregularly sampled socioeconomic processes.

6 Conclusion

In this study, we proposed a V-NSDE framework for modeling the temporal dynamics of multiple socioeconomic indicators related to household characteristics. The proposed approach formulates socioeconomic evolution as a continuous-time stochastic dynamical system, explicitly accounting for uncertainty arising from measurement errors, policy shocks, and unobserved heterogeneity.

The V-NSDE framework integrates VAEs to project high-dimensional, irregularly sampled socioeconomic data into a low-dimensional latent space, where the underlying dynamics are governed by a Neural Stochastic Differential Equation with neural network-parameterized drift and diffusion functions. This latent representation significantly reduces model complexity while preserving

essential structural information. The latent SDE is solved using standard numerical SDE solvers, and the observed variables are reconstructed through a decoder network. Model training is carried out by maximizing the Evidence Lower Bound (ELBO), providing a fully probabilistic learning objective that balances reconstruction accuracy and latent regularization.

Empirical results using district-level data from Odisha, India, demonstrate that the proposed model effectively captures both the long-term trends and short-term stochastic fluctuations of key socioeconomic indicators. In particular, the predicted mean trajectories closely follow the observed temporal patterns across all districts, while the associated predictive uncertainty—quantified through the estimated standard deviation bands—successfully encloses the observed data. This indicates that the model not only achieves accurate point predictions but also provides meaningful uncertainty estimates. Notably, stable convergence was achieved within 250 training epochs, with the ELBO reaching values around -11, highlighting the computational efficiency and practical feasibility of the proposed framework.

Beyond the specific case study of Odisha, the proposed V-NSDE framework is general and transferable. It is well suited to a broad class of socioeconomic and development-related dynamical problems, particularly in settings where data are scarce, noisy, or irregularly sampled, as is common in household surveys and administrative records in developing regions. By enabling continuous-time modeling with uncertainty quantification, the approach supports more reliable policy analysis, forecasting, and scenario evaluation.

Despite its advantages, the current study has certain limitations. The model assumes time-invariant district embeddings and does not explicitly incorporate exogenous covariates such as policy interventions or macroeconomic shocks. Moreover, model performance may be sensitive to the choice of latent dimension and network architecture, which were selected empirically.

Future work will focus on extending the framework to include time-varying embeddings, exogenous control variables, and hierarchical spatial dependencies across districts. Additionally, integrating causal constraints and exploring policy-counterfactual simulations within the V-NSDE framework represent promising directions. These extensions will further enhance the applicability of the proposed methodology for evidence-based socioeconomic planning and decision-making.

A Existence and Uniqueness of Latent NSDEs with District Embeddings

A.1 Probabilistic Setup

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ be a complete filtered probability space satisfying the usual conditions, and

$$W_t = (W_t^{(1)}, \dots, W_t^{(m)})$$

be an m -dimensional standard Brownian motion adapted to the filtration $\{\mathcal{F}_t\}$. All stochastic processes considered are assumed to be adapted to this filtration.

A.2 Latent NSDE with District Embeddings

We consider the latent NSDE as

$$\begin{aligned} dZ_t^{(d)} &= f_\theta(Z_t^{(d)}, t, e_d) dt + g_\theta(Z_t^{(d)}, t, e_d) dW_t^{(d)}, \\ Z_0^{(d)} &\sim q_\phi(Z_0 | Y_0^{(d)}, e_d), \end{aligned} \quad (21)$$

where $Z_t^{(d)} \in \mathbb{R}^k$ is the latent state for district d , $e_d \in \mathbb{R}^r$ is a fixed district embedding vector, $f_\theta : \mathbb{R}^k \times [0, T] \times \mathbb{R}^r \rightarrow \mathbb{R}^k$ is the neural drift, $g_\theta : \mathbb{R}^k \times [0, T] \times \mathbb{R}^r \rightarrow \mathbb{R}^{k \times m}$ is the neural diffusion, $W_t^{(d)}$ is an m -dimensional Brownian motion, and q_ϕ is an encoder-defined initial latent distribution.

A.3 Assumptions

Assumption A (Embedding Regularity). District embeddings are bounded, i.e.

$$\|e_d\| \leq M_e < \infty, \quad \forall d. \quad (22)$$

Assumption B (Measurability and Adaptedness). The functions $f_\theta(z, t, e_d)$ and $g_\theta(z, t, e_d)$ are Borel measurable in $(\mathbb{R}^k; [0, T]; \mathbb{R}^r)$, progressively measurable with respect to $\{\mathcal{F}_t\}$, and continuous in $\mathbb{R}^k$ for almost every t .

Assumption C (Global Lipschitz Continuity). There exists a constant $L > 0$ such that for all $z_1, z_2 \in \mathbb{R}^k$, all $t \in [0, T]$, and all embeddings e_d ,

$$\|f_\theta(z_1, t, e_d) - f_\theta(z_2, t, e_d)\| + \|g_\theta(z_1, t, e_d) - g_\theta(z_2, t, e_d)\| \leq L\|z_1 - z_2\|. \quad (23)$$

Assumption D (Linear Growth Condition). There exists $C > 0$ such that

$$\|f_\theta(z, t, e_d)\|^2 + \|g_\theta(z, t, e_d)\|^2 \leq C(1 + \|z\|^2 + \|e_d\|^2), \quad \forall z, t, d. \quad (24)$$

Assumption E (Finite Second Moment of Initial Condition). The initial latent variable satisfies

$$\mathbb{E}\|Z_0^{(d)}\|^2 < \infty. \quad (25)$$

A.3.1 Existence and Uniqueness Theorem

Theorem A.1 (Existence and Uniqueness of Latent NSDE). *Under Assumptions A–E, for each district d , the latent NSDE in Eq. (21) admits a **unique strong solution***

$$\{Z_t^{(d)}\}_{t \in [0, T]}$$

such that $Z_t^{(d)}$ is $\mathcal{F}_t$ -adapted and almost surely continuous and $\mathbb{E} \left[\sup_{t \in [0, T]} \|Z_t^{(d)}\|^2 \right] < \infty$.

Proof. In this proof we will follow standard arguments for Itô SDEs with parameterized coefficients. Therefore, Eq. (21) is equivalent to the following Itô integral

$$Z_t^{(d)} = Z_0^{(d)} + \int_0^t f_\theta(Z_s^{(d)}, s, e_d) ds + \int_0^t g_\theta(Z_s^{(d)}, s, e_d) dW_s^{(d)}. \quad (26)$$

Using Picards iteration, we define the sequence $\{Z^{(n)}\}_{n \geq 0}$ as

$$\begin{aligned} Z_t^{(0)} &= Z_0^{(d)}, \\ Z_t^{(n+1)} &= Z_0^{(d)} + \int_0^t f_\theta(Z_s^{(n)}, s, e_d) ds + \int_0^t g_\theta(Z_s^{(n)}, s, e_d) dW_s^{(d)}. \end{aligned} \quad (27)$$

To prove existence of solution, we will show that the Picard iteration is a Cauchy in $L^2(\Omega; C([0, T]; \mathbb{R}^k))$.

Existence of strong solution. From the Picard iteration, we have

$$\begin{aligned} Z_t^{(n+1)} - Z_t^{(n)} &= \int_0^t [f_\theta(Z_s^{(n)}, s, e_d) - f_\theta(Z_s^{(n-1)}, s, e_d)] ds \\ &\quad + \int_0^t [g_\theta(Z_s^{(n)}, s, e_d) - g_\theta(Z_s^{(n-1)}, s, e_d)] dW_s. \end{aligned}$$

Taking supremum over $t \in [0, T]$, expectation and using $\|a + b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$, we obtain

$$\begin{aligned} \mathbb{E} \left[\sup_{t \leq T} \|Z_t^{(n+1)} - Z_t^{(n)}\|^2 \right] &= 2\mathbb{E} \left[\sup_{t \leq T} \left\| \int_0^t [f_\theta(Z_s^{(n)}, s, e_d) - f_\theta(Z_s^{(n-1)}, s, e_d)] ds \right\|^2 \right] \\ &\quad + 2\mathbb{E} \left[\sup_{t \leq T} \left\| \int_0^t [g_\theta(Z_s^{(n)}, s, e_d) - g_\theta(Z_s^{(n-1)}, s, e_d)] dW_s \right\|^2 \right]. \end{aligned}$$

This can also be written as

$$\mathbb{E} \left[\sup_{t \leq T} \|Z_t^{(n+1)} - Z_t^{(n)}\|^2 \right] = 2\mathbb{E} \left[\sup_{t \leq T} \left\| \int_0^t \Delta f_s ds \right\|^2 \right] + 2\mathbb{E} \left[\sup_{t \leq T} \left\| \int_0^t \Delta g_s dW_s \right\|^2 \right]. \quad (28)$$

where

$$\Delta f_s = f_\theta(Z_s^{(n)}, s, e_d) - f_\theta(Z_s^{(n-1)}, s, e_d), \quad \Delta g_s = g_\theta(Z_s^{(n)}, s, e_d) - g_\theta(Z_s^{(n-1)}, s, e_d).$$

For drift term integral, using Cauchy-Schwarz inequality we get

$$\mathbb{E} \left[\sup_{t \leq T} \left\| \int_0^t \Delta f_s ds \right\|^2 \right] \leq T \cdot \mathbb{E} \left[\sup_{t \leq T} \int_0^t \|\Delta f_s\|^2 ds \right].$$

For the stochastic integral, the Itô Isometry inequality yields

$$\mathbb{E} \left[\sup_{t \leq T} \left\| \int_0^t \Delta g_s dW_s \right\|^2 \right] \leq \mathbb{E} \left[\sup_{t \leq T} \int_0^t \|\Delta g_s\|^2 ds \right]$$

Therefore Eq. (28) becomes

$$\begin{aligned} \mathbb{E} \left[\sup_{t \leq T} \|Z_t^{(n+1)} - Z_t^{(n)}\|^2 \right] &\leq 2T \cdot \mathbb{E} \left[\sup_{t \leq T} \int_0^t \|\Delta f_s\|^2 ds \right] + 2\mathbb{E} \left[\sup_{t \leq T} \int_0^t \|\Delta g_s\|^2 ds \right] \\ &\leq 2(T+1) \left(\mathbb{E} \left[\sup_{t \leq T} \int_0^t \|\Delta f_s\|^2 ds + \sup_{t \leq T} \int_0^t \|\Delta g_s\|^2 ds \right] \right) \\ &= 2(T+1) \left(\mathbb{E} \left[\sup_{t \leq T} \int_0^t (\|\Delta f_s\|^2 + \|\Delta g_s\|^2) ds \right] \right). \end{aligned} \quad (29)$$

Using the global Lipschitz continuity of Eq. (23), we obtain

$$\begin{aligned}
\mathbb{E} \left[\sup_{t \leq T} \|Z_t^{(n+1)} - Z_t^{(n)}\|^2 \right] &\leq 2L^2(T+1) \mathbb{E} \left[\sup_{t \leq T} \int_0^t \|Z_s^{(n)} - Z_s^{(n-1)}\|^2 ds \right] \\
&\leq 2L^2(T+1) \mathbb{E} \left[\int_0^T \|Z_s^{(n)} - Z_s^{(n-1)}\|^2 ds \right] \\
&= 2L^2(T+1) \int_0^T \mathbb{E} \left[\|Z_s^{(n)} - Z_s^{(n-1)}\|^2 \right] ds \\
&\leq 2L^2T(T+1) \mathbb{E} \left[\sup_{t \leq T} \|Z_t^{(n)} - Z_t^{(n-1)}\|^2 \right].
\end{aligned} \tag{30}$$

Choosing $T > 0$ such that $\alpha = 2L^2T(T+1) < 1$ gives

$$\mathbb{E} \left[\sup_{t \leq T} \|Z_t^{(n+1)} - Z_t^{(n)}\|^2 \right] \leq \alpha \mathbb{E} \left[\sup_{t \leq T} \|Z_t^{(n)} - Z_t^{(n-1)}\|^2 \right] \tag{31}$$

Hence, the Picard iteration is a contraction, i.e. $\{Z^{(n)}\}$ is Cauchy in $L^2(\Omega)$. Therefore, there exists a unique fixed point which is the strong solution of the Eq. (21).

Convergence and Uniqueness. We will show that the latent Neural SDE admits at most one strong solution. We will prove this by contradiction. Suppose that there exist two solutions $\{Z_t\}_{t \in [0, T]}$ and $\{\tilde{Z}_t\}_{t \in [0, T]}$ of Eq. (21) with the same initial condition Z_0 and driven by the same Brownian motion.

The difference process is defined as

$$D_t := Z_t - \tilde{Z}_t = \int_0^t \left(f_\theta(Z_s, s, e_d) - f_\theta(\tilde{Z}_s, s, e_d) \right) ds + \int_0^t \left(g_\theta(Z_s, s, e_d) - g_\theta(\tilde{Z}_s, s, e_d) \right) dW_s. \tag{32}$$

Taking the supremum over $t \in [0, T]$ and expectation, and applying the inequality $\|a + b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$, we obtain

$$\mathbb{E} \left[\sup_{t \leq T} \|D_t\|^2 \right] \leq 2 \mathbb{E} \left[\sup_{t \leq T} \left\| \int_0^t \Delta f_s ds \right\|^2 \right] + 2 \mathbb{E} \left[\sup_{t \leq T} \left\| \int_0^t \Delta g_s dW_s \right\|^2 \right], \tag{33}$$

where

$$\Delta f_s = f_\theta(Z_s, s, e_d) - f_\theta(\tilde{Z}_s, s, e_d), \quad \Delta g_s = g_\theta(Z_s, s, e_d) - g_\theta(\tilde{Z}_s, s, e_d).$$

Using the global Lipschitz continuity of f_θ and g_θ from Eq. (23), together with standard bounds for deterministic and stochastic integrals, we obtain

$$\mathbb{E} \left[\sup_{t \leq T} \|D_t\|^2 \right] \leq C \int_0^T \mathbb{E} \left[\sup_{u \leq s} \|D_u\|^2 \right] ds, \tag{34}$$

for some constant $C > 0$.

The Grönwall's inequality states that if $g(t) \leq A + C \int_0^t g(s)ds$ and if $A = 0$, then $g(t) = 0$ for all $t \in [0, T]$. It follows that

$$\mathbb{E} \left[\sup_{t \leq T} \|D_t\|^2 \right] = 0.$$

Therefore, our assumption was wrong. Hence,

$$Z_t = \tilde{Z}_t \quad \text{for all } t \in [0, T].$$

Hence it is proved that the strong solution of Eq. (21) is unique. $\square$

References

- [1] S. Panda and B. Padhi, "Poverty and inequality in odisha, india," *Journal of Public Affairs*, vol. 21, no. 2, p. e2220, 2021.
- [2] R. Kessels and G. Erreygers, "Structural equation modeling for decomposing rank-dependent indicators of socioeconomic inequality of health: an empirical study," *Health economics review*, vol. 6, no. 1, p. 56, 2016.
- [3] E. Odjesa and C. A. Ighodaro, "Socio-economic determinants of infant mortality rate in nigeria: Evidence from autoregressive distributed lag technique," *African Journal of Health Economics*, vol. 10, no. 1, pp. 15–34, 2021.
- [4] S. Kumar, S. Chakraverty, and N. Sethi, "Multidimensional poverty: Cmpi development, spatial analysis and clustering," *Social Indicators Research*, vol. 169, no. 1, pp. 647–670, 2023.
- [5] H.-A. H. Dang and P. F. Lanjouw, "Poverty dynamics in india between 2004 and 2012: Insights from longitudinal analysis using synthetic panel data," *Economic Development and Cultural Change*, vol. 67, no. 1, pp. 131–170, 2018.
- [6] X. Cheng, C.-m. Shuai, J. Wang, W.-j. Li, J. Shuai, and Y. Liu, "Building a sustainable development model for china's poverty-stricken reservoir regions based on system dynamics," *Journal of Cleaner Production*, vol. 176, pp. 535–554, 2018.
- [7] M. Łatuszyńska and S. Fate, "Combining system dynamics and agent-based simulation to study the effects of public interventions on poverty," *Procedia Computer Science*, vol. 207, pp. 3978–3987, 2022.
- [8] S. Kumar, S. Chakraverty, and N. Sethi, "Assessing changes in wealth index using primary survey data," *Socio-Economic Planning Sciences*, vol. 98, p. 102115, 2025.
- [9] S. M. Ross, *Stochastic processes*. John Wiley & Sons, 1995.
- [10] R. Gupta and S. Chakraverty, "Spectral collocation method for multidimensional stochastic integral equation involving fractional brownian motion," *Numerical Algorithms*, pp. 1–22, 2025.
- [11] A. K. Sahoo, S. Kumar, and S. Chakraverty, "An unsupervised scientific machine learning algorithm for approximating displacement of object in mass-spring-damper systems," *IEEE Access*, 2024.
- [12] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," *Advances in neural information processing systems*, vol. 31, 2018.

- [13] P. Kidger, J. Morrill, J. Foster, and T. Lyons, “Neural controlled differential equations for irregular time series,” *Advances in neural information processing systems*, vol. 33, pp. 6696–6707, 2020.
- [14] X. Liu, T. Xiao, S. Si, Q. Cao, S. Kumar, and C.-J. Hsieh, “Neural sde: Stabilizing neural ode networks with stochastic noise,” *arXiv preprint arXiv:1906.02355*, 2019.
- [15] J. Jia and A. R. Benson, “Neural jump stochastic differential equations,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [16] X. Li, T.-K. L. Wong, R. T. Chen, and D. K. Duvenaud, “Scalable gradients and variational inference for stochastic differential equations,” in *Symposium on Advances in Approximate Bayesian Inference*, pp. 1–28, PMLR, 2020.
- [17] X. Li, T.-K. L. Wong, R. T. Chen, and D. Duvenaud, “Scalable gradients for stochastic differential equations,” in *International Conference on Artificial Intelligence and Statistics*, pp. 3870–3882, PMLR, 2020.
- [18] Y. Oh, S. Kam, J. Lee, D.-Y. Lim, S. Kim, and A. Bui, “Comprehensive review of neural differential equations for time series analysis,” *arXiv preprint arXiv:2502.09885*, 2025.
- [19] B. Tzen and M. Raginsky, “Neural stochastic differential equations: Deep latent gaussian models in the diffusion limit,” *arXiv preprint arXiv:1905.09883*, 2019.
- [20] V. K. Rohatgi and A. M. E. Saleh, *An introduction to probability and statistics*. John Wiley & Sons, 2015.
- [21] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [22] B. Oksendal, *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.
- [23] J. R. Hershey and P. A. Olsen, “Approximating the kullback leibler divergence between gaussian mixture models,” in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP’07*, vol. 4, pp. IV–317, IEEE, 2007.
- [24] Government of India, “Population and household characteristics, dlhs-iii.” <https://www.data.gov.in/resource/population-and-household-characteristics-total-dlhs-iii>, 2014. Accessed: 2025-12-26.
- [25] The DHS Program, “Nfhs-4 and nfhs-5 survey raw data by dhs program.” <https://dhsprogram.com/data/available-datasets.cfm>, 2020. Accessed: 2025-12-26.
- [26] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2017.