

Document downloaded from:

<http://hdl.handle.net/10251/120535>

This paper must be cited as:

Burgos-Simon, C.; Cortés, J.; Lombana, IC.; Martínez Rodríguez, D.; Villanueva Micó, RJ. (2018). Modelling the dynamics of frequent users of electronic commerce in Spain using optimization techniques for inverse problems with uncertainty. *Journal of Optimization Theory and Applications*. 158(3):1-12. <https://doi.org/10.1007/s10957-018-1382-6>



The final publication is available at

<http://doi.org/10.1007/s10957-018-1382-6>

Copyright Springer-Verlag

Additional Information

Modeling the Dynamics of the Frequent Users of Electronic Commerce in Spain Using Optimization Techniques for Inverse Problems with Uncertainty

Clara Burgos · Juan-Carlos Cortés ·
Iván-Camilo Lombana · David
Martínez-Rodríguez · Rafael-J.Villanueva

Received: date / Accepted: date

Abstract In this paper, we retrieve data about the frequent users of electronic commerce during the period 2011- 2016 from the Spanish National Institute of Statistics. These data, coming from surveys, have intrinsic uncertainty that we describe using appropriate random variables. Then, we propose a stochastic model to study the dynamics of frequent users of electronic commerce. The goal of this paper is to solve the inverse problem that consists of determining the model parameters as suitable parametric random variables, in such a way the model output be capable to capture the data uncertainty, at the time instants where sample data are available, via adequate probability density functions. To achieve the aforementioned goal, we propose a computational procedure that involves building a nonlinear objective function, based on sta-

Clara Burgos
Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain
clabursi@posgrado.upv.com

Juan-Carlos Cortés, Corresponding author
Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain
jccortes@imm.upv.es

Iván-Camilo Lombana
Facultad de Ingeniería Civil - Campus Ibagué
Universidad Cooperativa de Colombia,
Calle 10 No 1-120 Ed. Urrutia, Ibagué - Colombia
ivan.lombana@campusucc.edu.co

David Martínez-Rodríguez
Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain
damarro3@etsii.upv.es

Rafael J.Villanueva
Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain
rjvillan@imm.upv.es

tistical moment measures, to be minimized using a variation of the Particle Swarm Optimization algorithm.

Keywords Inverse problem, Uncertainty Quantification, Random Optimization Computational Methods, Non-linear Stochastic Model, Probability Density Function.

Mathematics Subject Classification (2000) 65C20, 65C60, 65K10

1 Introduction

In the last years, many contributions have dealt with the study of ordinary differential equations and difference equations whose input data (initial/boundary conditions, forcing term and/or coefficients) are assumed to be random variables [1–6], for instance. In most of them, it is assumed that the probability distributions of the parameters are known and follow standard patterns (Uniform, Gaussian, Exponential, etc.), and the aim is to see their effect on the solution, which is a stochastic process. However, when we want to model and study a real phenomenon, to assume that the probability distribution of the parameters are known and standard, may be unreal. This fact is even more apparent when dealing with a mathematical model designed to describe phenomena in Social Sciences.

Therefore, to determine appropriate probability distributions of the model parameters is becoming a key part of the problem when dealing with real applications. In other words, when we try to describe real phenomena, usually it is not enough to build coherent models, but also consider and treat adequately the uncertainty involved in both sample data and model parameters as well as to control their effect on the solution. In this latter sense, a key issue is the computation of the probability distributions of the model parameters that make that the solution stochastic process of the model, at certain time instants, capture the uncertainty embedded in sample data.

There have been several attempts to approach this problem. For instance, in [7,6] the authors use inverse frequentist and least square parametric techniques, respectively, to describe the data uncertainty because the lack of information about how the sample data were obtained and treated. In both approaches, strong assumptions about model parameters are implicitly assumed. In [8], it is proposed a complete computational approach of the problem to obtain the probability distribution of the model parameters, the solution stochastic process and also to capture the data uncertainty via punctual (mean) and probabilistic (confidence intervals) information. To do that, the probability distributions of these data uncertainties need to be previously assigned.

In this paper we want to go beyond and, following the approach presented in [8], we want to capture the data uncertainty via their distributions of probability rather than using the mean and confidence intervals only. That is, given a model and the probability distributions of sample data, describing their

uncertainty, we want to find out the probability distributions of the model parameters such that the probability density functions of the solution stochastic process be as much similar as the probability density functions at the time instants where sampled data are available.

We should say that there are two issues increasing the difficulty of this approach. On the one hand, the data themselves. The sources where we retrieve the data do not always provide enough information to facilitate the assignation of probability distributions that describe their uncertainty. Anyway, if we were able to assign acceptable probability distributions using statistical techniques, this would not guarantee that we were accurately describing the uncertainty of the whole phenomenon. On the other hand, the model itself. Although the model gains flexibility when its parameters are described via probability distributions, we implicitly assume certain behavior once a specific model (pattern) is chosen. Thus, the choice of the model always involves certain rigidities that may prevent that the solution stochastic process does not capture successfully the data uncertainty. **Although the choice of a mathematical model can restrict the subsequent analysis, often making this decision is necessary to perform the corresponding study. Anyway, this choice must be based upon physical tenets supporting the modeling process. Otherwise, heuristic methods can be applied.**

The paper is organized as follows. In Section 2, we retrieve data about the frequent users of e-commerce (EC) in Spain from 2011 until 2016. Then, we will be able to assign probability distributions that capture the uncertainty of each one of the data. In Section 3, we build a nonlinear stochastic model that describes the dynamics of frequent users of EC in Spain. Also, we assign probability distributions to the model parameters, discussing what distributions should be more flexible and appropriate. In Section 4 we describe the computational procedures we implement with the aim to determine what probability distributions for the model parameters will allow the solution stochastic process be as much similar as possible to the data probability density functions at their corresponding time instants, using for this adjustment an inverse optimization technique whose objective function involves statistical moment measures. In Section 5, we describe relevant details of the implementation as well as we show the obtained results. Section 6 is devoted to discussion of the main conclusions derived from our study.

2 Available Data and e-Commerce Model

This section is devoted to present the available data about frequent users of EC that will be used throughout the study. According to source [9], frequent users of EC are those who have bought by the Internet in the last three months.

Table 1 shows available data from Spanish Statistics Institute (INE) [9] about the average proportions (p_k) of people that use EC in Spain in Dec 2011 ($k = 0$), Dec 2012 ($k = 1$), Dec 2013 ($k = 2$), Dec 2014 ($k = 3$), Dec 2015 ($k = 4$) and Dec 2016 ($k = 5$), as well as the sample size (n_k) of the corresponding surveys. The time instants t_k are given in months (see Table 1).

Table 1 Proportion p_k of Spanish people who have bought by the Internet at least once in the last three months during the period 2011 – 2016 and the sample size n_k of the corresponding surveys [9]. The time instants are expressed in months from Dec 2011 (t_0) to Dec 2016 (t_5).

| Date | Time instant (months) | Proportion of users of EC | Survey sample size |
|----------|-----------------------|---------------------------|--------------------|
| Dec 2011 | $t_0 = 0$ | $p_0 = 0.186936$ | $n_0 = 20647$ |
| Dec 2012 | $t_1 = 12$ | $p_1 = 0.218087$ | $n_1 = 20484$ |
| Dec 2013 | $t_2 = 24$ | $p_2 = 0.229510$ | $n_2 = 20815$ |
| Dec 2014 | $t_3 = 36$ | $p_3 = 0.276636$ | $n_3 = 20786$ |
| Dec 2015 | $t_4 = 48$ | $p_4 = 0.321732$ | $n_4 = 23887$ |
| Dec 2016 | $t_5 = 60$ | $p_5 = 0.350582$ | $n_5 = 24132$ |

A key point to quantify the data uncertainty shown in Table 1 is to assign adequate probability distributions of the data at each time instant t_k , $k = 0, 1, 2, 3, 4, 5$. To achieve this goal, we use the sample size of the surveys collected in Table 1 and we assume that people interviewed each year is different and, consequently, the survey outputs are independent. An individual can be a frequent user of EC or not, and these are two mutually exclusive events. Therefore, it is natural to consider that in each time instant t_k , $k = 0, 1, 2, 3, 4, 5$, the users of EC can be represented by a Binomial distribution, $\text{Bi}(n_k, p_k)$, where n_k is the corresponding survey sample size and $p_k \in]0, 1[$ is the proportion of frequent users of EC. For convenience, we will work with the percentage of people who buy or not by the Internet. Thus, it is necessary to scale the Binomial distributions to the interval $]0, 1[$ and we do this dividing by the number of survey samples. In Table 2, we show the scaled Binomial distributions P_k for every time instant t_k , $k = 0, 1, 2, 3, 4, 5$ and the moments of first and second order. According to the statistical properties of Scaled Binomial distribution, the first and second order moments of P_k have been calculated using the following expressions

$$m_k = \mathbb{E}[P_k] = p_k, \quad v_k = \mathbb{E}[P_k^2] = \frac{p_k(1-p_k)}{n_k} + p_k^2. \quad (1)$$

Table 2 Scaled Binomial (Bi) distributions for Spanish EC data collected in Table 1 and their moments of first (m_k) and second order (v_k), $k = 0, 1, \dots, 5$, defined in (1).

| Time instants | Scaled Binomial distribution | Moment of 1st order | Moment of 2nd order |
|---------------|--|---------------------|---------------------|
| $t_0 = 0$ | $P_0 = \frac{1}{20647} \text{Bi}(20647, 0.186936)$ | $m_0 = 0.186936$ | $v_0 = 0.034950$ |
| $t_1 = 12$ | $P_1 = \frac{1}{20484} \text{Bi}(20484, 0.218087)$ | $m_1 = 0.218087$ | $v_1 = 0.047570$ |
| $t_2 = 24$ | $P_2 = \frac{1}{20815} \text{Bi}(20815, 0.229510)$ | $m_2 = 0.229510$ | $v_2 = 0.052683$ |
| $t_3 = 36$ | $P_3 = \frac{1}{20786} \text{Bi}(20786, 0.276636)$ | $m_3 = 0.276636$ | $v_3 = 0.076537$ |
| $t_4 = 48$ | $P_4 = \frac{1}{23887} \text{Bi}(23887, 0.321732)$ | $m_4 = 0.321732$ | $v_4 = 0.103522$ |
| $t_5 = 60$ | $P_5 = \frac{1}{24132} \text{Bi}(24132, 0.350582)$ | $m_5 = 0.350582$ | $v_5 = 0.122917$ |

3 Building a Model to Study the Dynamics of the Frequent Users of e-Commerce in Spain

Now, we introduce the proposed model in order to describe the dynamics of the frequent users of EC in Spain. First, we consider the following subpopulations:

- $X_t^{\bar{B}}$ denotes the percentage or proportion of people who do not use frequently EC at the time instant t (in months),
- X_t^B denotes the percentage or proportion of people who use frequently EC at the time instant t (in months).

Therefore $X_t^{\bar{B}}, X_t^B \in]0, 1[$, for every t . In the period 2011 – 2016, according to [10], the birth and death rates are very close, and consequently, we assume constant population. Also, we are going to consider that non-users of EC may become users by peer influence of the EC users [11], and this is modeled by the term $\beta X_t^{\bar{B}} X_t^B$, where $\beta > 0$ is the transmission parameter. People can leave the users of EC subpopulation (X_t^B), when they do not purchase any item by the Internet in three months, at rate $\gamma > 0$, and it is modeled by the term γX_t^B . The dynamics of people who use and do not use EC are described by the following nonlinear system of difference equations

$$\begin{cases} X_{t+1}^{\bar{B}} = X_t^{\bar{B}} - \beta X_t^{\bar{B}} X_t^B + \gamma X_t^B, \\ X_{t+1}^B = X_t^B + \beta X_t^{\bar{B}} X_t^B - \gamma X_t^B. \end{cases} \quad (2)$$

In Figure 1, we show the model diagram corresponding to dynamic model (2).

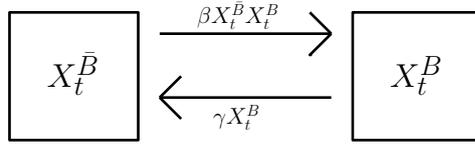


Fig. 1 Compartmental model for the dynamics of the proportion of Spanish frequent users of EC. The boxes represent the subpopulations and the arrows the transitions between subpopulations.

As $X_t^{\bar{B}}$ and X_t^B are probabilities, then $X_t^{\bar{B}} + X_t^B = 1$, for every t . Therefore, dynamic model (2) can be rewritten as:

$$\begin{cases} X_{t+1}^{\bar{B}} = 1 - X_{t+1}^B, \\ X_{t+1}^B = X_t^B + \beta X_t^{\bar{B}} (1 - X_t^B) - \gamma X_t^B. \end{cases} \quad (3)$$

Once we have formulated and simplified the proposed model, taking into account the sampling errors in the surveys, it is natural to consider the model parameters as random variables in order to describe data uncertainty. In a first stage, we are going to consider the parameter β as a positive random variable and γ as a positive constant to be determined.

In [12], the authors show that a reasonable probability distribution for parameter γ is the exponential. In the computer tests and simulations performed, we realized that a small percentage of users of EC moves to the non-users state, which agrees the observations collected in [9]. This means that the mean, and consequently the variance, of the exponential distribution are very small and it supports we can consider model parameter γ as a deterministic constant, rather than a random variable, to be determined.

We should say that transmission parameter β embeds social complex processes involving the frequent use of EC habit. These social processes are usually not studied in standard surveys, as it happens in [9]. Therefore, it is plausible to consider β as a random variable. Then, we are going to assume that random variable β follows a Beta Prime distribution $\text{Be}_p(\beta_1, \beta_2)$ because it is positive and flexible (biparametric) [13].

From [12, p. 352], parameter γ can be interpreted as the average time a user of EC becomes a non-user. Taking into account that the surveys consider that a frequent user of EC changes his/her current state if he/she does not buy by the Internet during three months and the model time step is a month, the model parameter γ will be upper bounded by $1/3$ and should lie in the interval $(0, 1/3)$.

Now, the goal is to seek the parameters $\beta_1 > 0$ and $\beta_2 > 0$ of the positive random variable β , and the real number $\gamma > 0$, such that the model output of (3) fits the best the probability distributions at the corresponding time instants shown in Table 2.

4 Procedure to Solve the Problem Using an Inverse Optimization Technique

In this section, we propose a computational algorithm to find the values of the parameters β_1 , β_2 and γ such that the probability distributions of the model (3) in the time instants $t_0, t_1, t_2, t_3, t_4, t_5$, be as much close as possible to the corresponding probability distributions in Table 2. To achieve this goal, a computational method to minimize the nonlinear error function of the associated inverse problem is proposed.

First, according to second equation of (3), we define the function

$$M(X, \beta, \gamma) = X + \beta X(1 - X) - \gamma X, \quad (4)$$

that allows to calculate the frequent users of EC in a month depending on the frequent users of EC in the previous month.

An important step in our procedure is to obtain the model output every month given the model parameter estimates. This will be done by the output function, say $O(\beta_1, \beta_2, \gamma, T)$, where T denotes the number of months to calculate the model output mo , defined by the following steps

- Initialize the vector mo with size $T + 1$.

- Let x_0 be a value of the initial condition sampled according to scaled Binomial distribution P_0 (see Table 2).
- Let b be a value sampled according to Beta Prime distribution $\beta = \text{Be}_p(\beta_1, \beta_2)$.
- Assign $mo(1) = x_0$.
- FOR $i = 1$ TO T
 - Calculate $mo(i + 1) = M(mo(i), b, \gamma)$.
- RETURN mo .

We will need to repeat several times the evaluation of the above function $O(\beta_1, \beta_2, \gamma, T)$. Then, we define a function that performs this task. Let us define the function $S(\beta_1, \beta_2, \gamma, N, T)$, where N is the number of times we evaluate the function O , defined by

- Initialize the matrix OUT of size $N \times T$.
- FOR $i = 1$ to N
 - Calculate the row $OUT(i) = O(\beta_1, \beta_2, \gamma, T)$.
- RETURN OUT

The above algorithm returns a matrix of size $N \times T$, where each row is the vector-realization (model output) obtained with a given sampled values of x_0 and β .

Now, we introduce the fitness function that will help us to determine the best model parameters β_1 , β_2 and γ . The fitness function applies the above algorithm to obtain a matrix OUT with N vector-realizations. However, we do not need all the elements of the matrix OUT , we need all the rows but only in the time instants t_k , for k from 0 to 5, where we have data and then we can compare. Let us define the fitness function $F(\beta_1, \beta_2, \gamma)$ as follows

- Let N be a fixed value, large enough.
- Set $T = 60$.
- Calculate the matrix $OUT = S(\beta_1, \beta_2, \gamma, N, T)$.
- FOR $k = 0$ to 5
 - Take the column t_k of the matrix OUT , that is, the N model outputs corresponding to the same time instant t_k $\{OUT(1, t_k), \dots, OUT(N, t_k)\}$ and calculate M_k , its mean (moment of first order), and V_k , its moment of second order.
- Calculate

$$E = |M_0 - m_0| + \dots + |M_5 - m_5| + |V_0 - v_0| + \dots + |V_5 - v_5|.$$

- RETURN E

The fitness function $F(\beta_1, \beta_2, \gamma)$ is a positive real valued function defined in the domain $\mathbb{R}^+ \times \mathbb{R}^+ \times (0, 1/3)$. The initial domain is too wide and, after some runs, it will be reduced.

Now, to find the appropriate values of β_1 , β_2 and γ , we have to minimize the fitness function F . To do that, we use the Random Particle Swarm Optimization (rPSO) presented in [14]. This optimization algorithm is a variation

of the classical PSO algorithm where randomness is introduced in the calculation of the new velocity of the particles. rPSO has shown to perform well fitting models with data uncertainty [15].

Then, we apply the above procedure to fit probabilistically the model to the data distributions in Table 2. Now, we describe some extra features included in our version of the rPSO algorithm:

1. We include 10% of mutation probability.
2. We apply rPSO with 90 particles. 60 of them follow the classical PSO updating procedure. The remainder 30, called *explorers*, and their updates are stored in a repository. Therefore, the number of explorers stored in the repository increases as the times goes on. Thus, in each iteration (an iteration corresponds to the evaluation of a whole generation of particles), we calculate the center of mass of a sample of 1/3 of the particles in the repository and update the velocities of the explorers taking directions moving away the center of mass.

Observe that the above features will allow us to explore more extensively the space of parameters.

5 Results

For the evaluations of the fitness function $F(\beta_1, \beta_2, \gamma)$, we have made the decision of taking $N = 1000$ and the parameters search space limited to $\beta_1 \in [1000, 10000]$, $\beta_2 \in [50000, 150000]$ and $\gamma \in (0, 1/3)$. The interval for γ was justified previously. The intervals for β_1 and β_2 have been chosen after some runs and taking into account that the Beta Prime function has a big variance if its parameters are small.

Respect to the rPSO algorithm, as we said before, we consider 90 particles with a maximum number of 5000 iterations. **The procedure was run in a PC with Intel i5 CPU and 16GB of RAM under MS-Windows 10. The code was implemented in Python 2.7 and the fitting procedure took 1 hour and 52 minutes.**

The values of the parameters with best fitness have been

$$\beta_1 = 1985.7534, \beta_2 = 13820.0903, \gamma = 7.02012261 \times 10^{-5}, \quad (5)$$

being the fitness

$$F(1985.7534, 13820.0903, 7.02012261 \times 10^{-5}) = 0.0435104687349.$$

Graphically, the result of the fitting can be seen in Figure 2. Here, we evaluate the function $S(\beta_1, \beta_2, \gamma, 1000, 60)$ with β_1, β_2, γ given in (5), and we build the histograms of the $N = 1000$ model output values in the time instants $t_k, \{OUT(1, t_k), \dots, OUT(N, t_k)\}$, for $t_k = 12, 24, 36, 48, 60$ months (in blue). We compare these histograms with the histograms of the data probability

distributions P_1, P_2, P_3, P_4, P_5 of Table 2 (in red). Except for Dec 2013, the similarity of the histograms in each time instant is remarkable because both histograms share most part of their uncertainties.

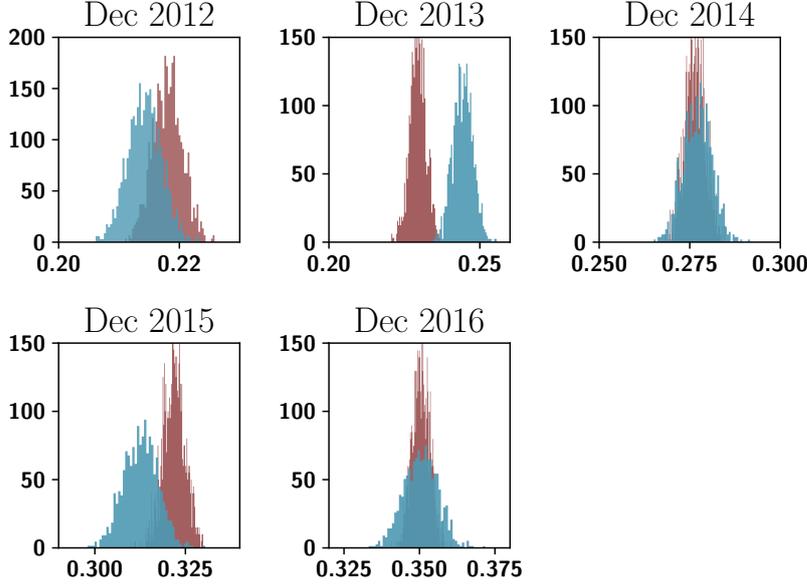


Fig. 2 Comparison between the histograms of the $N = 1000$ model outputs in the time instants $t_1 = \text{Dec 2012}$, $t_2 = \text{Dec 2013}$, $t_3 = \text{Dec 2014}$, $t_4 = \text{Dec 2015}$, $t_5 = \text{Dec 2016}$ (in blue) and their corresponding data histograms of the scaled Binomial distributions given in Table 2 (in red).

In Figure 3, we can see the obtained probability density function of the transmission parameter β , the Beta Prime distribution of the random variable $\beta = \text{Be}_p(1985.7534, 13820.0903)$.

As the fitting has been good, we can evaluate $S(\beta_1, \beta_2, \gamma, 1000, 108)$ with β_1, β_2, γ given in (5) and then build the histograms of the $N = 1000$ model output values in the time instants $t_k, \{OUT(1, t_k), \dots, OUT(N, t_k)\}$, for $t_k = 72, 84, 96, 108$ to perform model predictions of the probability distributions of the frequent users of EC over Dec 2017, Dec 2018, Dec 2019 and Dec 2020, respectively. Taking $T = 108$ months, we reach Dec 2020. The graphical results are illustrated in the Figure 4.

We conclude this section showing one more graph. In Figure 5, we show the 95% confidence intervals of the data distribution (scaled Binomial random variables in Table 2) represented by the red points, and the model 95% confidence intervals represented by the green lines at every month from 2011–

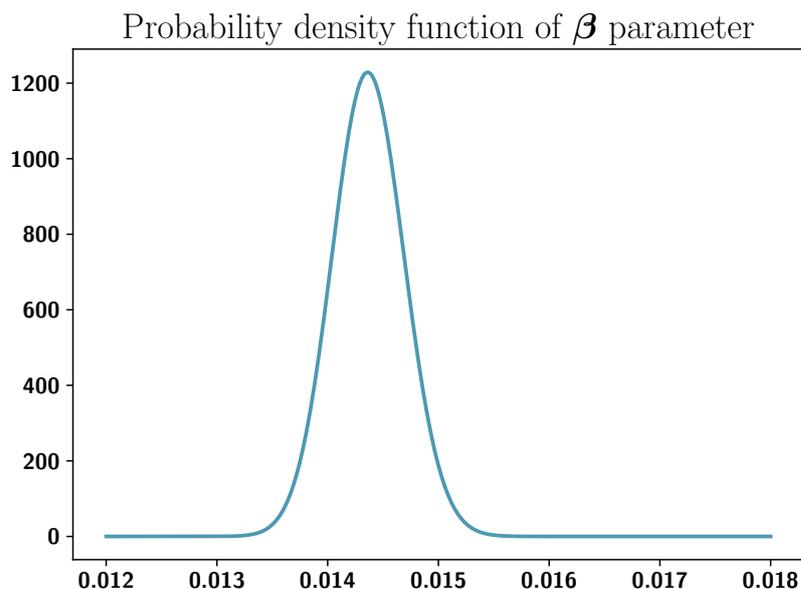


Fig. 3 Probability density function of the random variable transmission parameter $\beta = \text{Be}_p(\beta_1, \beta_2)$ where $\beta_1 = 1985.7534$ and $\beta_2 = 13820.0903$ are the values that provide the best fitness. Note that the variability is quite small.

2016 (estimation) and from 2017–2020 (prediction). This is another form to visualize the goodness-of-fit of the model and the prediction.

6 Conclusions

In this paper we have described a computational procedure to determine the probability distribution we can assign to the model parameters in such a way the solution stochastic process of a random dynamic discrete model can capture the data uncertainty expressed through probability distributions. The study has been conducted via real data of the use of electronic commerce in Spain. The obtained estimation has been validated and it allowed us to perform reliable predictions over the next four years using histograms. Nevertheless, we must say that the capture of the uncertainty related to the datum in Dec 2013 has not been good. This makes us return to the discussion in the Introduction section: maybe, the data uncertainty follow different distributions to the ones we proposed; or the model is not as flexible as we expected and it is not able to capture accurately the whole data uncertainty; or something happened in the year 2013 (something we do not know) such that the datum of Dec 2013, and consequently its uncertainty, did not follow the same pattern as the other data.

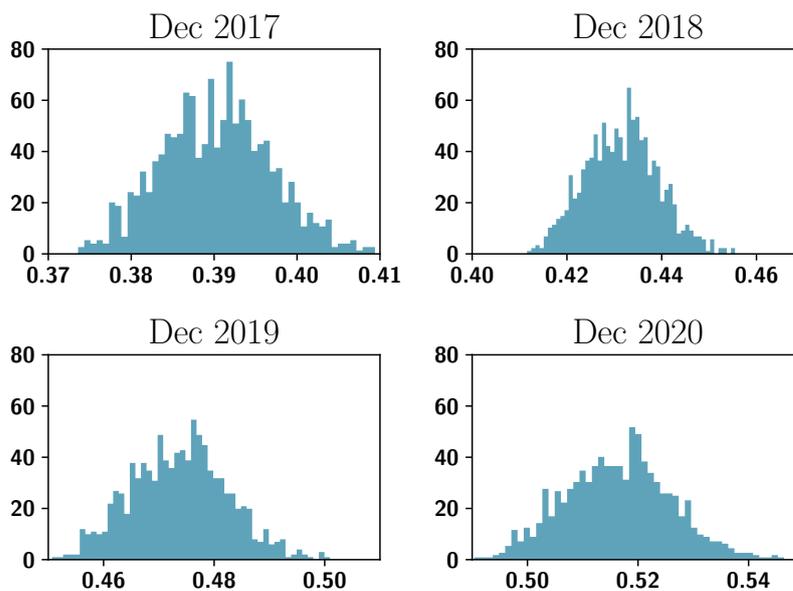


Fig. 4 Histograms of the $N = 1000$ model outputs in the future time instants Dec 2017, Dec 2018, Dec 2019 and Dec 2020.

Also, we have to take into account that the proposed nonlinear stochastic model has only two parameters, being one of them a random variable and the other a deterministic value. Furthermore, previously to the application of the optimization procedure, we had to narrow the space of parameters and to determine the appropriate search area. Anyway, the search space has been very wide. Then, if we want to extend this procedure to more complex models, a suitable reduction of the area to seek the model parameters should be considered.

In the future, we expect to design new computational techniques more general where we can avoid the assignment of probability distributions to data and/or model parameters and then, to obtain a solution stochastic process able to completely capture the data uncertainty.

Acknowledgements This work has been supported by the Ministerio de Economía, Industria y Competitividad grant MTM2017-89664-P.

References

1. Dorini, F.A., Cecconello, M.S., Dorini, M.B.: On the logistic equation subject to uncertainties in the environmental carrying capacity and initial population density. Com-

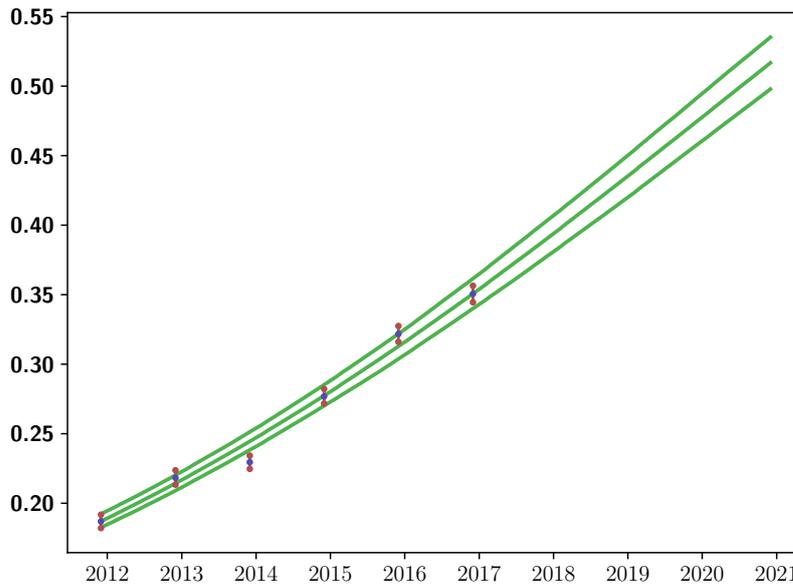


Fig. 5 In this figure, the red and blue points represent, respectively, the mean and the 95% confidence interval of the data corresponding to Dec 2011, Dec 2012, Dec 2013, Dec 2014, Dec 2015 and Dec 2016. The green lines represent, respectively, the average and 95% confidence interval for all the months from Dec 2011 until Dec 2020. From Jan 2017, the green lines represent the model prediction.

- munications in *Nonlinear Science and Numerical Simulation* **33**, 160–173 (2016). DOI 10.1016/j.cnsns.2015.09.009
2. Hussein, A., Selim, M.M.: Solution of the stochastic radiative transfer equation with Rayleigh scattering using RVT technique. *Applied Mathematics and Computation* **218**(13), 7193–7203 (2012). DOI 10.1016/j.amc.2011.12.088
 3. Dorini, F., Cunha, M.: Statistical moments of the random linear transport equation. *Journal of Computational Physics* **227**(19), 8541–8550 (2008). DOI 10.1016/j.jcp.2008.06.002
 4. Xu, Z., Tipireddy, R., Lin, G.: Analytical approximation and numerical studies of one-dimensional elliptic equation with random coefficients. *Applied Mathematical Modelling* **40**(9–10), 5542–5559 (2016). DOI 10.1016/j.apm.2015.12.04
 5. Mourad, K., Debbouche, A.: Complete controllability of nonlocal fractional stochastic differential evolution equations with Poisson jumps in Hilbert spaces. *International Journal of Advances in Applied Mathematics and Mechanics* **3**(1), 41–48 (2015)
 6. Casabán, M.C., Cortés, J.C., Navarro-Quiles, A., Romero, J.V., Roselló, M.D., Villanueva, R.J.: A comprehensive probabilistic solution of random SIS-type epidemiological models using the random variable transformation technique. *Communications in Nonlinear Science and Numerical Simulation* **32**, 199–210 (2016). DOI 10.1016/j.cnsns.2015.08.009
 7. Casabán, M.C., Cortés, J.C., Navarro-Quiles, A., Romero, J.V., Roselló, M.D., Villanueva, R.J.: Computing probabilistic solutions of the Bernoulli random differential equation. *Journal of Computational and Applied Mathematics* **309**, 396–407 (2017). DOI 10.1016/j.cam.2016.02.034

8. Cortés, J.C., Santonja, F.J., Tarazona, A.C., Villanueva, R.J., Villanueva-Oller, J.: A probabilistic estimation and prediction technique for dynamic continuous social science models: The evolution of the attitude of the basque country population towards ETA as a case study. *Applied Mathematics and Computation* **264**, 13–20 (2015). DOI 10.1016/j.amc.2015.03.128. URL <https://doi.org/10.1016/j.amc.2015.03.128>
9. Spanish INE. Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares (Survey on equipment and use of the information technologies and communication in the household). URL http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176741&menu=resultados&idp=1254735576692
10. Spanish INE. Indicadores demográficos básicos (Basic demographic indicators). URL http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177003&menu=resultados&idp=1254735573002
11. Christakis, N.A., Fowler, J.H.: *Connected. The surprising power of our social networks and how they shape our lives*. Little, Brown & Company (2009). URL <http://www.connectedthebook.com>
12. Brauer, F., Castillo-Chávez, C.: *Mathematical Models in Population Biology and Epidemiology*. Springer New York (2001). DOI 10.1007/978-1-4757-3516-1. URL <https://doi.org/10.1007/978-1-4757-3516-1>
13. Norman, L., Kotz, S., Balakrishnan, N.: *Continuous univariate distributions* (1994)
14. Khemka, N., Jacob, C.: Exploratory toolkit for evolutionary and swarm-based optimization. *The Mathematica Journal* **11**(3), 376–391 (2010). DOI 10.3888/tmj.11.3-5. URL <https://doi.org/10.3888%2Ftmj.11.3-5>
15. Acedo, L., Burgos, C., Hidalgo, J.I., Sánchez-Alonso, V., Villanueva, R.J., Villanueva-Oller, J.: Calibrating a large network model describing the transmission dynamics of the human papillomavirus using a particle swarm optimization algorithm in a distributed computing environment. *The International Journal of High Performance Computing Applications* p. 109434201769786 (2017). DOI 10.1177/1094342017697862. URL <https://doi.org/10.1177/1094342017697862>