sciendo

CrossMark

Moroccan J. of Pure and Appl. Anal. (MJPAA)
Volume 9(1), 2023, Pages 127–140
ISSN: Online 2351-8227 - Print 2605-6364
DOI: 10.2478/mjpaa-2023-0009

Stochastic Pareto diffusion process : Statistical analysis and computational issues. Simulation and Application

Ahmed Nafidi[1], Ilyasse Makroz[2], Boujemâa Achchab[3] and Ramón Gutiérrez-Sánchez[4]

Abstract. We propose a novel diffusion process having a mean function equal to the Pareto probability density function up to a constant of proportionality. We examine the probabilistic properties of the proposed model. Then, referring to the problem of statistical inference, we describe the approach employed to tackle the issue of obtaining parameter estimates by maximizing the likelihood function based on discrete sampling. This estimation reduces to solving a set of complex equations, that is accomplished using the simulated annealing algorithm. A simulation study is also given to validate the methodology presented. Finally, using a real-world example of the Moroccan child mortality rate, we obtain the fits and forecasts by employing the suggested stochastic process and nonlinear regression model.

Mathematics Subject Classification (2020). 60H35, 65C30, 62M86, 62M20, 60H30.
Key words and phrases. Stochastic diffusion model, Pareto distribution, Statistical inference, Simulated annealing method, Nonlinear regression, Child mortality rate

Received : December 21, 2022 - Accepted : January 22, 2023.

[1,2,3] *Laboratory of Analysis and Modeling Systems for Decision Support, Hassan First University of Settat, National School of Applied Sciences, 26100, Berrchid, Morocco*
[1]*e-mail: ahmed.nafidi@uhp.ac.ma*
[2]*e-mail: i.makroz@uhp.ac.ma*
[3]*e-mail: boujemaa.achchab@uhp.ac.ma*
[4]*Departamento de Estadística e Investigación Operativa, Faculty of sciences, Campus de Fuentenueva, Universidad de Granada, 18071, Granada, Spain*
e-mail: ramongs.ugr.es .

1. Introduction

Stochastic diffusion processes (SDP) are useful tools in the modeling and analysis of phenomena that evolve randomly and continuously in time. Their fields of application range from traditional applications such as economics, physics, and biology, to news technology and the emissions of greenhouse gases. Several SDPs have been developed to model a wide range of problems, such as, Pearson [7]; Weibull [19]; Lomax [17]; Brody [18]; γ-power of the Lundqvist-Korf [6].

In studies of statistical inference in these processes, if we are confident about the parametric specification of the model under investigation, then maximum likelihood is the method of choice, as these estimators offer many desirable properties [27]. However, in most cases the transition densities of diffusion processes are unknown and as a result, the maximum likelihood function is often difficult to implement. Accordingly, estimation methods are often needed, and in many cases, these are based on approximations of the maximum likelihood. This theory has been extensively reviewed by [8], [22] and [13], among others.

Many real-world systems depend not only on the previous states of the process but also on time. For example, stock prices are known to behave differently according to the day of the week, the month, and the year. Various non-homogeneous stochastic processes, with exogenous factors, have been suggested to depict this sort of time-dependent behavior, such as Lognormal [25], Gompertz [12], Gamma [11]. In this study, We develop a novel diffusion process based on the density function of the Pareto distribution as a time inhomogeneous extension of the lognormal process.
The Pareto distribution arises in many contexts, such as the size of earthquakes [9], disk drive sector errors [24] among other applications [20]. Furthermore many extensions can be found throughout the literature, for example, beta-Pareto distribution [1], Weibull-Pareto distribution [3], Gamma-Pareto distribution [2], among others.
This distribution was originally introduced by Vilfredo Pareto [16], who used it to model the unequal distribution of wealth. The Pareto distribution which is referred to in the literature as $P(I)(\beta, \alpha)$, is characterized by a long right-skewed tail which approaches but never touches the horizontal axis.
The pdf of a Pareto distribution $P(I)(\beta, \alpha)$ defined on $[\beta, +\infty)$ is :

$$f(t) = \begin{cases} \frac{\alpha \beta^{\alpha}}{t^{\alpha+1}} & \text{if } t \geq \beta \\ 0 & \text{if } t < \beta \end{cases} \tag{1.1}$$

Examples of (approximate) Pareto distributions in real-life phenomena include the sizes of human settlements, and the sizes of sand particles [23].

2. The Pareto diffusion process

2.1. The model. The Stochastic Pareto diffusion process (SPDP) is the time dependent diffusion process $\{y(\tau) : \tau \in [s, T], s > 0\}$ with values on the positive real line $\mathbb{R}_+^*$, and with

$$M_1(\tau, y) = -\frac{\gamma}{\tau}\, y \qquad \text{and} \qquad M_2(\tau, y) = \nu^2 y^2 \quad , \tag{2.1}$$

where $\nu > 0$, $\gamma > 1$ are real numbers.
alternatively, the SPDP can be defined by means of Itô's SDE as follows:

$$dy(\tau) = M_1(\tau, y)d\tau + \sqrt{M_2(\tau, y)}dB(\tau) \quad , \quad y(s) = y_s \quad , \tag{2.2}$$

With $B(\tau)$ denoting the standard Brownian motion and $y_s \in \mathbb{R}_+^*$.

2.2. Analytical expression of the process. We can show that $\exists K \in \mathbb{R}^{*+}$, i.e $K = \frac{|\gamma|}{s} + \nu$, such that $\forall a, y \in \mathbb{R}^+$ and $\tau \in [s, T]$, we have

$$\mid M_1(\tau, a) - M_1(\tau, y) \mid + \mid \sqrt{M_2(\tau, a)} - \sqrt{M_2(\tau, y)} \mid \leq K \mid a - y \mid$$
$$\mid M_1(\tau, y) \mid^2 + \mid \sqrt{M_2(\tau, y)} \mid^2 \leq K^2(1 + \mid y \mid^2).$$

Therefore, (2.2) has a strong solution $\{y(\tau) : \tau \in [s, T], s > 0\}$ that is continuous almost surely, and which satisfies the initial condition $y(s) = y_s$.
Using the continuity of $M_1(\tau, y)$ with respect to τ, we deduce that the solution $\{y(\tau) : \tau \in [s, T], \quad s > 0\}$ is a one-dimensional diffusion process with drift $M_1(\tau, y)$ and the diffusion coefficient $M_2(\tau, y)$ (see [4]), and their expression is obtained by using Itô's lemma applied to the time independent transform $\xi(\tau) = log(y(\tau))$, then (2.2) is written as follows:

$$d\xi(\tau) = \left(-\frac{\gamma}{\tau} - \frac{\nu^2}{2}\right) d\tau + \nu dB(\tau) \quad , \quad \xi(s) = \log(y_s).$$

By integrating and substituting, the strong solution of the SDE (2.2) is expressed as follows:

$$y(\tau) = y_\zeta \left(\frac{\zeta}{\tau}\right)^\gamma \exp\left[-\frac{\nu^2}{2}(\tau - \zeta) + \nu\left(B(\tau) - B(\zeta)\right)\right], \tag{2.3}$$

from which we deduce that

$$y(\tau) \mid y(\zeta) = y_\zeta \sim \Lambda\left(log(y_\zeta) + \gamma log\left(\frac{\zeta}{\tau}\right) - \frac{\nu^2}{2}(\tau - \zeta), \nu^2(\tau - \zeta)\right), \tag{2.4}$$

therefore, the TPDF of the process is given by:

$$f(y, \tau \mid y_\zeta, \zeta) = \frac{1}{y\sqrt{2\pi(\tau - \zeta)\nu^2}} \times \exp\left(-\frac{\left[log\left(\frac{y}{y_\zeta}\right) + \gamma log\left(\frac{\tau}{\zeta}\right) + \frac{\nu^2}{2}(\tau - \zeta)\right]^2}{2\nu^2(\tau - \zeta)}\right). \tag{2.5}$$

2.3. Mean functions. Based on the well known Lognormal distribution's properties, we have:

$$E(y^r \mid y(\zeta) = y_\zeta) = exp\left[r\left(log(y_\zeta) + \gamma log\left(\frac{\zeta}{\tau}\right) - \frac{\nu^2}{2}(\tau - \zeta)\right) + \frac{r^2}{2}\nu^2(\tau - \zeta)\right]. \tag{2.6}$$

The conditional mean function of the process is obtained by setting r in (2.6) to one, i.e:

$$E(y(\tau) \mid y(\zeta) = y_\zeta) = y_\zeta \left(\frac{\zeta}{\tau}\right)^\gamma. \tag{2.7}$$

Furthermore, considering that $P(y(s) = y_s) = 1$, the unconditional mean function of the SPDP is:

$$E(y(\tau)) = y_s \left(\frac{s}{\tau}\right)^\gamma. \tag{2.8}$$

We observe that:

- The mean function of the SPDP defined in (2.8) is equal to the Pareto probability density function (1.1), up to a constant of proportionality.
- On the other hand, when $\nu = 0$, (2.2) reduces to an ordinary differential equation for which the solution is, $y(\tau) = y_\zeta(\frac{\zeta}{\tau})^\gamma$, and which is equal to the Pareto (Type I) [16] density function $P(I)(\zeta, \gamma)$, up to a constant of proportionality.

3. Parameters Estimation

The unknown parameters in the process are estimated by ML method. Therefore, we consider a discrete sampling of the process $y(\tau_1), y(\tau_2), ..., y(\tau_N)$ at times $\tau_1, \tau_2, ..., \tau_N$, and we denote $y(\tau_j) = y_j$, for $j = 1, ..., N$ in the following. In addition, for simplicity, we suppose that the time gap between two consecutive observations is constant (i.e., $\tau_j - \tau_{j-1} = \delta$, for $j = 2, ..., N$) and $\theta = (\gamma, \nu)$ is the vector of parameters.
Hereafter, by supposing that $P[y(\tau_1) = y_{\tau_1}] = 1$, the corresponding likelihood function $l(y; \theta)$ is:

$$\begin{aligned} l(y,\theta) &= \prod_{j=2}^{N} f(y_j, \tau_j \mid y_{j-1}, \tau_{j-1}) \\ &= \prod_{j=2}^{N} \frac{1}{y\sqrt{2\pi\delta\nu^2}} \exp\left(- \frac{\left[log\left(\frac{y_j}{y_{j-1}}\right) + \gamma log\left(\frac{\tau_j}{\tau_{j-1}}\right) + \frac{\nu^2}{2}\delta \right]^2}{2\nu^2\delta} \right). \end{aligned}$$

By setting $K_{j,\gamma} = log\left(\frac{y_j}{y_{j-1}}\right) + \gamma log\left(\frac{\tau_j}{\tau_{j-1}}\right)$, the log-likelihood function is given by:

$$\begin{aligned} \log(l(y,\theta)) &= \sum_{j=2}^{N} -log(y_j) - \frac{1}{2}log(2\pi\delta) - \frac{1}{2}log(\nu^2) \\ &\quad - \frac{1}{2\nu^2\delta}\left[log\left(\frac{y_j}{y_{j-1}}\right) + \gamma log\left(\frac{\tau_j}{\tau_{j-1}}\right) + \frac{\nu^2}{2}\delta \right]^2 \\ &= -\frac{N-1}{2}log(\nu^2) - \frac{N-1}{2}log(2\pi\delta) - \sum_{j=2}^{N}\left[log(y_j) + \frac{1}{2\nu^2\delta}\left(K_{j,\gamma} + \frac{\nu^2}{2}\delta \right)^2 \right]. \end{aligned}$$

This function is then differentiated with respect to the vector θ and after several operations and simplifications, the likelihood equations are:

$$\sum_{j=2}^{N}\left(K_{j,\gamma} + \frac{\nu^2\delta}{2} \right) log\left(\frac{\tau_j}{\tau_{j-1}}\right) = 0 \tag{3.1}$$

$$\left(\sum_{j=2}^{N} K_{j,\gamma}^2 \right) - (N-1)\nu^2\delta - \frac{N-1}{4}\nu^4\delta^2 = 0. \tag{3.2}$$

Note that the trinomial of (3.2) has two roots. Since we know that ν^2 corresponds to the non-negative root, the estimator $\hat{\nu}^2$, can be expressed as

$$\hat{v}^2 = \frac{2}{\delta}\left(\left(1+\frac{1}{N-1}\sum_{j=2}^{N} K_{j,\gamma}^2\right)^{1/2} - 1\right). \tag{3.3}$$

After replacing v^2 by $\hat{v}^2$ in (3.1), the estimator of γ satisfies the following non-linear equation :

$$\sum_{j=2}^{N}\left(K_{j,\gamma} + \left(1+\frac{1}{n-1}\sum_{j=2}^{N} K_{j,\gamma}^2\right)^{1/2} - 1\right) log\left(\frac{\tau_j}{\tau_{j-1}}\right) = 0. \tag{3.4}$$

The solution to this equation may be arduous to find, and so this problem is addressed using numerical resolution methods.

4. Computational aspects

4.1. Approximate likelihood estimators. In the present work, simulated annealing is used to solve (3.4), as described below.

Simulated annealing is a commonly used random optimization method that converges to the extremum of a given function using stochastic exploration. In theory, this enables us to obtain the global extremum of the function, in contrast to most of the classical (deterministic) methods which are limited to a local extremum. The likelihood of a physical system possessing an energy E when thermodynamic equilibrium is attained at a temperature T is proportional to the Boltzmann factor: $exp(-E/KT)$, where K represents the Boltzmann constant. To model the evolution of a physical system towards thermodynamic equilibrium at a temperature T, we use the Metropolis algorithm: Starting from an initial configuration of the system, we apply an elementary modification, if it decreases the energy E, we accept this modification. Otherwise, if $\Delta E > 0$, we accept the modification with probability $exp(-\Delta E/T)$. At low temperatures, only very small increases in energy are accepted. The series of configurations converges to a state of thermodynamic equilibrium at temperature T. We slightly decrease the temperature and then re-converge towards thermodynamic equilibrium. If the temperature is lowered too quickly, convergence is slowed down. When the temperature is close to zero, the energy is close to a minimum.

Analogically if E is replaced by a cost function, this algorithm allows us to find the global extremum of this function: When T is high the algorithm can, with a high probability, temporarily accept a value that degrades the cost function, making it possible to escape the local extremes, when T approaches 0 a degradation is never accepted.

4.2. Estimated mean functions. By substituting the parameters with their estimators in (2.7) and (2.8), the estimated mean function (EMF) and the estimated conditional mean function (ECMF) are derived. Thus the EMF is :

$$\hat{E}(y(\tau)) = y_s\left(\frac{s}{\tau}\right)^{\hat{\gamma}}. \tag{4.1}$$

and the ECMF is:

$$\hat{E}(y(\tau) \mid y(\zeta) = x_\zeta) = y_\zeta \left(\frac{\zeta}{\tau}\right)^{\hat{\gamma}}. \tag{4.2}$$

4.3. Goodness of fit. The mean absolute percentage error (MAPE) is a commonly used measure for assessing the accuracy of the forecasts obtained by a model. The MAPE measures the average absolute value of percentage errors. Let x_i be the observed values, $\hat{x}_i$ the model's forecasted values, and n and the number of forecasts. Thus, The MAPE is given by:

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \frac{\mid x_i - \hat{x}_i \mid}{x_i}.$$

Although the MAPE is the most commonly used measure for accuracy, it has several shortcomings [26], [15]. To overcome those problems we consider two additional accuracy measures, the symmetric mean absolute percentage error (SMAPE) and the mean squared error (MSE), which are defined as follows:

$$SMAPE = \frac{100}{n} \sum_{i=1}^{n} \frac{\mid x_i - \hat{x}_i \mid}{(\mid \hat{x}_i \mid + \mid x_i \mid)/2}$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2.$$

4.4. Confidence interval of the process. As the process is inhomogeneous, it's not possible to find the stationary distribution and the asymptotic confidence intervals of the parameters of the process, for example, in the case of the Gompertz process [10] and others. An alternative procedure that makes it possible to find bounds that contain almost all the trajectories of the process is used. The principle of this procedure is to find an estimation of the confidence interval of the variable $y(\tau)$. By using the expression (2.2) and the standard brownian motion properties, the random variable Z is given by:

$$Z = \frac{log\left(\frac{y(\tau)}{y(s)}\right) - \mu(\gamma, \nu^2, \tau, s)}{\nu\sqrt{\tau - s}} \sim \mathcal{N}(0, 1).$$

where $\mu(\gamma, \nu^2, \tau, s) = -\gamma log\left(\frac{\tau}{s}\right) + \frac{\nu^2}{2}(\tau - s)$.

An α% confidence interval for z is given by $\mathrm{P}(\mid z \mid \leq z_\alpha) = \alpha$. From this, we can obtain a confidence interval of with following form:

$$y_l(\tau) = y_s exp\left[\mu(\gamma, \nu^2, \tau, s) - z_\alpha \nu\sqrt{\tau - s}\right] \tag{4.3}$$

$$y_u(\tau) = y_s exp\left[\mu(\gamma, \nu^2, \tau, s) + z_\alpha \nu\sqrt{\tau - s}\right]. \tag{4.4}$$

With $z_\alpha = \phi^{-1}\left(\frac{1+\alpha}{2}\right)$ and where ϕ denote the cumulative normal standard distribution.

Then, by substituting the parameters by their estimators in (4.3) and (4.4) the estimated confidence intervals are given by:

$$\hat{y}_l(\tau) = y_s exp\left[\mu(\hat{\gamma}, \hat{\nu}^2, \tau, s) - z_\alpha \hat{\nu}\sqrt{\tau - s}\right]. \tag{4.5}$$

$$\hat{y}_u(\tau) = y_s exp\left[\mu(\hat{\gamma}, \hat{v}^2, \tau, s) + z_\alpha \hat{v}\sqrt{\tau - s}\right]. \tag{4.6}$$

5. Simulation

In order to obtain an illustration of the proposed model, we discretize the interval $[t, s]$ with $\tau_j = \tau_{j-1} + (j-1)\delta$ for $j = 2, ..., n$. Assume ($\tau_1 = t$) and a discretization step of $\delta = \frac{s-t}{n}$, where n is the sample size. Let $s = 0.2$, $t = 0.9$, $n = 400$, and $x_t = 135.171$, twenty-five trajectories of the process given by (2.2) were simulated.

Figure 1 shows the simulation results, as well as the EMF of the SPDP.

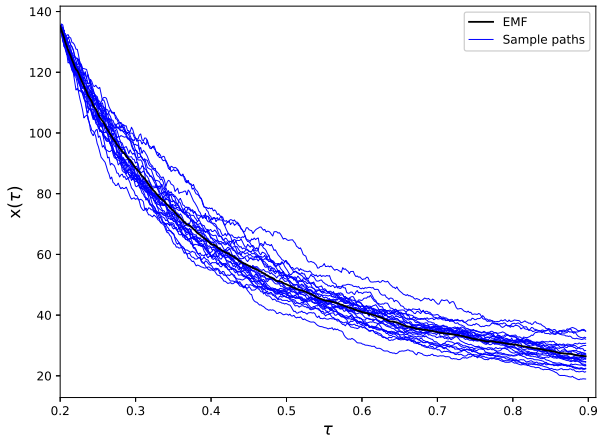

FIGURE 1. Simulated trajectories vs the EMF for $\gamma = 1.1$ and $\nu = 0.2$

we obtained the estimators $\hat{\gamma}_j$ and ${\hat{v}_j}^2$ of each trajectory i using the SA method to solve (3.4), with a logarithmic cooling rate, an initial temperature of 1000. We then considered the average value of each estimator, i.e: $\overline{\gamma} = \frac{1}{25}\sum_{j=1}^{25}\hat{\gamma}_j$ and $\overline{v}^2 = \frac{1}{25}\sum_{j=1}^{25}{\hat{v}_j}^2$.
The results obtained were : $\overline{\gamma} = 1.107303765890008$ and $\overline{v}^2 = 0.199945288360394$. Denoting the i^{th} sample path by x_i, at each time point τ_j, the simulated trajectories average value ($\bar{x}(\tau_j)$) was computed using the fomula $\bar{x}(\tau_j) = \frac{1}{d}\sum_{i=i}^{25} x^i(\tau_j)$. Figure 2 shows the calculated EMF of the SPDP and the average value of the of the simulated trajectories.

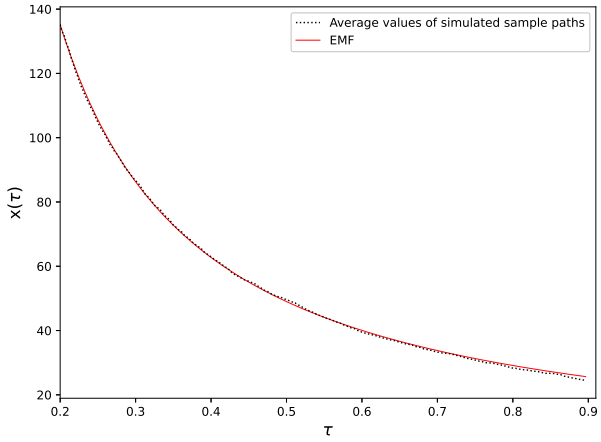

FIGURE 2. EMF vs the average of the simulated sample paths

6. Application

Mortality rates for various age groups are important measures of a country's health status. Mortality rates can be used to identify vulnerable populations since data on disease frequency and severity is often unavailable. Furthermore, these metrics are commonly utilized in worldwide socioeconomic development comparisons.

For example, statistics on under-five mortality may reflect the consequences of gender discrimination more effectively than newborn mortality statistics, because starvation and medical interventions have a greater influence on this age group. Under-five mortality rates are greater for boys than for girls in nations where parental gender preferences are insignificant.

Complete vital registration systems are scarce in developing countries. As a result, estimates must be produced from sample surveys or registration, census, or survey data using indirect estimation techniques [5].

In our application of the SPDP described above, we consider the variable $x(t)$ defined as the under-five mortality rate, i.e. the probability per 1,000 that a newborn baby will die before reaching the age of five years, if subject to age-specific mortality rates of the specified year. These data are of annual periodicity and are available on [5]. For Morocco, the average value in this respect during the period from 1983 to 2014 was 47.4188. The values recorded for this period along with the EMF and ECMF are shown in Table 1.

6.1. Fits and forecasts using nonlinear regression. The standard strategy to detect the power-law behaviour in empirical data make use of the fact that if $f(x)$ follows a power law then $log(f(x)) = log(C) - \gamma log(x)$. Therefore if we plot the function $log(f(x))$ on a logarithmic scale, we would expect to obtain a straight line of slope $-\gamma$. In order to apply this method to the present dataset, we used the linear regression implementation from scikit-learn API

TABLE 1. Yearly child mortality rate in Morocco

Year	Data	EMF	ECMF	Year	Data	EMF	ECMF
1983	83.2	83.2	83.2	1999	43.5	42.6514	43.454
1984	79.9	79.7844	79.7844	2000	41.7	40.9141	41.7282
1985	76.5	76.5106	76.6215	2001	40	39.2485	40.0023
1986	73.4	73.3728	73.3626	2002	38.5	37.6514	38.3723
1987	70.4	70.365	70.3912	2003	36.9	36.1201	36.9341
1988	67.5	67.4821	67.5156	2004	35.5	34.6517	35.4
1989	64.9	64.7185	64.7358	2005	34	33.2438	34.0576
1990	62.5	62.0695	62.2435	2006	32.7	31.8937	32.6192
1991	60.1	59.5301	59.943	2007	31.3	30.599	31.3726
1992	57.9	57.0959	57.6424	2008	30	29.3575	30.0301
1993	55.6	54.7623	55.5335	2009	28.8	28.167	28.7834
1994	53.5	52.5252	53.3287	2010	27.5	27.0253	27.6327
1995	51.3	50.3805	51.3155	2011	26.3	25.9305	26.3859
1996	49.3	48.3244	49.2064	2012	25.2	24.8804	25.235
1997	47.3	46.3532	47.289	2013	24	23.8734	24.1801
1998	45.3	44.4633	45.3715	2014	22.9	22.9077	23.0291

TABLE 2. Forecast accuracy for the deterministic Pareto model

Forecast accuracy error measurements	Values
MAPE	2.5667282278198185
SMAPE	2.514343279863844
MSE	0.9992024240811666

in Python [21]. The estimate obtained for γ using this method is 81.5907155, and the sum of squared error is 0.002. A popular approach is to rely on visualizations of the data, as in Figure 3, where the logarithm of each data point was plotted against the logarithm of time expressed in years to confirm the linearity of the logarithm of the data and therefore to deduce the power-law behavior in data.

The results obtained using this nonlinear regression are as follows: the fits and the predictions are illustrated in Figure 4, and the MAPE, SMAPE, and MSE are shown in Table 2.

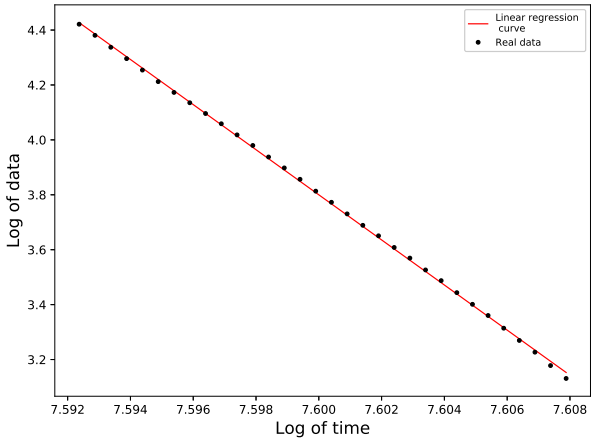

FIGURE 3. Logarithm of the data with linear regression curve

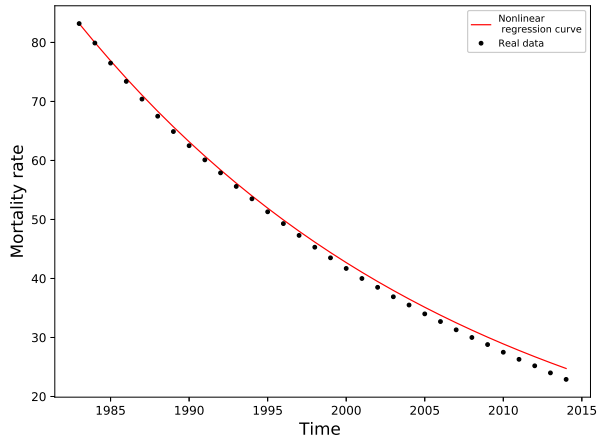

FIGURE 4. Real data vs predictions of the deterministic model

6.2. Fits and forecasts using the stochastic Pareto model. A MatLab program was implemented to perform the necessary calculations. The values of the estimators corresponding to parameters γ and ν obtained by the above-described method are $\hat{\gamma} = 83.18889$ and $\hat{v}^2 = 7.52494 \times 10^{-6}$. It can be seen that $\hat{v}^2 > 0$, which justifies the choice of a stochastic model to model these data, because $\nu \approx 0$ would be proof that a deterministic model would have been sufficient. Figures 5 and 6 depict the real data with the EMF and the ECMF of the fitted model respectively.

TABLE 3. Forecasting results.

Year	Data	EMF	ECMF	Lower bound	Upper bound
2015	21.9	21.9814	21.9741	21.3205	22.6577
2016	20.9	21.093	21.0149	20.4491	21.7521
2017	20.1	20.241	20.0557	19.6139	20.8831
2018	19.2	19.4237	19.2885	18.8133	20.0490

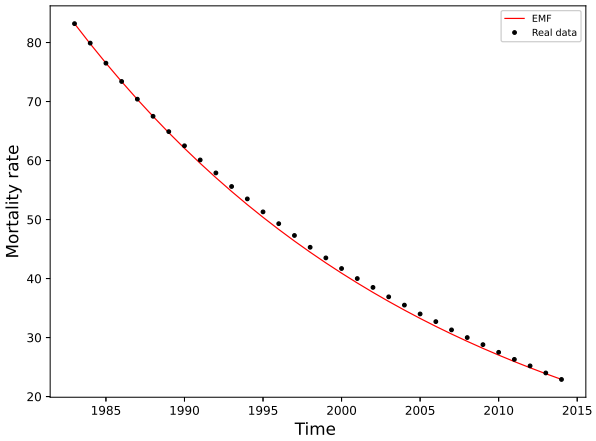

FIGURE 5. Real data vs EMF

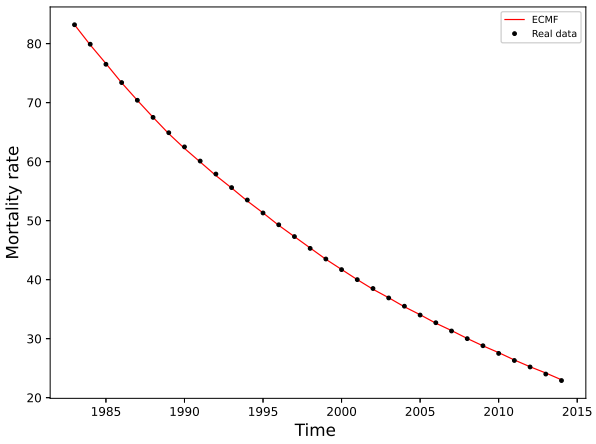

FIGURE 6. Real data vs ECMF

The MAPE and the SMAPE for the proposed model are shown in Table 4.

TABLE 4. Forecast accuracy for the stochastic Pareto model

Forecast accuracy error measurements	Values
MAPE	0.204817854615773
SMAPE	0.204747440507338
MSE	0.0283929225

The MAPE value is less than 10, indicating that the forecasts of the SPD model are "highly accurate"[14], in addition, the value of the MSE is very close to 0 which indicated that the forecast of the model is very accurate. Furthermore, the values of the MAPE, SMAPE, and MSE obtained using the stochastic model are significantly lower than those obtained using the deterministic model.

Finally, to demonstrate the predictive capabilities of the model used, we plotted the data together with the upper bound and lower bound from (4.5) and (4.6), the results of which are shown in Table 3 and Figure 7.

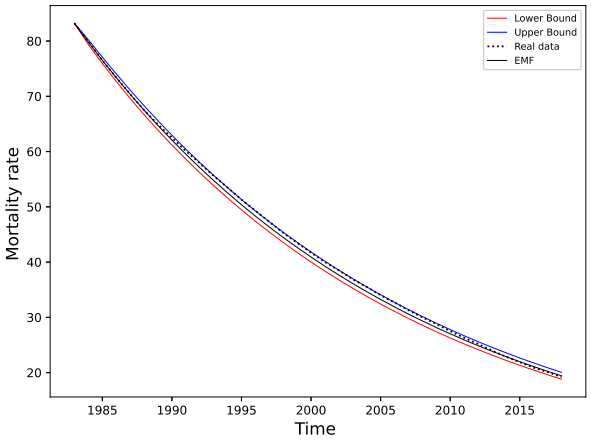

FIGURE 7. Real data vs EMF and upper and lower bounds

Conclusion

We determined the essential probabilistic aspects of the Stochastic Pareto diffusion process and derived its parameter estimators in this work. We derived a set of non-linear equations using the maximum likelihood method based on discrete sampling, which was solved using numerical methods, specifically the simulated annealing method. Finally, the SPDP is used to fit and forecast under-five mortality in Morocco.

Funding

This research has been funded by " A-FQM-228-UGR18. Proyectos I+D+i del Programa Operativo FEDER 2018 Junta de Andalucia ".

References

[1] Alfred Akinsete, Felix Famoye, and Carl Lee. "The beta-Pareto distribution". In: *Statistics* 42.6 (2008), pp. 547–563.
[2] Ayman Alzaatreh, Felix Famoye, and Carl Lee. "Gamma-Pareto distribution and its applications". In: *Journal of Modern Applied Statistical Methods* 11.1 (2012), p. 7.
[3] Ayman Alzaatreh, Felix Famoye, and Carl Lee. "Weibull-Pareto distribution and its applications". In: *Communications in Statistics-Theory and Methods* 42.9 (2013), pp. 1673–1691.
[4] Ludwig Arnold. "Stochastic differential equations". In: *New York* (1974).
[5] World Bank. *Mortality rate, infant (per 1,000 live births)*. 2018. URL: https://data.worldbank.org/indicator/SP.DYN.IMRT.IN.
[6] Abdenbi El Azri and Ahmed Nafidi. "A γ- power stochastic Lundqvist-Korf diffusion process: Computational aspects and simulation". In: *Moroccan Journal of Pure and Applied Analysis (MJPAA)* 8.3 (2022), pp. 364–374.
[7] Julie Lyng Forman and Michael Sørensen. "The Pearson diffusions: A class of statistically tractable diffusion processes". In: *Scandinavian Journal of Statistics* 35.3 (2008), pp. 438–465.
[8] Christiane Fuchs. *Inference for diffusion processes: with applications in life sciences*. Springer Science & Business Media, 2013.
[9] Beno Gutenberg and Charles F Richter. "Frequency of earthquakes in California". In: *Bulletin of the Seismological Society of America* 34.4 (1944), pp. 185–188.
[10] R Gutiérrez, R Gutiérrez-Sánchez, and A Nafidi. "Modelling and forecasting vehicle stocks using the trends of stochastic Gompertz diffusion models: The case of Spain". In: *Applied Stochastic Models in Business and Industry* 25.3 (2009), pp. 385–405.
[11] R Gutiérrez, R Gutiérrez-Sánchez, and A Nafidi. "The trend of the total stock of the private car-petrol in Spain: Stochastic modelling using a new gamma diffusion process". In: *Applied energy* 86.1 (2009), pp. 18–24.
[12] R Gutiérrez et al. "Inference in gompertz-type nonhomogeneous stochastic systems by means of discrete sampling". In: *Cybernetics and Systems* (2005).
[13] Yu A Kutoyants. "Estimation of a parameter of a diffusion process". In: *Theory of Probability & Its Applications* 23.3 (1979), pp. 641–649.
[14] Colin David Lewis. *A Radical Guide to Exponential Smoothing and Curve Fitting*. 1982.
[15] Spyros Makridakis. "Accuracy measures: theoretical and practical concerns". In: *International journal of forecasting* 9.4 (1993), pp. 527–529.
[16] Ed Meynial. *Recueil publié par la Faculté de droit, à l'occasion de l'exposition nationale suisse de Genève*. 1898.
[17] Ahmed Nafidi, Ilyasse Makroz, and Ramón Gutiérrez Sánchez. "A Stochastic Lomax Diffusion Process: Statistical Inference and Application". In: *Mathematics* 9.1 (2021), p. 100.

[18] Ahmed Nafidi, Oussama Rida, Boujemaa Achchab, et al. “Stochastic Diffusion Process Based on Generalized Brody Curve: Application to Real Data”. In: *Journal of Mathematics and Statistics Studies* 2.1 (2021), pp. 01–11.

[19] Ahmed Nafidi et al. “The stochastic Weibull diffusion process: Computational aspects and simulation”. In: *Applied Mathematics and Computation* 348 (2019), pp. 575–587.

[20] Mark E J Newman. “Power laws, Pareto distributions and Zipf’s law”. In: *Contemporary physics* 46.5 (2005), pp. 323–351.

[21] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[22] SP Preston and Andrew TA Wood. “Approximation of transition densities of stochastic differential equations by saddlepoint methods applied to small-time Ito-Taylor sample-path expansions”. In: *Statistics and Computing* 22.1 (2012), pp. 205–217.

[23] William J Reed and Murray Jorgensen. “The double Pareto-lognormal distribution: A new parametric model for size distributions”. In: *Communications in Statistics-Theory and Methods* 33.8 (2004), pp. 1733–1753.

[24] Bianca Schroeder, Sotirios Damouras, and Phillipa Gill. “Understanding latent sector errors and how to protect against them”. In: *ACM Transactions on storage (TOS)* 6.3 (2010), pp. 1–23.

[25] Gerhard Tintner and Jati K Sengupta. *Stochastic economics: stochastic processes, control, and programming*. Elsevier, 2014.

[26] Chris Tofallis. “A better measure of relative prediction accuracy for model selection and model estimation”. In: *Journal of the Operational Research Society* 66.8 (2015), pp. 1352–1362.

[27] Nakahiro Yoshida. “Estimation for diffusion processes from discrete observation”. In: *Journal of Multivariate Analysis* 41.2 (1992), pp. 220–242.