

A new stochastic diffusion process to model and predict electricity production from natural gas sources in the United States

Safa' Alsheyab

Department of Mathematics and Statistics
Jordan University of Science and Technology
P.O.Box 3030, Irbid 22110, Jordan
e-mail:smalsheyab6@just.edu.jo

Abstract

This paper introduces a new stochastic diffusion process to model the electricity production from natural gas sources (as a percentage of total electricity production) in the United States. The method employs trend function analysis to generate fits and forecasts with both conditional and unconditional estimated trend functions. Parameters are estimated using the maximum likelihood (ML) method, based on discrete sampling paths of the variable "electricity production from natural gas sources in the United States" with annual data from 1990 to 2021. The results show that the proposed model effectively fits the data and provides dependable medium-term forecasts for 2022-2023.

Math. Subject Classification: 62M86, 60H30, 65C30

Key Words and Phrases: Stochastic diffusion process; maximum likelihood estimate; fit and forecast; trend function; prediction accuracy; electricity production from natural gas

1 Introduction

Electricity is a fundamental part of modern life, greatly improving our standard of living by powering homes, businesses, and technologies. However, generating electricity can also harm the environment. The extent of this damage largely depends on how electricity is produced. For instance, burning coal releases twice as much carbon dioxide as burning an equal amount of natural gas. Considering the many benefits of natural gas, such as its relatively low cost and low carbon footprint, it is considered a good source of electricity. Plus, gas plants can be built quickly, unlike nuclear facilities. The percentage of electricity produced from natural gas, % of total, represents the share of natural gas in overall electricity generation, measured in total gigawatt hours (GWh) produced by power plants.

Electricity production in the United States has undergone a structural transformation, with natural gas becoming one of the main sources of power generation. Understanding the dynamics of its share in total electricity production is essential for policy design, risk management, and long-term planning. Electricity generation from natural gas in the United States has grown steadily over the past thirty years. As reported in Table 1, production increased from 11.85815 units in 1990 to over 41.90 units in 2023, more than tripling. However, this growth has not been steady; the data show fluctuations and irregular deviations. For example, while production rose steadily during the early 2000s, it experienced a temporary dip in 2003, a sharp acceleration after 2009, and more recently, a slight decline in 2021 before rebounding in 2022–2023. Such variability indicates the influence of both long-term growth trends and short-term stochastic disturbances. Modelling such behaviour re-

quires a framework that captures both deterministic dynamics and random shocks. Stochastic differential equations (SDEs) provide a flexible approach that combines deterministic trends with random variability to meet these challenges.

Previous studies have extensively used stochastic models in finance and natural sciences, with growing applications in energy markets. For example, Schwartz [1] developed stochastic models for commodity prices, which were later adapted for electricity markets. Geman and Roncoroni [2] highlighted the importance of capturing seasonality and jumps in power prices, Benth et al. [3] used stochastic processes to describe electricity spot prices with periodic components, while Weron [4] provided a thorough review of stochastic models designed specifically for power markets, emphasizing the need to include both cyclical and random dynamics to accurately represent electricity price behavior. In the context of electricity, up to our knowledge, most of the literature focuses on prices, or consumption, or optimization models, or uses different methodological approaches such as stochastic time change, hybrid diffusions, or optimization frameworks. For example, Borovkova and Schmeck [5] have primarily focused on modelling electricity prices using a time-changed jump diffusion that adapts volatility and jump intensity over time. Mert et al. [6] develop a hybrid stochastic diffusion ensemble specifically for forecasting natural-gas prices. Leonel et al. [7] focus on optimization-based decision-making for large industrial self-producers using natural gas, aiming to minimize energy costs under uncertainty through a two-stage stochastic optimization framework. While for the natural gas production in the United States, Cai and Deng [8] propose a grid-optimized fractional grey model to forecast U.S. natural gas production; their focus is on total production forecasting using deterministic/grey system techniques. Despite these advancements, limited attention has been given to electricity production shares by fuel source within a stochastic diffusion framework. This creates a gap in the literature. Our goal is to use a new stochastic differential equation model to analyze U.S. electricity production from natural gas between 1990 and 2021 and

to forecast for 2022-2023. The stochastic component captures external shocks and random variability. Building on this foundation, we consider the following SDE of the form

$$dX(t) = \left(\frac{2}{t} - \lambda - \frac{\lambda\pi}{2t^2} e^{-\lambda/t} \cot\left(\frac{\pi}{2} e^{-\lambda/t}\right) \right) X(t)dt + \sigma X(t)dW(t), \quad (1)$$

which is particularly relevant for modelling U.S. electricity production from natural gas (% of total), here W is the standard Brownian motion, and λ and σ are unknown indexed parameters. For more information about Brownian motion, SDEs, and their applications, we refer the reader to [9], [10], [11], [12], and [13].

The remainder of this paper is organized as follows: In the subsequent section, the proposal model is defined as the solution to a stochastic differential equation (SDE), from which the explicit expression of the process, the transition probability distribution function (TPDF), and the moments(particularly, the conditional and unconditional trends of the process) are derived. Section 3 is dedicated to estimating the model's parameters using the Maximum Likelihood Estimation (MLE) method, based on discrete sampling of the process. In Section 4, the application of the proposed model to specific data is presented: electricity production from natural gas sources (% of total) in the United States for the period 1990-2023. The accuracy of the forecast is assessed using standard performance metrics, including mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). The performance of the proposed model is compared with that of the Gompertz model using the same dataset. Furthermore, a small Monte Carlo experiment is conducted to assess the robustness of the estimation procedure. Finally, Section 5 offers concluding remarks.

2 The model and its probabilistic characteristics

In this section, we present a novel Sine-Like stochastic diffusion process. We derive and analyze the critical properties of this process, encompassing the existence of solutions, transition probability distributions, and moments.

2.1 Sine-Like (SL) stochastic diffusion model

The SL process is the non-homogeneous diffusion process $\{X(t), t \in [t_1, T], t_1 > 0\}$ with values in $(0, \infty)$, and satisfies the following Itô's SDE

$$dX(t) = a(t, X(t), \theta)dt + b^{1/2}(t, X(t), \theta)dW(t); \quad X(t_1) = X_1 \quad (2)$$

with infinitesimal moments given by

$$\begin{aligned} a(t, X(t), \theta) &= \left(\frac{2}{t} - \lambda - \frac{\lambda\pi}{2t^2} e^{-\lambda/t} \cot\left(\frac{\pi}{2} e^{-\lambda/t}\right) \right) X(t), \\ b^{1/2}(t, X(t), \theta) &= \sigma X(t) \end{aligned} \quad (3)$$

where $\sigma > 0$, and λ is a non-zero real constant. The solution of (2)-(3) is found, by applying Itô's integral, as follows

$$\begin{aligned} X(t) &= X_1 + \int_{t_1}^t \left(\frac{2}{s} - \lambda - \frac{\lambda\pi}{2s^2} e^{-\lambda/s} \cot\left(\frac{\pi}{2} e^{-\lambda/s}\right) \right) X(s)ds \\ &\quad + \sigma \int_{t_1}^t X(s)dW(s) \end{aligned}$$

2.2 Existence and uniqueness

In this part, we show the existence and uniqueness of the solution for the SL process given in (2)-(3). To achieve this goal, it is sufficient to verify uniform Lipschitz and linear growth conditions for the infinitesimal moments; see [14], which are realized in the following theorem.

Theorem 1. *The SDE in (2)-(3) has a unique solution.*

Proof. On the one hand, consider $x, y \in \mathbb{R}^+$ and $t \in [t_1, T]$. It then follows that

$$\begin{aligned} & |a(t, x) - a(t, y)| + |\sqrt{b(t, x)} - \sqrt{b(t, y)}| \\ &= |a(t, x - y)| + |\sqrt{b(t, x - y)}|, \\ &= \left| \left(\frac{2}{t} - \lambda - \frac{\lambda\pi}{2t^2} e^{-\lambda/t} \cot\left(\frac{\pi}{2} e^{-\lambda/t}\right) \right) (x - y) \right| + |\sigma(x - y)|, \\ &= \left(\left| \left(\frac{2}{t} - \lambda - \frac{\lambda\pi}{2t^2} e^{-\lambda/t} \cot\left(\frac{\pi}{2} e^{-\lambda/t}\right) \right) \right| + |\sigma| \right) |x - y|, \\ &\leq \left(\sup_{t_1 \leq t \leq T} \left| \left(\frac{2}{t} - \lambda - \frac{\lambda\pi}{2t^2} e^{-\lambda/t} \cot\left(\frac{\pi}{2} e^{-\lambda/t}\right) \right) \right| + |\sigma| \right) |x - y| \end{aligned}$$

therefore, the SL process satisfies a uniform Lipschitz. On the other hand, this process satisfies linear growth condition as for $y = 0$, we have

$$\begin{aligned} & |a(t, x)|^2 + |\sqrt{b(t, x)}|^2 \leq \left(|a(t, x)| + |\sqrt{b(t, x)}| \right)^2, \\ &\leq \left[\left(\sup_{t_1 \leq t \leq T} \left| \left(\frac{2}{t} - \lambda - \frac{\lambda\pi}{2t^2} e^{-\lambda/t} \cot\left(\frac{\pi}{2} e^{-\lambda/t}\right) \right) \right| + |\sigma| \right) |x| \right]^2, \\ &\leq \left(\sup_{t_1 \leq t \leq T} \left| \left(\frac{2}{t} - \lambda - \frac{\lambda\pi}{2t^2} e^{-\lambda/t} \cot\left(\frac{\pi}{2} e^{-\lambda/t}\right) \right) \right| + |\sigma| \right)^2 (1 + |x|)^2. \end{aligned}$$

Thus, there is an almost surely (a.s.) continuous process $\{x(t), t \in [t_1, T]; t_1 > 0\}$ that is a.s. the unique solution of the SDE (2)-(3). $\square$

2.3 The probability distribution and moments of the process

The explicit solution of the SDE (2)-(3) can be obtained by means of the appropriate transformation; $Y(t) = \log(X(t))$, and by applying

Itô's formula [15, Theorem 4.57] to Y . Then, we have the following

$$\begin{aligned} dY(t) &= \frac{1}{X(t)} dX(t) - \frac{1}{2X^2(t)} \sigma^2 X^2(t) dt \\ &= \left(\frac{2}{t} - \lambda - \frac{\lambda\pi}{2t^2} e^{-\lambda/t} \cot\left(\frac{\pi}{2} e^{-\lambda/t}\right) \right) dt + \sigma dW(t) - \frac{\sigma^2}{2} dt \\ &= \left(\frac{2}{t} - \lambda - \frac{\lambda\pi}{2t^2} e^{-\lambda/t} \cot\left(\frac{\pi}{2} e^{-\lambda/t}\right) - \frac{\sigma^2}{2} \right) + \sigma dW(t) \end{aligned}$$

Where $Y(t_1) = \log(X_1)$. By integrating the above equation, we obtain,

$$\begin{aligned} Y(t) - Y(t_1) &= \int_{t_1}^t \left(\frac{2}{s} - \lambda - \frac{\lambda\pi}{2s^2} e^{-\lambda/s} \cot\left(\frac{\pi}{2} e^{-\lambda/s}\right) - \frac{\sigma^2}{2} \right) ds \\ &\quad + \sigma(W(t) - W(t_1)), \end{aligned}$$

and hence

$$\begin{aligned} Y(t) &= Y(t_1) + 2\log(t/t_1) - \lambda(t - t_1) + \log\left(\sin\left(\frac{\pi}{2} e^{-\lambda/t}\right)\right) \\ &\quad - \log\left(\sin\left(\frac{\pi}{2} e^{-\lambda/t_1}\right)\right) - \frac{\sigma^2}{2}(t - t_1) + \sigma(W(t) - W(t_1)), \end{aligned}$$

Therefore, the solution in terms of the original SL process

$$X(t) = X_1 \left(\frac{t}{t_1} \right)^2 \frac{\sin\left(\frac{\pi}{2} e^{-\lambda/t}\right)}{\sin\left(\frac{\pi}{2} e^{-\lambda/t_1}\right)} e^{-\lambda(t-t_1) - \frac{\sigma^2}{2}(t-t_1)} e^{\sigma(W(t)-W(t_1))} \quad (4)$$

Observe that since the initial condition $Y(t_1)$ is constant a.s., and $Y(t)$ is a Markovian process, it then follows that the finite dimensional distribution of $Y(t)$ is normal. Then, the finite dimensional distribution of $X(t)$ is log-normal distribution, and the transition probability distribution (TPDF) of $X(t)$ given $X(s)$ where $s < t$ follows a log-normal distribution denoted by $\Lambda_1(\mu(s, t, x_s), \sigma^2(t - s))$,

where $\mu(s, t, x_s)$ is given by

$$\begin{aligned}\mu(s, t, x) = & \log(x) + 2 \log(t/s) - \lambda(t - s) + \log\left(\sin\left(\frac{\pi}{2}e^{-\lambda/t}\right)\right) \\ & - \log\left(\sin\left(\frac{\pi}{2}e^{-\lambda/s}\right)\right) - \frac{\sigma^2}{2}(t - s).\end{aligned}$$

Therefore, the TPDF of the process has the following form

$$f(y, t|x_s, s) = \frac{1}{y} [2\pi\sigma^2(t - s)]^{-1/2} \exp\left(-\frac{[\log(y) - \mu(s, t, x_s)]^2}{2\sigma^2(t - s)}\right). \quad (5)$$

Taking into account that $X(t)|X(s) = x_s$ is distributed according to $\Lambda_1(\mu(s, t, x_{ts}), \sigma^2(t - s))$, and taking into account the properties of this distribution, the n^{th} conditional moment of $X(t)$ given $X(s)$ is

$$E[X^n(t)|X(s) = x_s] = \exp\left(n\mu(s, t, x_s) + \frac{n^2\sigma^2}{2}(t - s)\right)$$

Then, the conditional mean, which is considered as the trend function ($n = 1$) of the process, is

$$E[X(t)|X(s) = x_s] = x_s \left(\frac{t}{s}\right)^2 e^{-\lambda(t-s) + \log\left(\sin\left(\frac{\pi}{2}e^{-\lambda/t}\right)\right) - \log\left(\sin\left(\frac{\pi}{2}e^{-\lambda/s}\right)\right)} \quad (6)$$

On the other hand, the mean function or the unconditional trend of the process under the assumption that $P(x(t_1) = x_1) = 1$ is given by

$$\begin{aligned}E[X(t)] &= x_1 \left(\frac{t}{t_1}\right)^2 e^{-\lambda(t-t_1) + \log\left(\sin\left(\frac{\pi}{2}e^{-\lambda/t}\right)\right) - \log\left(\sin\left(\frac{\pi}{2}e^{-\lambda/t_1}\right)\right)} \\ &= x_1 \left(\frac{t}{t_1}\right)^2 \frac{\sin\left(\frac{\pi}{2}e^{-\lambda/t}\right)}{\sin\left(\frac{\pi}{2}e^{-\lambda/t_1}\right)} e^{-\lambda(t-t_1)}\end{aligned} \quad (7)$$

Since the trend function of $X(t)$ incorporates a sinusoidal component, we refer to the resulting model as a Sine-Like stochastic diffusion process. The variance of the process is

$$\begin{aligned} Var[X(t)] &= E[X^2(t)] - (E[X(t)])^2 \\ &= x_1^2 \left(\frac{t}{t_1} \right)^4 \frac{\sin^2 \left(\frac{\pi}{2} e^{-\lambda/t} \right)}{\sin^2 \left(\frac{\pi}{2} e^{-\lambda/t_1} \right)} e^{-2\lambda(t-t_1)} \left(e^{\sigma^2(t-t_1)} - 1 \right). \end{aligned}$$

3 Inference on the process

In this section, we examine the ML estimation of the parameters of the model from which we can obtain, from Zenha's theorem [16], the corresponding aforementioned estimated trend function and the conditional estimated trend function.

3.1 ML estimates

As long as we obtain the explicit expression of the TPDF of the process, we can estimate the two parameters involved in the drift and diffusion functions using the ML method. Let us consider a discrete sampling of the process $x(t_1), x(t_2), \dots, x(t_n)$ at times $t_1, t_2, \dots, t_n = T$. For simplicity, put $t_{j+1} - t_j = h$ and use x_i to refer to $x(t_i) = x_i$. Assuming the initial condition $P(X(t_1) = x_1) = 1$. The likelihood function of the parameter $\theta = (\lambda, \sigma^2)^T$ can be obtained from equation (5) as

$$\begin{aligned} \mathcal{L}(\theta) &= \prod_{j=1}^{n-1} f(x_{j+1}, t_{j+1} | x_j, t_j), \\ &= \prod_{j=1}^{n-1} \frac{1}{x_{j+1}} [2\pi\sigma^2(t_{j+1} - t_j)]^{-\frac{1}{2}} \exp \left(\frac{-[\log(x_{j+1}) - \mu(j, j+1, x_j)]^2}{2\sigma^2(t_{j+1} - t_j)} \right) \\ &= \prod_{j=1}^{n-1} \frac{1}{x_{j+1}} [2\pi\sigma^2 h]^{-1/2} \exp \left(-\frac{[\mathcal{H}_{\lambda,j} + \frac{\sigma^2}{2} h]^2}{2\sigma^2 h} \right), \end{aligned}$$

where

$$\mathcal{H}_{\lambda,j} = \log\left(\frac{x_{j+1}}{x_j}\right) - 2\log\left(\frac{t_{j+1}}{t_j}\right) + \lambda h - \log\left(\frac{\sin\left(\frac{\pi}{2}e^{-\lambda/t_{j+1}}\right)}{\sin\left(\frac{\pi}{2}e^{-\lambda/t_j}\right)}\right).$$

The log-likelihood equation is

$$\begin{aligned} \ell(\lambda, \sigma^2) = & -\frac{n-1}{2}\log(2\pi h) - \frac{n-1}{2}\log(\sigma^2) - \sum_{j=1}^{n-1}\log(x_{j+1}) \\ & - \frac{1}{2\sigma^2 h} \sum_{j=1}^{n-1} \left[\mathcal{H}_{\lambda,j} + \frac{\sigma^2}{2}h \right]^2. \end{aligned} \quad (8)$$

The log-likelihood function $\ell(\lambda, \sigma^2)$ can be maximized by solving the nonlinear likelihood equation obtained by differentiating with respect to the parameter vector $\boldsymbol{\theta} = (\lambda, \sigma^2)^T$. The first-order partial derivatives of $\ell(\lambda, \sigma^2)$ are given by

$$\begin{aligned} \frac{\partial \ell(\lambda, \sigma^2)}{\partial \lambda} = & \frac{-1}{\sigma^2 h} \sum_{j=1}^{n-1} \left(\mathcal{H}_{\lambda,j} + \frac{\sigma^2}{2}h \right) \left[h + \frac{\pi e^{-\lambda/t_{j+1}} \cot\left(\frac{\pi}{2}e^{-\lambda/t_{j+1}}\right)}{2t_{j+1}} \right] \\ & - \frac{1}{\sigma^2 h} \sum_{j=1}^{n-1} \left(\mathcal{H}_{\lambda,j} + \frac{\sigma^2}{2}h \right) \left[-\frac{\pi e^{-\lambda/t_j} \cot\left(\frac{\pi}{2}e^{-\lambda/t_j}\right)}{2t_j} \right] = 0. \end{aligned} \quad (9)$$

$$\frac{\partial \ell(\lambda, \sigma^2)}{\partial \sigma^2} = -\frac{n-1}{2\sigma^2} + \frac{1}{2\sigma^4 h} \sum_{j=1}^{n-1} \mathcal{H}_{\lambda,j}^2 - \sum_{j=1}^{n-1} \frac{h}{8} = 0, \quad (10)$$

Let $S(\boldsymbol{\theta}) = (\partial \ell(\boldsymbol{\theta})/\partial \lambda, \partial \ell(\boldsymbol{\theta})/\partial \sigma^2)^T$ be the score function. The MLE of $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\sigma}^2)$ can be obtained by solving the system of equations $S(\boldsymbol{\theta}) = 0$. Since closed-form solutions are not available, numerical methods are required to compute these estimates. From equation (10) we obtain

$$\frac{\hat{\sigma}^2}{2} = \frac{1}{h} \left[\left(1 + \frac{1}{n-1} \sum_{j=1}^{n-1} \mathcal{H}_{\hat{\lambda},j}^2 \right)^{1/2} - 1 \right] \quad (11)$$

Then, by substituting the expression of $\hat{\sigma}^2/2$ in (11) into Equation (9), we obtain the estimator $\hat{\lambda}$ from the following nonlinear equation:

$$\sum_{j=1}^{n-1} \left(\mathcal{H}_{\hat{\lambda},j} + \frac{\hat{\sigma}^2 h}{2} \right) \left[h + \frac{\pi e^{-\hat{\lambda}/t_{j+1}} \cot \left(\frac{\pi}{2} e^{-\hat{\lambda}/t_{j+1}} \right)}{2t_{j+1}} \right] - \sum_{j=1}^{n-1} \left(\mathcal{H}_{\hat{\lambda},j} + \frac{\hat{\sigma}^2 h}{2} \right) \left[\frac{\pi e^{-\hat{\lambda}/t_j} \cot \left(\frac{\pi}{2} e^{-\hat{\lambda}/t_j} \right)}{2t_j} \right] = 0. \quad (12)$$

3.2 Estimated trend functions and confidence bounds

We provide estimates for the conditional mean and the mean of the process by replacing the parameters in equations (6) and (7) by their ML estimators, due to the invariance property of the ML estimates, see for example, Theorem [5.28, 308] in [17]. Let $\hat{\lambda}$ and $\hat{\sigma}^2$ be the ML estimates of λ and σ^2 respectively, then the estimated conditional mean of the process (ECMF) is

$$\hat{E}(X(t)|X(s)) = x_s \left(\frac{t}{s} \right)^2 \frac{\sin \left(\frac{\pi}{2} e^{-\hat{\lambda}/t} \right)}{\sin \left(\frac{\pi}{2} e^{-\hat{\lambda}/s} \right)} e^{-\hat{\lambda}(t-s)} \quad (13)$$

Similarly, the estimated mean function (EMF) of the process is

$$\hat{E}(X(t)|X(t_1)) = x_1 \left(\frac{t}{t_1} \right)^2 \frac{\sin \left(\frac{\pi}{2} e^{-\hat{\lambda}/t} \right)}{\sin \left(\frac{\pi}{2} e^{-\hat{\lambda}/t_1} \right)} e^{-\hat{\lambda}(t-t_1)}, \quad (14)$$

taking into account the assumption $P(X(t_1) = x(t_1)) = 1$. In addition, we can obtain a confidence band for the CMF and MF of the process, using the procedure described in [18]. From Equation (5), we have that for $t > s$, $X(t)|X(s)$ follows $\Lambda_1(\mu(s, t, x_s), \sigma^2(t-s))$.

Therefore, we have that

$$Z = \frac{\ln(X(t)) - \mu(s, t, x_s)}{\sigma\sqrt{t-s}} \sim \mathcal{N}(0, 1).$$

Consequently, a $(1 - \alpha)100\%$ confidence band for z is determined by $P(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}) = 1 - \alpha$, for $\alpha \in (0, 1)$. and hence,

$$\mathbb{P} \left[-Z_{\frac{\alpha}{2}} \leq \frac{\mathcal{H}_{\lambda,t} + \frac{\sigma^2}{2}(t-t_1)}{\sigma\sqrt{t-t_1}} \leq Z_{\frac{\alpha}{2}} \right] \approx 1 - \alpha.$$

here

$$\mathcal{H}_{\lambda,t} = \log \left(\frac{x(t)}{x_1} \right) - 2 \log(t/t_1) + \lambda(t-t_1) - \log \left(\frac{\sin \left(\frac{\pi}{2} e^{-\lambda/t} \right)}{\sin \left(\frac{\pi}{2} e^{-\lambda/t_1} \right)} \right).$$

From this, we can obtain a $(1 - \alpha)100\%$ confidence bound (CB) for $x(t)$ as

$$x_{lower}(t) \leq x(t) \leq x_{upper}(t) \quad (15)$$

where,

$$x_{lower}(t) = x_1 e^{-Z_{\alpha/2} \sigma \sqrt{t-t_1} + 2 \log(\frac{t}{t_1}) - \lambda(t-t_1) + \log \left(\frac{\sin \left(\frac{\pi}{2} e^{-\lambda/t} \right)}{\sin \left(\frac{\pi}{2} e^{-\lambda/t_1} \right)} \right) - \frac{\sigma^2}{2}(t-t_1)}$$

$$x_{upper}(t) = x_1 e^{Z_{\alpha/2} \sigma \sqrt{t-t_1} + 2 \log(\frac{t}{t_1}) - \lambda(t-t_1) + \log \left(\frac{\sin \left(\frac{\pi}{2} e^{-\lambda/t} \right)}{\sin \left(\frac{\pi}{2} e^{-\lambda/t_1} \right)} \right) - \frac{\sigma^2}{2}(t-t_1)}.$$

4 Applications

This section focuses on giving an application of our model to real data, where we give an application in real data for the electricity production from natural gas sources (% of total) in the United States during the period 1990 to 2023. Raw data are shown in Table 1; these data are annualized and available in the World Bank's database. We compare our model with an existing model, Gompertz, which is suitable for modelling increasing trend data.

4.1 Application to electricity production from natural gas sources in the United States (% of total)

Using data from 1990 to 2021, we train the model and estimate its parameters: $\hat{\lambda} = -0.03828096$, and $\hat{\sigma} = 0.0673062$. To forecast electricity production from natural gas sources in the United States (% of total) for 2022 and 2023, we apply EMF and ECMF, replacing the estimated parameters in (7) and (6), respectively. See Table 2. Figure 1 shows the performance of the stochastic SL diffusion process for our forecasts. The left panel displays the observed data, estimated trend function (EMF), and estimated confidence bands, illustrating how well the SL process fits the current data and makes accurate predictions. The right panel also shows the conditional trend function estimation, forecasts, and estimated confidence bands. All computations in this paper were performed using R software [19].

Table 1: Electricity production from natural gas sources in the United States.

Year	1990	1991	1992	1993	1994	1995
Data	11.85815	12.27868	12.97155	12.92770	14.16742	14.76346
Year	1996	1997	1998	1999	2000	2001
Data	13.02046	13.67453	14.57905	14.93086	15.65117	17.07274
Year	2002	2003	2004	2005	2006	2007
Data	17.58605	16.41918	17.52281	18.22920	19.59561	21.03976
Year	2008	2009	2010	2011	2012	2013
Data	20.83612	22.67735	23.24735	24.03179	29.47220	26.90093
Year	2014	2015	2016	2017	2018	2019
Data	26.75654	31.79336	32.81091	31.20787	34.09804	37.33868
Year	2020	2021	2022	2023		
Data	39.43958	37.35339	38.71585	41.90800		

Table 2: Predictions with EMF and ECMF of the process.

Year	Real Data	EMF	ECMF
2022	38.71585	41.67541	38.84946
2023	41.90800	43.34456	40.26646

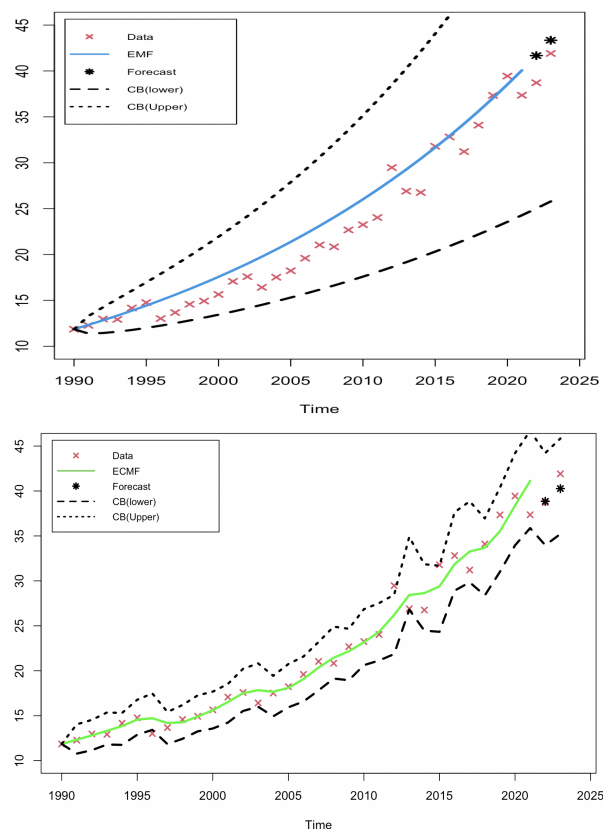


Figure 1: Observed data, EMF, ECMF, and the forecasted values.

We compare the proposed diffusion process with the Gompertz process [20] using the existing data. Table 3 shows the estimated

parameters and the AIC criterion. The SL has a lower AIC and thus excels the Gompertz diffusion process.

Table 3: SL, Gompertz: Parameters and AIC.

Model	λ	β	σ	AIC
SL	-0.03828096	NA	0.06730620	112.3892
Gompertz	0.060000022	0.006881754	0.067492691	114.3477

Figure 2 shows the fits made using the methods mentioned in Table 3.

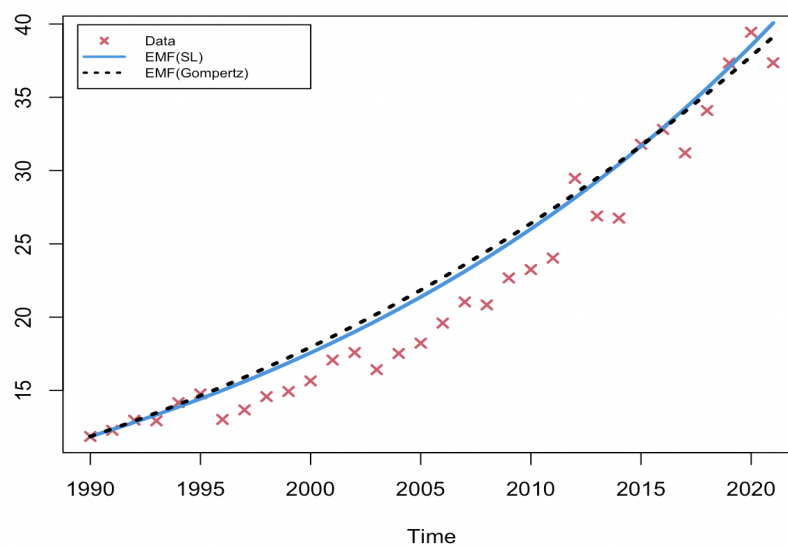


Figure 2: Observed data, EMF using SL, and Gompertz diffusion processes.

4.2 Goodness of fit

Forecast accuracy can be measured using metrics such as mean absolute error (MAE), the root mean square error (RMSE), and the mean absolute percentage error (MAPE). We consider the observed data as $x(t)$, the predicted data $\hat{x}(t)$ which is obtained by substituting the estimated parameter $\hat{\lambda} = -0.03828096$ in equation (14), and the difference between them $x(t) - \hat{x}(t)$ over the years 1990-2021. We calculate the mean absolute error (MAE), the root mean square error (RMSE), and the mean absolute percentage error (MAPE) as follows

$$\begin{aligned} \text{MAE} &= \frac{1}{32} \sum_{i=1}^{32} |x(t_i) - \hat{x}_i(t)| = 1.718274, \\ \text{RMSE} &= \sqrt{\frac{1}{32} \sum_{i=1}^{32} |x(t_i) - \hat{x}_i(t)|^2} = 66.1315, \\ \text{MAPE} &= \frac{1}{32} \sum_{i=1}^{32} \frac{|x(t_i) - \hat{x}_i(t)|}{x(t)} \times 100 = 7.559711. \end{aligned}$$

The value obtained for MAPE is less than 10, and this indicates that we obtained a high-accuracy prediction according to [21].

4.3 Simulation

The sample trajectories were simulated using Equation (4) with values of α , σ , and x_1 that approximate the values of these parameters in the real example in the application on which this study is conducted in section 4.1. Ten trajectories were generated, each containing 500 values. Figure 3 shows the simulated trajectories of the SL process versus to MF for the particular case of $\lambda = -0.03828096$, $\sigma = 0.0673062$, $t_1 = 1990$, $\Delta t = 0.066$, $x_{t_1} = 11.85815$ which respectively correspond to values close to those obtained in the study of $X(t)$. We used the same algorithm as in [22] to simulate the trajectories of the SL process.

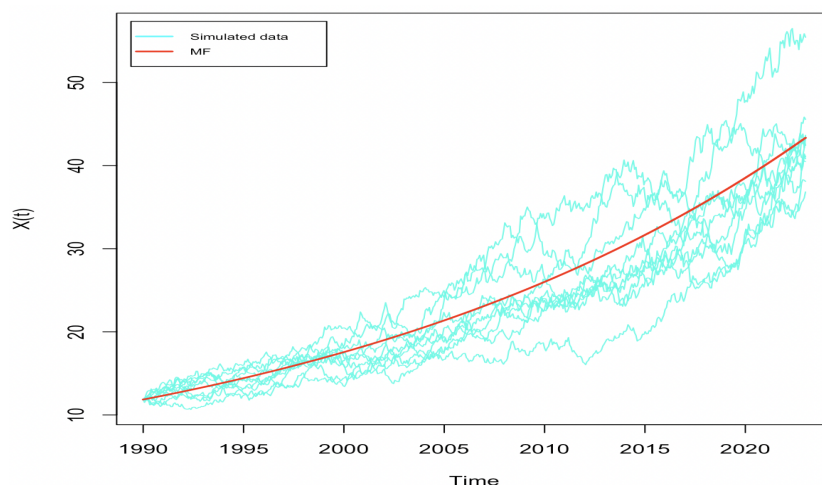


Figure 3: Sine-like (SL) process simulated along with the mean function (MF).

5 Conclusions

Based on the results obtained (see Table 2, Figure 1), we conclude that modeling the U.S. data on electricity generated from natural gas (% of total) over the period 1990–2021 using the novel Sine-Like model- estimated according to the procedure outlined in Section 3- yields a high level of accuracy in the historical trend and produces reliable medium-term forecasts for 2022–2023. The conditioned trend function provides much better descriptions and forecasts than the basic trend function. Furthermore, a comparison with the Gompertz model using the same dataset and sample period demonstrates that the sine-like stochastic diffusion model achieves superior overall performance.

Future work could build on this study by incorporating external time-varying factors-such as natural gas prices, renewable energy adoption, or seasonal demand- directly into the drift and/or diffusion terms of the SDE. This would require moving from a univariate

to a multivariate stochastic framework, such as coupled SDE systems or stochastic regression models.

Acknowledgments

The author appreciates the efforts of the Editor and the journal's editorial team.

References

- [1] E. S. Schwartz, The stochastic behavior of commodity prices: Implications for valuation and hedging, *J. Finance*, **52** (1997), 923–973.
- [2] H. Geman, A. Roncoroni, Understanding the fine structure of electricity prices, *J. Bus.*, **79** (2006), 1225–1261.
- [3] F. E. Benth, J. Š. Benth, S. Koekebakker, *Stochastic modelling of electricity and related markets*, World Sci. Publ., (2008).
- [4] R. Weron, Electricity price forecasting: A review of the state-of-the-art with a look into the future, *Int. J. Forecast.*, **30** (2014), 1030–1081.
- [5] Borovkova, S., & Schmeck, M. D. . Electricity price modeling with stochastic time change. *Energy Economics*, **63**, (2017), 51–65.
- [6] Mert, K., Koc, E., & Selcuk-Kestel, A. S.. Modelling Natural Gas Future Prices via Hybrid Stochastic Diffusion Processes. In *Proceedings of the International Conference on Computational and Financial Econometrics* . Springer, Cham(2025), 297–324.

- [7] L. D. Leonel, M. H. Balan, L. A. S. Camargo, D. S. Ramos, R. Castro, & F. S. Clemente . Stochastic Decision-Making Optimization Model for Large Electricity Self-Producers Using Natural Gas in Industrial Processes: An Approach Considering a Regret Cost Function. *Energies*, **17(21)**, (2024), 5389.
- [8] Cai, Y., & Deng, Y.. Modelling Natural Gas Energy Production of United States. *E3S Web of Conferences*, **248**, (2021), 02034 .
- [9] I. Karatzas, S. E. Shreve, *Brownian motion and stochastic calculus*, 2 Eds., New York: Springer, (1991).
- [10] L. C. Evans, *An introduction to stochastic differential equations*, Amer. Math. Soc., (2012).
- [11] B. Øksendal, *Stochastic differential equations: An introduction with applications*, Springer, (2013).
- [12] X. Mao, *Stochastic differential equations and applications*, 2 Eds.: Elsevier, (2007).
- [13] K. Sobczyk, *Stochastic differential equations: With applications to physics and engineering*, Springer, (2001).
- [14] P. E. Kloeden, E. Platen, *Numerical solution of stochastic differential equations*, Springer, (1992).
- [15] J. Jacod, A. Shiryaev, *Limit theorems for stochastic processes*, 2 Eds.: Springer, (2002).
- [16] P. W. Zehna, Invariance of maximum likelihood estimators, *Ann. Math. Stat.*, 37 (1966), 744–748.
- [17] M. J. Schervish, *Theory of statistics*, New York: Springer, (1999).

- [18] A. Katsamaki, C. H. Skiadas, Analytic solution and estimation of parameters on a stochastic exponential model for a technological diffusion process, *Appl. Stoch. Models Data Anal.*, 11 (1995), 59–75.
- [19] The R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, (2016). Available from:
https://web.mit.edu/r_v3.3.1/fullrefman.pdf.
- [20] R. Gutiérrez, R. Gutiérrez-Sánchez, A. Nafidi, Modelling and forecasting vehicle stocks using the trends of stochastic Gompertz diffusion models: The case of Spain, *Appl. Stoch. Model. Bus. Ind.*, 25 (2009), 385–405.
- [21] C. D. Lewis, *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*, Butterworth-Heinemann, (1982).
- [22] S. Alsheyab, K. Shakhatreh, A new stochastic diffusion process based on generalized Gamma-like curve: inference, computation, with applications, *AIMS Math.*, 9 (2024), 27687–27703.