Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Statistical Estimation Based on Generalized Order Statistics from Kumaraswamy Distribution

Samir K. Safi
Department of Statistics, The Islamic University of Gaza, Palestine
(E-mail: **samirsafi@gmail.com**)

Rehab H. Al Sheikh Ahmed
1Statistician Consultant, Gaza, , Palestine
(E-mail: **rehab4ps@yahoo.com**)

Abstract. The Kumaraswamy distribution is similar to the Beta distribution but has the key advantage of a closed-form cumulative distribution function. In this paper we present the estimation of Kumaraswamy distribution parameters based on Generalized Order Statistics (GOS) using Maximum Likelihood Estimators (MLE). We proved that the parameters estimation for Kumaraswamy distribution can not be obtained in explicit forms, and therefore it has been implemented using the simulated data for illustrative purposes. We compare the performances of parameters estimation through an extensive numerical simulation for different sample sizes. These simulations examine the sensitivity of estimation to different sample sizes. In particular, how do estimations perform for small, moderate and large sample sizes? The main findings are: First, the worst performance estimation for small sample size selection for different values of the parameters estimation. Secondly, as the sample size increases the MSE of the estimation decreases. Finally, the estimation accuracy reaches its superiority for large sample sizes.

Keywords: Kumaraswamy Distribution, Generalized Order Statistics, Simulation, Maximum Likelihood Estimators.

1 Introduction

Poondni Kumaraswamy was a leading Indian engineer and hydrologist. Kumaraswamy[9] introduced the distribution for variables that are lower and upper bounded. With its two non-negative shape parameters p and q, it was originally conceived to model hydrological phenomena, (See for example Mitnik[10]). The Kumaraswamy distribution is a continuous probability distribution with double-bounded support, defined on the interval [0,1] differing in the values of their two non-negative shape parameters p and q. It is similar to the Beta distribution but has the key advantage of a closed form cumulative distribution function, Carrasco *et al.*[4].

Generalized Order Statistics (GOS) concept was introduced by Kamps[8] as a unified approach to several models of ordered random variables such as upper order statistics, upper record values, sequential order statistics, ordering via truncated distributions, censoring schemes, among others. Ateya and Ahmad[3], Jaheen[7], Habibullah and Ahsanullah[6], Raqab and Ahsanullah [11]among others, utilized the GOS in their works.

Abu El-Fotouh and Nassar[1] have investigated the estimation problem for the unknown parameters of Weibull extension model based on GOS by

Maximum Likelihood Estimators (MLE). Alkasasbeh and Raqab[2] considered the MLE of the different parameters of a generalized logistic distribution and compared the performances of these procedures through an extensive numerical simulation.

This paper is organized as follows. Section 2 presents some basic definitions; Section 3 demonstrates the estimation of Kumaraswamy distribution parameters based on GOS Using MLE, the main results of this paper are stated and proved. Simulation study is shown in section 4; and Section 5 summarizes the important results.

2 Preliminaries

In this section, we introduce some basic definitions.

Definition 2.1 The random variables $X(1,n,\underline{m},k),...,X(n,n,\underline{m},k)$ are called GOS based on the Cumulative Distribution Function (*cdf*), $F(x)$, if their joint probability density function *(pdf)* is given by Kamps[8].

$$f(x_1,...,x_n)=k\left(\prod_{j=1}^{n-1}\gamma_j\right)\left[\prod_{i=1}^{n-1}(1-F(x_i))^{m_i}f(x_i)\right](1-F(x_n))^{k-1}f(x_n), \quad (2.1)$$

on the cone $F^{-1}(0)<X_1\le X_2\le ...\le X_n<F^{-1}(1)$ of $\mathbb{R}^n$, with parameters $n\in\mathbb{N}$, $n\ge 2$, $k>0$, $\underline{m}=(m_1,...,m_{n-1})\in\mathbb{R}^{n-1}$, $M_r=\sum_{j=r}^{n-1}m_j$, such that $\gamma_r=k+n-r+M_r>0$ for all $r\in\{1,...,n-1\}$, let $c_{r-1}=\prod_{j=1}^{r}\gamma_j$, $r=1,2,\mathrm{K},n-1$ and $\gamma_n=k$.

Special Case

Given Definition 2.1, let $k=1$ and $m_1=m_2=...=m_{n-1}=zero$, then the joint *pdf* of all Ordinary Order Statistics (OOS) is $f(x_1,...,x_n)=\left[\prod_{j=1}^{n-1}\gamma_j\right]\left[\prod_{i=1}^{n}f(x_i)\right]$

$\left[\prod_{j=1}^{n-1}\gamma_j\right]=\prod_{j=1}^{n-1}(k+n-j+M_j)$, $M_j=\sum_{r=j}^{n-1}m_r=zero$. Then

$\left[\prod_{j=1}^{n-1}\gamma_j\right]=\prod_{j=1}^{n-1}(1+n-j)=n(n-1)(n-2)...3\times2\times1=n!$.

Therefore, the joint *pdf* of all OOS $f(x_1,...,x_n)=n!\prod_{i=1}^{n}f(x_i)$, *which is the well known pdf of all OOS*.

Definition 2.2 The *pdf* of the Kumaraswamy distribution is given by

$$f_Z(z)=\frac{1}{(b-c)}pq\left(\frac{z-c}{b-c}\right)^{p-1}\left[1-\left(\frac{z-c}{b-c}\right)^{p}\right]^{q-1} \quad , c<z<b, \qquad (2.2)$$

with shape parameters $p>0$ and $q>0$, and boundary parameters *c* and *b*. The standard form of the Kumaraswamy density function $(c=0, b=1)$, $kum(p,q)$ is given by

$$f_X(x;p,q)=pq\,x^{p-1}\left(1-x^{p}\right)^{q-1} \qquad (2.3)$$

The closed form of the *cdf* of the Kumaraswamy distribution is given by

$$F(x;p,q)=1-\left(1-x^{p}\right)^{q} \qquad (2.4)$$

Definition 2.3 The joint *pdf* of $X(1,n,\vec{m},k)$, $X(2,n,\vec{m},k),\ldots,X(n,n,\vec{m},k)$ for Kumaraswamy distribution is

$$f(x_1,\ldots,x_n)=k\left[\prod_{j=1}^{n-1}\gamma_j\right]\prod_{i=1}^{n-1}\left[\left(1-x_i^{\,p}\right)^{q m_i}\,pq\,x_i^{\,p-1}\left(1-x_i^{\,p}\right)^{q-1}\right]\left(1-x_n^{\,p}\right)^{q(k-1)}\times$$

$$pq\,x_n^{\,p-1}\left(1-x_n^{\,p}\right)^{q-1} \qquad (2.5)$$

3. Estimation of Kumaraswamy Distribution Parameters Based on GOS Using MLE

The method of MLE is, by far, the most popular technique for deriving estimators and a reasonable choice for an estimator. The MLE is the parameter point for which the observed sample is most likely. In general, the MLE is and a good point estimator, possessing some of the optimality properties such as invariance property of MLE, Casella and Berger[5]. In this section, the estimation of Kumaraswamy distribution parameters based on GOS using MLE will be derived. Furthermore, the estimation of Kumaraswamy distribution parameters based on OOS will be derived as special case when $k=1$ and $m=0$.

Theorem 3.1 Let $X(1,n,\vec{m},k)$, $X(2,n,\vec{m},k),\ldots,X(n,n,\vec{m},k)$ be *n* GOS for Kumaraswamy with parameters *p* and *q*, i.e. *X* has *kum(p,q)*. The estimation of Kumaraswamy distribution parameters based on GOS for *p* and *q* are given by

$$\hat{p}=-n\left[\sum_{i=1}^{n-1}\left(\frac{\ln x_i\left(1-\hat{q}\left(m_i+1\right)x_i^{\hat{p}}\right)}{1-x_i^{\hat{p}}}\right)+\frac{\ln x_n\left(1-qkx_n^{\hat{p}}\right)}{1-x_n^{\hat{p}}}\right]^{-1} \qquad (3.1)$$

and

$$\hat{q} = -n\left[\sum_{i=1}^{n-1}(m_i+1)\ln\left(1-x_i^{\hat{p}}\right)+k\,\ln\left(1-x_n^{\hat{p}}\right)\right]^{-1}, \qquad (3.2)$$

respectively.

Proof. The likelihood function for GOS for *kum(p,q)* is defined by

$$L(p,q;x) = k\left[\prod_{j=1}^{n-1}\gamma_j\right]\prod_{i=1}^{n-1}\left[\left(1-x_i^{\,p}\right)^{q m_i}\, p\,q\,x_i^{\,p-1}\left(1-x_i^{\,p}\right)^{q-1}\right]\left(1-x_n^{\,p}\right)^{q(k-1)}\times$$

$$p\,q\,x_n^{\,p-1}\left(1-x_n^{\,p}\right)^{q-1}$$

Collecting terms, $L(p,q;x)$ can be written as

$$L(p,q;x) = k\left[\prod_{j=1}^{n-1}\gamma_j\right]\left[\prod_{i=1}^{n-1}(1-x_i^{\,p})^{q(m_i+1)-1}\,pqx_i^{\,p-1}\right](1-x_n^{\,p})^{qk-1}\,pqx_n^{\,p-1} \qquad (3.3)$$

While this function in (3.3) is not all that hard to differentiate, it is much easier to differentiate the log likelihood. Now take logarithm on both sides of (3.3) to get

$$ln\left[L(p,q;x)\right] =$$

$$ln\,k + \sum_{j=1}^{n-1} ln\,\gamma_j + \sum_{i=1}^{n-1}\left[\left(q(m_i+1)-1\right)ln\left(1-x_i^{\,p}\right)+ln\,p+ln\,q+(p-1)ln\,x_i\right]$$

$$+(qk-1)ln\left(1-x_n^{\,p}\right)+ln\,p+ln\,q+(p-1)ln\,x_n$$

Take the first partial derivatives, with respect to p and q, and collecting terms, we find that

$$\frac{\partial ln\left[L(p,q;x)\right]}{\partial p} = \sum_{i=1}^{n-1}\left[\frac{\ln x_i\left(1-q(m_i+1)x_i^{\,p}\right)}{1-x_i^{\,p}}\right]+\frac{\ln x_n\left(1-qkx_n^{\,p}\right)}{1-x_n^{\,p}}+\frac{n}{p},\ \text{and}$$

$$\frac{\partial ln\left[L(p,q;x)\right]}{\partial q} = \sum_{i=1}^{n-1}\left[(m_i+1)\ln\left(1-x_i^{\,p}\right)\right]+k\,\ln(1-x_n^{\,p})+\frac{n}{q},$$

respectively.

Setting these first partial derivatives equal to zero and solving for p and q, yield the solution

$$\hat{p} = -n\left[\sum_{i=1}^{n-1}\left(\frac{\ln x_i\left(1-\hat{q}(m_i+1)x_i^{\hat{p}}\right)}{1-x_i^{\hat{p}}}\right)+\frac{\ln x_n\left(1-qkx_n^{\hat{p}}\right)}{1-x_n^{\hat{p}}}\right]^{-1},\ \text{and}$$

$$\hat{q} = -n\left[\sum_{i=1}^{n-1}(m_i+1)\ln\left(1-x_i^{\hat{p}}\right)+k\,\ln\left(1-x_n^{\hat{p}}\right)\right]^{-1},$$

respectively.

Evaluating the second derivative at $p=\hat{p}$ and $q=\hat{q}$ yield $\left.\frac{\partial^2 ln\left[L(p,q;x)\right]}{\partial p^2}\right|_{p=\hat{p}} = -\frac{n}{\hat{p}^2} < 0$, and $\left.\frac{\partial^2 ln\left[L(p,q;x)\right]}{\partial q^2}\right|_{q=\hat{q}} = -\frac{n}{\hat{q}^2} < 0$, then each of $\hat{p}$ and $\hat{q}$ is the local maximum, and since they are the only values obtained when the first partial derivatives are equal to zero, then $\hat{p}$ and $\hat{q}$ are the global maximum for the likelihood function $ln\left[L\left(p,q;x\right)\right]$. This completes the proof of the Theorem. □

Corollary 3.1 The estimation of Kumaraswamy distribution parameters based on OOS for p and q are given by

$$\hat{p} = -n\left(\sum_{i=1}^{n}\left[\frac{\ln x_i\left(1-\hat{q}x_i^{\hat{p}}\right)}{1-x_i^{\hat{p}}}\right]\right)^{-1}, \qquad (3.4)$$

and

$$\hat{q} = -n\left(\sum_{i=1}^{n}\ln\left(1-x_i^{\hat{p}}\right)\right)^{-1}, \qquad (3.5)$$

respectively.

Proof. Let $m=0$ and $k=1$ in (3.1), then $\hat{p} = -n\left[\sum_{i=1}^{n-1}\left[\frac{\ln x_i\left(1-\hat{q}x_i^{\hat{p}}\right)}{1-x_i^{\hat{p}}}\right]+\frac{\ln x_n\left(1-\hat{q}x_n^{\hat{p}}\right)}{1-x_n^{\hat{p}}}\right]^{-1}$ and collecting terms, we get (3.4).

Let $m=0$ and $k=1$ in (3.2), then $\hat{q} = -n\left[\sum_{i=1}^{n-1}\ln\left(1-x_i^{p}\right)+\ln\left(1-x_n^{\hat{p}}\right)\right]^{-1}$ and collecting terms, we get (3.5). This completes the proof of the Corollary. □

Equations (3.1), (3.2), (3.4) and (3.5) are complicated and consequently computer facilities and numerical solutions are needed to compute $\hat{p}$ and $\hat{q}$.

4 Simulation Study

In this section, since there are no closed forms for the estimation of Kumaraswamy distribution parameters, we consider the simulation technique for the estimation of Kumaraswamy distribution parameters p and q for different sample sizes. These simulations examine the sensitivity of estimation to different sample sizes. In particular, how do estimations perform for small, moderate and large sample sizes?

Definition 1. The efficiency of the parameter estimation for sample size n_1 relative to that of n_2 in terms of the Mean Squared Error (MSE) of the parameter p, $RE(\hat{p})$, is given by

$$RE(\hat{p}) = \frac{n_2 \sum_{i=1}^{r} \left(\hat{p}_{i.n_1} - p_{n_1}\right)^2}{n_1 \sum_{i=1}^{r} \left(\hat{p}_{i.n_2} - p_{n_2}\right)^2}, \qquad (4.1)$$

where r represents the number of simulations. Note that, p_{n_1} and p_{n_2} are the true parameters values for the two samples sizes n_1 and n_2, respectively. A ratio greater than one indicates that the parameter estimation for the sample size n_1 is less efficient than sample size n_2 estimate, and if $RE(\hat{p})$ is close to one, then the parameter estimation for the sample size n_1 is nearly as efficient as sample size n_2 estimate. We will look for different pairs of parameters for Kumaraswamy distribution that we can use to characterize the efficiency ratio, such as $p = q = 1$, $p = 1, q = 2$, $p = 2, q = 1$ and $p = q = 2$. We will try to find an answer to the following question: How robust are Kumaraswamy parameter estimations for different sample sizes?

4.1 The Simulation Setup

Three finite sample sizes (50, 200, and 500) and four values for the parameters p and q. We also generated a simulation of length 500 observations for each of the selected parameters; (p,q): (1,1), (1,2), (2,1) and (2,2).

4.2 The Simulation Results for $RE(\hat{p})$

Table (4.1) shows the complete simulation results for all selected parameters for Kumaraswamy distribution; (p,q): (1,1), (1,2), (2,1) and (2,2) for three finite sample small, moderate and large sizes (50, 200, and 500). The estimated values of the parameters and their corresponding MSEs are given. In addition, the ratios of parameter estimation p, $RE(\hat{p})$, for sample size n_1 relative to that of n_2 in terms of the mean squared error are shown.

Looking at the Table (4.1), we see that for the parameter $p = 1$, the relative efficiency of the MSE for estimating the parameter $p = 1$ with sample size 50 with respect to 200 equals 19.3. This means the parameter estimation error for small sample size ($n = 50$) is about 19 times for moderate sample size ($n = 200$). While the relative efficiency of the MSE for estimating the parameter $p = 1$ with sample size 50 with respect to 500 equals 113.0. This means the parameter estimation error for small sample size ($n = 50$) is about 113 times for large sample size ($n = 500$). This result is the worst performance

of small sample size selection as compared to large sample size for estimating the parameter p . In addition, the relative efficiency of the MSE for estimating the parameter $p=1$ with sample size 200 with respect to 500 equals 5.9. This means the parameter estimation error for moderate sample size ($n=200$) is about 6 times for large sample size ($n=500$). This result indicates that as the sample size increases the MSE of the estimated parameters decreases. This indicates that the MLE tend to its true parameters values. In other words, the estimation accuracy reaches its superiority as the sample size gets larger and larger. Results for the other sample sizes and different parameter choices for Kumaraswamy distribution demonstrate a similar pattern as shown in Table (4.1).

Table 4.1. Estimation of the Parameters for Kumaraswamy Distribution p and q for Different Sample Sizes

Parameters	Sample Size	$\hat{p}$	$MSE(\hat{p})$	$RE(\hat{p})$	$\hat{q}$	$MSE(\hat{q})$	$RE(\hat{q})$
p = q =1	50	1.00115	0.000379	19.3*	0.988747	0.000166691	6.2*
	200	0.994804	0.000020	113.0**	0.988783	0.000027	37.4**
	500	0.994319	0.000003	5.9***	0.985543	0.000004	6.1***
p = 1 , q = 2	50	0.852075	0.002560	17.9*	1.447103	0.020512	27.2*
	200	0.945802	0.000143	387.6**	1.767986	0.000755	173.9**
	500	0.971881	0.000007	21.7***	1.877155	0.000118	6.4***
p = 2 , q =1	50	1.638147	0.005442	7.9*	0.931092	0.000379	5.1*
	200	1.660116	0.000692	29.7**	0.893285	0.000074	9.6**
	500	1.718804	0.000183	3.8***	0.91861	0.000039	1.9***
p = q =2	50	1.320175	0.019067	50.4*	1.291178	0.050228212	77.1*
	200	1.852513	0.000379	299.5**	1.829603	0.000652	462.1**
	500	1.940557	0.000064	5.9***	1.894831	0.000109	6.0***

* RE estimate of n_1=50 relative to n_2=200.
**RE estimate of n_1=50 relative to n_2=500.
*** RE estimate of n_1=200 relative to n_2=500.

5. Conclusions

This paper deals with the estimation of Kumaraswamy distribution parameters using maximum likelihood estimators. Statistical estimation of Kumaraswamy distribution parameters have been derived based on generalized order statistics. Special cases are also deduced for ordinary order statistics. The resulting

equations are complicated and numerical solutions for parameters p and q is recommended. The simulation technique is discussed.

References

1. S. Abu El-Fotouh and M. Nassar. Estimation for the Parameters of the WeibullExtension Model Based on Generalized Order Statistics. *Int. J. Contemp. Math. Sciences*, 6, 36, 1749 – 1760, 2011.
2. M. Alkasasbeh and M. Raqab. Estimation of the Generalized Logistic Distribution Parameters: Comparative Study. *Statistical Methodology*, 6, 3, 262-279, 2009.
3. S. Ateya and A. Ahmad. Inferences based on generalized order statistics under truncated type I generalized logistic distribution, *Statistics,* 45, 389-402*,* 2011.
4. J. Carrasco, S. Ferrari and G. Cordeiro. A New Generalized Kumaraswamy Distribution, 2010.
5. G. Casella and R. Berger. *Statistical Inference*, second edition, Duxbury, California, 2002.
6. M. Habibullah and M. Ahsanullah. Estimation of parameters of a Pareto distribution by generalized order statistics, *Commun. Statist.-Theor. Meth., 29, 1597-1609,* 2000.
7. Z. Jaheen. Estimation based on generalized order statistics from the Burr model, *Commun. Statist.- Theor. Meth.,* 34, 785-794, 2005.
8. U. Kamps. A Concept of Generalized Order Statistics, *ELSEVIER Journal of Statistical planning and inference,* 48, 1995.
9. P. Kumaraswamy. A generalized probability density function for double bounded random processes. *Journal of Hydrology*. 46, 79-88, 1980.
10. P. Mitnik. New Facts About Kumaraswamy Distribution. Department of Sociology, University of Wisconsin-Madison, USA, 2008.
11. M. Raqab and M. Ahsanullah. Estimation of the location and scale parameters of the generalized exponential distributions based on order statistics. *J. Statist. Comput. Simul.* 69, 109- 123, 2001.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

The Coral Reefs Optimization Algorithm: An Efficient Meta-heuristic for Solving Hard Optimization Problems

S. Salcedo-Sanz[1], J. Del Ser[2], I. Landa-Torres[2], S. Gil-López[2], and A. Portilla-Figueras[1]

[1] Department of Signal Processing and Communications, Universidad de Alcalá, 28805 Alcalá de Henares, Madrid, Spain.
(E-mail: sancho.salcedo@uah.es)
[2] Tecnalia Research & Innovation. Parque Tecnológico de Bizkaia, 48170 Zamudio, Bizkaia, Spain.

Abstract. This paper presents a novel bio-inspired algorithm to tackle complex optimization problems: the Coral Reefs Optimization (CRO) algorithm. The CRO algorithm artificially simulates a coral reef, where different corals (namely, solutions to the optimization problem considered) grow and reproduce in coral colonies, fighting by choking out other corals for space in the reef. This fight for space, along with the specific characteristics of the corals' reproduction, produces a robust meta-heuristic algorithm, shown to be powerful for solving hard optimization problems. In this research the CRO algorithm is detailed and tested in several continuous and discrete optimization problems, obtaining advantages over other existing meta-heuristic techniques. The obtained results confirm the excellent performance of the proposed algorithm.
Keywords: Coral Reefs Optimization algorithm, optimization problems, modern meta-heuristics, bio-inspired algorithms.

1 Introduction

In the last years, huge research efforts have been conducted towards solving hard optimization problems, by well balancing the trade-off between the complexity incurred by the utilized method and the optimality of the produced solutions. These problems, often characterized by search spaces of high dimensionality (either discrete or continuous), non-linear objective functions and/or stringent constraints, arise frequently in Science and Engineering applications. In such fields, classical optimization approaches do not provide in general good solutions to these problems, or are just not applicable, due to the unmanageable search space structure or its huge size.

In this context, modern optimization heuristics and meta-heuristics have been lately the core of research, aimed at solving the aforementioned lack of efficient methods. A good number of such algorithms are *bio-inspired* techniques such as evolutionary algorithms (EA), which includes a whole family of techniques such as Genetic Algorithms [1], Evolutionary Strategies [2], Evolutionary Programming [3], Differential Evolution [4], among others. These schemes are based on concepts borrow from natural evolution and survival of the fittest individuals in Nature. Likewise, Ant Colonies Optimization (ACO) [5] are based

on the social behavior of ants, Particle Swarm Optimization (PSO) approaches [6] are in essence elegant algorithms specially well-suited for continuous optimization problems. They imitate the behavior of birds flocks or fish schools. There have been more research activity on bio-inspired meta-heuristics, with approaches such as Artificial Bee Colony [7], the Invasive Weed Optimization Algorithm (IWO), [8], based on weed growth and their invasive properties, or the so-called Cuckoo search approach [9], built upon the reproduction and breeding of the cuckoo bird, among others.

In this paper we present a novel bio-inspired meta-heuristic for optimization problems, which will be hereafter coined as the Coral Reefs Optimization (CRO) algorithm. The CRO algorithm is based on an artificial simulation of the process of coral reefs' formation and reproduction. During this process, the CRO algorithm emulates different phases of coral reproduction and fight for space in the reef, which ultimately renders an efficient algorithm for solving difficult optimization problems. The proposed CRO approach can be regarded as a cellular-type evolutionary scheme, with superior exploration-exploitation properties thanks to the particularities of the emulated reef structure and coral reproduction. The performance of the proposed approach has been tested in different benchmark problems obtaining very good results in comparison with alternative approaches in the literature.

The rest of this article is structured as follows: for the sake of self-completeness of the manuscript, the next section provides an introduction to coral reefs and corals' structure and reproduction. Next, Section 2 presents the CRO algorithm in detail, including an analysis of similarities and differences with other existing meta-heuristics. Section 3 shows the performance of the CRO algorithm in different optimization problems. Finally, Section 4 ends the paper by giving some concluding remarks.

2 The Coral Reefs Optimization algorithm

The CRO is a novel meta-heuristic approach based on corals' reproduction and coral reefs formation. Basically, the CRO is based on the artificial modeling of a coral reef, Λ, consisting of a $N \times M$ square grid. We assume that each square (i, j) of Λ is able to allocate a coral (or colony of corals) $\Xi_{i,j}$, representing a solution to a given optimization problem, which is encoded as a string of numbers in a given alphabet $\mathcal{I}$. The CRO algorithm is first initialized at random by assigning some squares in Λ to be occupied by corals (i.e. solutions to the problem) and some other squares in the grid to be empty, i.e. holes in the reef where new corals can freely settle and grow in the future. The rate between free/occupied squares in Λ at the beginning of the algorithm is an important parameter of the CRO algorithm, which is denoted as rho, and note that $0 < \rho_0 < 1$. Each coral is labeled with an associated *health* function $f(\Xi_{ij}) : \mathcal{I} \to \mathbb{R}$, that represents the problem's objective function. The CRO is based on the fact that reef will progress, as long as healthier (stronger) corals (which represent better solutions to the problem at hand) survive, while less healthy corals perish.

After the reef initialization described above, a second phase of reef formation is artificially simulated in the CRO algorithm: a simulation of the corals' reproduction in the reef is done by sequentially applying different operators. This sequential set of operators is then applied until a given stop criteria is met. Several operators to imitate corals' reproduction are defined, among them: a modeling of corals' sexual reproduction (broadcast spawning and brooding), a model of asexual reproduction (budding), and also some catastrophic events in the reef, i.e. polyps depredation. After the sexual and asexual reproduction, the set of larvae formed (new solutions to the problem), try to locate a place to grow in the reef. It could be in a free space, or in an occupied once, by fighting against the coral actually located in that place. If larvae are not successful in locate a place to grow in a given number of attempts, they are depredated in this phase. This second phase of the CRO can be detailed as follows:

1. *Broadcast Spawning (external sexual reproduction)*: the modeling of coral reproduction by *broadcast spawning* consists of the following steps:

   1.a. In a given step k of the reef formation phase, select uniformly at random a fraction of the existing corals ρ_k in the reef to be broadcast spawners. The fraction of broadcast spawners with respect to the overall amount of existing corals in the reef will be denoted as F_b. Corals that are not selected to be broadcast spawners (i.e. $1 - F_b$) will reproduce by brooding later on, in the algorithm.

   1.b. Select couples out of the pool of broadcast spawner corals in step k. Each of such couples will form a coral larva by sexual crossover, which is then released out to the water. Note that, once two corals have been selected to be the parents of a larva, they are not chosen anymore in step k (i.e. two corals are parents only once in a given step). These couple selection can be done uniformly at random or by resorting to any fitness proportionate selection approach (e.g. roulette wheel).

2. *Brooding (internal sexual reproduction)*: as previously mentioned, at each step k of the reef formation phase in the CRO algorithm, the fraction of corals that will reproduce by brooding is $1 - F_b$. The brooding modeling consists of the formation of a coral larva by means of a random mutation of the brooding-reproductive coral (self-fertilization considering hermaphrodite corals). The produced larva is then released out to the water in a similar fashion than that of the larvae generated in step 1.b.

3. *Larvae setting*: once all the larvae are formed at step k either through broadcast spawning (1.) or by brooding (2.), they will try to set and grow in the reef. First, the health function of each coral larva is computed. Second, each larva will randomly try to set in a square (i, j) of the reef. If the square is empty (free space in the reef), the coral grows therein no matter the value of its health function. By contrast, if a coral is already occupying the square at hand, the new larva will set only if its health function is better than that of the existing coral. We define a number κ of

attempts for a larva to set in the reef: after κ unsuccessful tries, it will be depredated by animals in the reef.

4. *Asexual reproduction*: in the modeling of asexual reproduction (budding or fragmentation), the overall set of existing corals in the reef are sorted as a function of their level of healthiness (given by $f(\Xi_{ij})$), from which a fraction F_a duplicates itself and tries to settle in a different part of the reef by following the setting process described in Step 3. Note that a maximum number of identical corals (μ) are allowed in the reef.

5. *Depredation in polyp phase*: corals may die during the reef formation phase of the CRO algorithm. At the end of each reproduction step k, a small number of corals in the reef can be depredated, thus liberating space in the reef for next coral generation. The depredation operator is applied with a very small probability P_d at each step k, and exclusively to a fraction F_d of the worse health corals in Λ. For the sake of simplicity in the parameter setting of the CRO algorithm, the value of this fraction may be set to $F_d = F_a$. Any other assignment may also apply provided that $F_d + F_a \leq 1$ (i.e. no overlap between the asexually reproduced and the depredated coral sets).

3 Experiments and Numerical Results

In this paper we carry out a first performance assessment of the proposed CRO algorithm in different test problems. Specifically, different well-known continuous and discrete benchmark problems are under consideration: continuous analytical functions and several instances of the *Max-Ones* and *3-bit Deceptive*.

We have selected other meta-heuristic algorithms for comparison: Evolutionary Algorithms, Genetic Algorithms (EA and GA, [1]) and Harmony Search (HS, [10]), which have obtained excellent results in a wide range of optimization problems during the last years. Regarding the continuous benchmark problems, we have compared the results obtained by the CRO in the same problems tackled in [11].

Following this rationale, the encoding strategy used to represent the produced solutions for the aforementioned problems is set identical for all the algorithms under comparison. Specifically, real encoding has been adopted for the continuous benchmark problems, whereas the *Max Ones* and *3-bit Deceptive* problems resort to standard binary encoding. On the other hand, values of all parameters controlling the CRO approach have been set to be comparable to that of its counterparts tested in every benchmark function. Therefore we have kept the number of function evaluations constant for all the compared algorithms in *Maxones* (15000), whereas for the *3-bit Deceptive* problem the total number of function evaluations is set to 50000 for GA and HS, and 30000 for the proposed CRO. In the continuous benchmark functions, we have set the number of function evaluations to be comparable with the results in [11]. For every simulation instance, 30 executions of each algorithm have been launched so as to obtain well-sampled performance statistics (best, average and standard

deviation of the metric after all iterations are done). Note that the size of the population – $N \times M$ for CRO, population length L for the GA, and harmony memory size HM for HS – have been set equal for all the experiments for the sake of fairness in the comparison of the algorithms: in the *Max Ones* problem $N \times M = 5 \times 10$, $L = 50$ and $HM = 50$ and in the *3-bit Deceptive* problem $N \times M = 10 \times 10$, $L = 100$ and $HM = 100$. The CRO parameters F_b and ρ has been set to $F_b = 0.9$ and $\rho = 0.7$, unless otherwise stated in the discussion on the specific simulated application.

3.1 CRO Evaluation in Continuous Benchmark Problems

This first round of experiments includes four well-known benchmark functions, on which the proposed CRO is comparatively assessed with respect to different hybrid evolutionary algorithms described in [11]. In these experiments we have incorporated Gaussian and Cauchy mutations [3] in the internal reproduction (brooding) of the corals in order to accommodate the corresponding operator to the real encoding of the solutions. In the Gaussian mutation we have established a fixed standard deviation $\sigma = (max - min)/100$, where max and min are the maximum and minimum values that each component of the solution can take, whereas in the Cauchy mutation the value of the τ parameter has been fixed to 1 following the guidelines in [3]. The rest of operators in the CRO are the ones shown in Section 2.

Table 1 lists the results obtained by three different versions of the CRO (with Gaussian, Cauchy and Gaussian-Cauchy internal reproduction) in the benchmark functions tackled in this first round of experiments. Also included are the results for different versions of the hybrid evolutionary algorithm proposed in [11], labelled as Hybrid Adaptive Evolutionary Algorithm (HAEA) in what follows. It is straightforward to note that the CRO approach is able to obtain better results than the different versions of HAEA consistently – and with statistical significance positively checked through Kruskal-Wallis tests – in all the functions under consideration. The inclusion of both Gaussian and Cauchy mutations in the brooding coral reproduction (always maintaining the number of functions evaluations) appears to improve the performance of the CRO solver.

3.2 CRO Evaluation in Discrete Benchmark Problems

The first discrete benchmark problem considered is the well-known *Max Ones* problem, often used in a number of previous works aimed at evaluating different approaches of genetic algorithms (e.g. see [**?**,11] and references therein). This optimization problem is defined in a binary search space $\mathcal{S} = \{0, 1\}^n$, where n stands for the dimension of the space. The *One Max* problem is then defined as

$$\max_{\mathbf{x} \in \mathcal{S}} f(\mathbf{x}) = \frac{100}{n} \sum_{i=1}^{n} x_i \qquad [\%]. \tag{1}$$

Table 1. Results (mean/standard deviation) obtained in the different continuous benchmark functions tested.

Algorithm	Rosenbrock	Schwefel	Rastrigin	Griewank
CRO (G)	$7.0 \cdot 10^{-5}/5.0 \cdot 10^{-6}$	$1.3 \cdot 10^{-4}/2.0 \cdot 10^{-6}$	$7.1 \cdot 10^{-3}/3.0 \cdot 10^{-3}$	0.22/0.05
CRO (C)	$7.2 \cdot 10^{-5}/5.6 \cdot 10^{-6}$	$1.3 \cdot 10^{-4}/1.6 \cdot 10^{-6}$	$6.7 \cdot 10^{-3}/2.3 \cdot 10^{-3}$	0.03/0.02
CRO (G+C)	$2.3 \cdot 10^{-5}/1.0 \cdot 10^{-6}$	$1.3 \cdot 10^{-4}/1.4 \cdot 10^{-6}$	$4.3 \cdot 10^{-3}/1.5 \cdot 10^{-3}$	0.05/0.02
HAEA (XUG)	$7.0 \cdot 10^{-4}/1.0 \cdot 10^{-5}$	$5.6 \cdot 10^{-3}/0.01$	0.05/0.02	0.055/0.03
HAEA (XU)	$4.1 \cdot 10^{-3}/4.0 \cdot 10^{-3}$	1.3/0.93	0.24/0.15	0.5/0.2
HAEA (XG)	$1.3 \cdot 10^{-3}/3.6 \cdot 10^{-3}$	140.5/123.7	7.7/3.2	0.05/0.02
HAEA (GU)	$1.4 \cdot 10^{-4}/2.5 \cdot 10^{-3}$	201.9/81.2	6.3/1.4	1.6/0.38

Despite the evident simplicity of its definition, this problem is challenging for optimization algorithms when dealing with large values of the space dimensionality n.

Table 2 summarizes the results (maximum, average and standard deviation) obtained by CRO, GA and HS in *Max Ones* instances of varying size from $n = 50$ to $n = 500$. As one may expect, in the scenarios of smallest dimension all the utilized heuristic approaches are able to obtain the optimum solution (100%) in every run of the algorithm. However, when the dimensions of the simulated problem increase, the differences between the CRO and the other tested algorithms become more significant. Specially remarkable is the fact that the CRO obtains the best value in all the instances with a very high probability (over 99% of the times in which the algorithm was run). HS also obtains good solutions, but notably worse than the GA even in the smallest instances. By contrast, the CRO clearly dominates GA and HS, specially in the largest *Max Ones* instances.

Table 2. Results obtained by CRO, GA and HS in *Max Ones* problems of increasing size. The results are shown in best/average/standard deviation over 30 runs of the algorithms.

n	CRO	GA	HS
50	100/100/0	100/100/0	100/100/0
100	100/100/0	100/100/0	98/95.67/0.92
150	100/100/0	100/100/0	94.67/90.84/1.13
200	$100/99.98/9.12 \cdot 10^{-4}$	100/99.93/0.17	90/87.32/0.88
250	$100/99.97/7.3 \cdot 10^{-4}$	100/99.81/0.25	86.80/84.64/1.04
300	$100/99.96\ /8.45 \cdot 10^{-4}$	100/99.61/0.39	83.67/82.0700/0.62
350	$100/99.96/9.8 \cdot 10^{-4}$	100/99.21/0.46	81.4300/80.03/0.71
400	$100/99.95/7.3 \cdot 10^{-4}$	99.50/98.67/0.58	79.50/78.45/1.04
450	100/99.93/0.13	99.55/98.11/0.67	78.67/76.97/0.99
500	100/99.92/0.1	98.60/97.04/0.75	78/75.99/0.69

The second discrete benchmark problem addressed is the maximization of the aforementioned *3-bit Deceptive* function, which has been previously utilized to evaluate different improvements in genetic and evolutionary heuristics [11]. The 3-bit deceptive function is a binary optimization problem defined in blocks of 3 bits. Each 3-bit block is assigned a value according to Table 3. The optimization of the function is known to be computationally hard for heuristic algorithms, since the `111` block (optimum since it is assigned the highest value) is "surrounded" by low-valued blocks of two 1s (i.e. with small Hamming distance with respect to `111`). Different size functions (integer multiple of 3) are considered in this study, i.e. $n = \{15, 30, 45, 60, 75, 90, 105, 120\}$.

Table 4 shows the results obtained by the CRO in the considered 3-bit Deceptive functions, and its comparison to those of HS and GA. In this problem the CRO clearly obtains the best results among all the compared algorithms. Indeed, it is able to obtain the optimum (maximum) value in all the instances and in almost every executed run. The performance of the alternative algorithms degrades significantly in the largest instances, though in the smallest ones the GA is able to obtain the optimum value.

Table 3. Value assignment in the considered *3-bits Deceptive* function.

Groups of 3 bits	Value
1 1 1	80
0 0 0	70
0 0 1	50
0 1 0	49
1 0 0	30
1 1 0	3
1 0 1	2
0 1 1	1

Table 4. Results obtained by CRO, HS and GA in the considered *3-bit Deceptive* instances. The results are shown in best/average/standard deviation over 30 runs of the algorithms.

n	CRO	HS	GA	Upper Bound
15	400/400/0	400/399.66/1.82	400/400/0	400
30	800/800/0	800/792/8.05	800/795/6.82	800
45	1200/1200/0	1190/1159/14.93	1200/1179.3/13.37	1200
60	1600/1600/0	1560/1517.3/21.96	1590/1562.70/18.74	1600
75	2000/2000/0	1910/1882/20.97	1990/1940.30/22.04	2000
90	2400/2400/0	2280/2243/22.63	2340/2297.30/21.16	2400
105	2800/2799.70/1.82	2660/2598.80/34.20	2730/2687.70/27.75	2800
120	3200/3200/0	2990/2924.8/37.90	3090/3049/21.22	3200

4 Conclusions

In this paper we have presented a novel algorithm to solve optimization problems, inspired by the process of coral reefs formation, and guided by coral reproduction, reef expansion and fight for the space in the reef. The algorithm, named as the Coral Reef Optimization (CRO) algorithm, is a kind of cellular evolutionary algorithm rendering very good properties of convergence to global optima. In this paper we have studied the main characteristics of the proposed CRO and analyzed its comparison to other existing meta-heuristic approaches in different benchmark problems. The promising obtained results encourage the application of the proposed CRO approach to other practical optimization paradigms of high complexity.

Acknowledgements

This work has been partially supported by Spanish Ministry of Science and Innovation, under project number ECO2010-22065-C03-02.

References

1.A. E. Eiben and J. E. Smith, "Introduction to evolutionary computing," Springer-Verlag, Natural Computing Series 1st edition, 2003.
2.H. G. Beyer and H. P. Schwefel, "Evolution strategies - A comprehensive introduction," *Natural Computing*, vol. 1, no. 1, pp. 3-52, 2002.
3.X. Yao, Y. Liu and G. Lin, "Evolutionary Programming made faster," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 2, pp. 82-102, 1999.
4.R. Storn and K. Price, "Differential Evolution - A simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, pp. 341-359, 1997.
5.M. Dorigo, V. Maziezzo and A. Colorni, "The ant system: optimization by a colony of cooperating ants," *IEEE Transactions on Systems, Man and Cybernetics B*, vol. 26, no. 1, pp. 29-41, 1996.
6.J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. of the 4th IEEE International Conference on Neural Networks*, pp. 1942-1948, 1995.
7.D. Karaboga and B. Basturk, "On the performance of the artificial bee colony (ABC) algorithm," *Applied Soft Computing*, vol. 8, pp. 687-697, 2008.
8.A. R. Mehrabian and C. Lucas, "A novel numerical optimization algorithm inspired from weed colonization," *Ecological Informatics*, vol. 1, pp. 355-366, 2006.
9.X. S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *Proc. of the World Conference on Nature & Biologically Inspired Computing*, pp. 210-214, 2009.
10.Z. W. Geem, J. Hoon Kim and G. V. Loganathan, A New Heuristic Optimization Algorithm: Harmony Search, *Simulation*, vol. 76, no. 2, 2001. 60-68.
11.J. Gómez, "Self Adaptation of operator rates in evolutionary algorithms," GECCO 2004, *Lecture Notes in Computer Science*, vol. 3102, pp. 1162-1173, 2004.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Estimating the Density and Hazard Rate Functions Using the Reciprocal Inverse Gaussian Kernel

Raid B. Salha
Department of Mathematics, The Islamic University, Gaza, Palestine
rbsalha@iugaza.edu.ps

Abstract. In this paper, we use the Reciprocal Inverse Gaussian (RIG) kernel to estimate nonparametrically the probability density function (pdf) and the hazard rate function for independent and identically distributed (iid) data. The estimator uses adaptive weights depending on the points at which we estimate the functions. We derive the strong consistency, the asymptotic normality and the asymptotic mean squared error (AMSE) of the proposed estimator. Also, the selection of the optimal bandwidth is investigated. The performance of the proposed estimator is compared to that of the Gaussian kernel.

Keywords: Reciprocal Inverse Gaussian kernel, hazard rate function, kernel estimation, asymptotic mean squared error, boundary bias.

2000 MSC: 62G07, 62G08

1 Introduction

Estimators of the hazard rate function based on kernel estimation have been studied extensively in literature. For example, see Watson and Leadbetter [7], Rice and Rosenblatt [2] and Salha [4, 5]. However, when the support of the curve under estimation is bounded, many nonparametric estimators appear to be biased more than the usual in regions near the endpoints. Boundary bias is due to weight allocation by the fixed symmetric kernel outside the density support when smoothing is carried out near the boundary. To solve this problem, boundary kernels are used only within the boundary region. This is an efficient way to correct boundary bias but it requires complicated adjustments to the estimator. To solve this problem, Chen [1] has replaced the symmetric kernels by asymmetric Gamma kernel which never assigns weight outside the support.

In Salha [4], the estimation of the hazard rate function using the Inverse Gaussian (IG) kernel has been considered. In this paper, we consider the RIG kernel estimation of the hazard rate function. As Gamma kernel estimator, the RIG kernel estimator is free of boundary bias, always non-negative and achieves the optimal rate of convergence for the mean integrated squared error (MISE) within the class of nonnegative kernel density estimators, see Scaillet [3].

This paper is organized as follows. In Section 2, some basic definitions and conditions are stated. In Section 3, the main results of this paper are stated and proved. The AMSE of the proposed estimator and the selection of the optimal bandwidth are investigated in Section 4. In Section 5, the performance of the proposed estimator is tested and compared to that of the Gaussian kernel estimator. Section 6, contains some concluding remarks.

2 Preliminaries

In this section, we state the conditions under which the results of this paper will be proved. Also, we introduce some basic definitions.

Conditions

1. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with an unknown probability density function f defined on $[0,\infty)$, such that f is twice continuously differentiable, and $\int_0^\infty \left(x^3 f''(x)\right)^2 dx < \infty$.
2. h is a smoothing parameter satisfying $h + \frac{1}{nh} \to 0$, and $nh^{\frac{5}{2}} \to 0$, as $n \to \infty$.

Definition 1. Scaillet [3] defined the RIG kernel estimator of the pdf $f(\cdot)$,

$$\hat{f}_{RIG}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{RIG}\left(\frac{1}{x-h}, \frac{1}{h}\right)(X_i), \text{ where} \quad (1)$$

$$K_{RIG}\left(\frac{1}{x-h}, \frac{1}{h}\right)(u) = \frac{1}{\sqrt{2\pi h\, u}} \exp\left(-\frac{x-h}{2h}\left(\frac{u}{x-h} - 2 + \frac{x-h}{u}\right)\right). \quad (2)$$

Definition 2. Let X be a random variable with pdf $f(x)$ and cdf $F(x)$, the hazard rate function $r(x)$ of X is defined as

$$r(x) = \lim_{\Delta x \to 0} \frac{P(X \le x + \Delta x \mid X > x)}{\Delta x} = \frac{f(x)}{S(x)}, \; x > 0, \text{ where}$$

$S(\cdot) = 1 - F(\cdot)$ is called the survivor function.

Definition 3. The proposed kernel estimator for the hazard rate function is given by $\hat{r}_{RIG}(x) = \frac{\hat{f}_{RIG}(x)}{\hat{S}_{RIG}(x)}$, where $\hat{S}_{RIG}(x) = 1 - \int_0^x \hat{f}_{RIG}(u)\,du$.

Lemma 1. Under the conditions (1) and (2), the following hold

(i) $Bias\left(\hat{f}_{RIG}(x)\right) = \frac{1}{2} x f''(x) h + o(h)$.

(ii) $Var\left(\hat{f}_{RIG}(x)\right) = \frac{1}{2n\sqrt{\pi h}} x^{-\frac{1}{2}} f(x) + o(n^{-1}h^{-\frac{1}{2}})$.

Proof. See Proposition 1 and 2 in Scaillet [3].

3. Main Results

The asymptotic normality of the RlG kernel estimator of the pdf is given in Theorem 1.

Theorem 1. Under the conditions (1) and (2), the following holds

$$\sqrt{nh^{\frac{1}{2}}}(\hat{f}_{RIG}(x)-f(x))\xrightarrow{d}N\left(0,\frac{1}{2\sqrt{\pi}}x^{-\frac{1}{2}}f(x)\right).$$

Proof. Let $V_{ni}=K_{IG}(x,\frac{1}{h})(X_i)$, $i=1,2,...,n$, then $\hat{f}(x)=\frac{1}{n}\sum_{i=1}^{n}V_{ni}$.
Now, we show that Liapounov condition is satisfied, that is for some $\delta>0$,

$$\lim_{n\to\infty}\frac{E\,|V_n-E(V_n)|^{2+\delta}}{n^{\frac{\delta}{2}}\sigma^{2+\delta}(V_n)}=0.$$

Let η_x be a $RIG(\frac{1}{x-h},\frac{2+\delta}{h})$ distributed random variable. Hence

$$\mu_x=E(\eta_x)=x-h+\frac{h}{2+\delta}\text{ and }T_x=Var(\eta_x)=\frac{(x-h)h}{2+\delta}+\frac{2h^2}{(2+\delta)^2}.$$

$$E\,|V_n|^{2+\delta}=E\left[\left(\frac{1}{\sqrt{2\pi hy}}\right)^{2+\delta}\exp\left(-\frac{(2+\delta)(x-h)}{2h}\right)\left(\frac{y}{x-h}-2+\frac{x-h}{y}\right)\right]$$

$$=\frac{\sqrt{2\pi h}}{\sqrt{2+\delta}(2\pi h)^{1+\frac{\delta}{2}}}E\left(\eta_x^{-\frac{3}{2}(1+\delta)}f(\eta_x)\right).$$

By using the Taylor's series to expand $f(\eta_x)$ about μ_x, we obtain

$$E\left(\eta_x^{-\frac{3}{2}(1+\delta)}f(\eta_x)\right)=x^{-\frac{3}{2}(1+\delta)}f(x)+\frac{1}{2}(x^{-\frac{3}{2}(1+\delta)}f''(x)-\frac{3}{2}x^{-\frac{5}{2}(1+\delta)}f'(x)$$

$$-\frac{3}{2}x^{-\frac{5}{2}(1+\delta)}f'(x)+\frac{15}{4}x^{-\frac{7}{2}(1+\delta)}f(x))x^3h+o(h)$$

$$=x^{-\frac{3}{2}(1+\delta)}f(x)+o(h).$$

This implies that

$$\mathrm{E}|\mathrm{V_n}|^{2+\delta} = \frac{1}{\sqrt{2+\delta}\,(2\pi \mathrm{h})^{\frac{1+\delta}{2}}}\, x^{-\frac{3}{2}(1+\delta)} f(x) + o\left(h^{-\frac{(1+\delta)}{2}}\right) \to 0.$$

Lemma 2. Under the conditions (1) and (2), the following holds

$$\sqrt{nh^{\frac{1}{2}}}\,|\hat{F}(x) - F(x)| \xrightarrow{p} 0.$$

Proof. From the definition of $\hat{F}(x)$, the following Relations (3) and (4) hold.

$$E\hat{F}(x) = \int_0^{\infty}\int_0^{x} K_{RIG}\left(\frac{1}{u-h}, \frac{1}{h}\right)(y)\,du\, f(y)\,dy = \int_0^{x} E(f(\xi_u))\,du$$

$$= \int_0^{x} \left(f(u) + \frac{1}{2} u f''(u)\, h\right) du + o(h) = F(x) + o(h).$$

This implies that,

$$\sqrt{nh^{\frac{1}{2}}}\,|E\hat{F}(x) - F(x)| = o\left((nh^{\frac{5}{2}})^{\frac{1}{2}}\right) \to 0. \qquad (3)$$

Now, $\hat{F}(x)$ can be written in the following form

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{x} K_{RIG}\left(\frac{1}{u-h}, \frac{1}{h}\right)(X_i)\,du = \frac{1}{n}\sum_{i=1}^{n} W_i(x).$$

Let $\varepsilon > 0, \delta > 0$ be given.

$$P\left[(nh^{\frac{1}{2}})^{\frac{1}{2}}\,|\hat{F}(x) - EF(x)| > \varepsilon\right] \le \varepsilon^{-2-2\delta}(nh^{\frac{1}{2}})^{1+\delta} E\,\left|\frac{1}{n}\sum_{i=1}^{n}\left[W_i(x) - EW_i(x)\right]\right|^{2+\delta}$$

$$= \varepsilon^{-2-2\delta} h^{\frac{1+\delta}{2}} n^{-1-\delta} E\,\left|\sum_{i=1}^{n}\left[W_i(x) - EW_i(x)\right]\right|^{2+\delta} \le 2^{1+\delta}\varepsilon^{-2-2\delta}(n^{-1}h^{\frac{1}{2}})^{1+\delta}\sum_{i=1}^{n} E\,|W_i(x)|^{2+2\delta}$$

$$+2^{1+\delta}\varepsilon^{-2-2\delta}(n^{-1}h^{\frac{1}{2}})^{1+\delta}\sum_{i=1}^{n}|EW_i(x)|^{2+2\delta} \to 0.$$

This implies that,

$$\sqrt{nh^{\frac{1}{2}}}\,|\hat{F}(x) - EF(x)| \xrightarrow{p} 0. \qquad (4)$$

Now, using Relations (3), (4) and the following fact,

$|\hat{F}(x) - F(x)| \le |\hat{F}(x) - E\hat{F}(x)| + |E\hat{F}(x) - F(x)|$, we obtain that

$$\sqrt{nh^{\frac{1}{2}}}\,|\hat{F}(x) - F(x)| \le \sqrt{nh^{\frac{1}{2}}}\,|\hat{F}(x) - E\hat{F}(x)| + \sqrt{nh^{\frac{1}{2}}}\,|E\hat{F}(x) - F(x)| \xrightarrow{p} 0.$$

This completes the proof of the lemma. □

Now, the asymptotic normality of the proposed etimator is given in Theorem 2.

Theorem 2. Under the conditions (1) and (2), the following holds

$$\sqrt{nh^{\frac{1}{2}}}\left(\hat{r}(x)-r(x)\right)\xrightarrow{d}N\left(0,\frac{1}{2\sqrt{\pi}}x^{-\frac{1}{2}}\frac{r(x)}{S(x)}\right).$$

Proof.

$$\begin{aligned}\sqrt{nh^{\frac{1}{2}}}\left(\hat{r}(x)-r(x)\right)&=\sqrt{nh^{\frac{1}{2}}}\left(\frac{\hat{f}(x)}{\hat{S}(x)}-\frac{f(x)}{S(x)}\right)\\&=\sqrt{nh^{\frac{1}{2}}}\left(\frac{\hat{f}(x)}{\hat{S}(x)}-\frac{f(x)}{\hat{S}(x)}-\frac{f(x)}{S(x)}+\frac{f(x)}{\hat{S}(x)}\right)\\&=\frac{\sqrt{n\,h^{\frac{1}{2}}}}{\hat{S}(x)}\left[\hat{f}(x)-f(x)\right]+\frac{\sqrt{n\,h^{\frac{1}{2}}}f(x)}{S(x)\hat{S}(x)}\left[\hat{S}(x)-S(x)\right].\end{aligned}\quad(5)$$

The proof is completed by a combination of Theorem 1, Lemma 2 and Equation (5). Since by Theorem 1, the first term in Equation (5) is asymptotically normally distributed and the second term vanishes by Lemma 2. □

From Theorem 1 and 2, we get that

$$E\left(\hat{r}(x)\right)=\frac{E\left(\hat{f}(x)\right)}{E\left(\hat{S}(x)\right)}=\frac{f(x)+\frac{1}{2}xf''(x)h}{S(x)}+o(h)=r(x)+\frac{\frac{1}{2}xf''(x)h}{S(x)}+o(h).$$

This implies that

$$Bias\left(\hat{r}(x)\right)=\frac{\frac{1}{2}xf''(x)h}{S(x)}+o(h)\text{ and }Var\left(\hat{r}(x)\right)=\frac{1}{2n\sqrt{\pi h}}x^{-\frac{1}{2}}\frac{r(x)}{S(x)}+o(n^{-1}h^{-\frac{1}{2}}).$$

4 Bandwidth Selection

The selection of the bandwidth in kernel estimation plays an important role. It depends on choosing a value of the bandwidth that minimizes the AMSE. Using the same techniques of Scaillet [3], the AMSE is given by

$$\text{AMSE} = \left(\frac{\frac{1}{2} x f''(x) h}{S(x)} \right)^2 + \frac{1}{2n\sqrt{\pi h}} x^{-\frac{1}{2}} \frac{r(x)}{S(x)}. \qquad (6)$$

Differentiate the AMSE with respect to h, the n equating it to zero, we obtain

$$2 \left(\frac{\frac{1}{2} x f''(x)}{S(x)} \right)^2 h - \frac{1}{2n\sqrt{\pi}} x^{-\frac{1}{2}} \frac{r(x)}{S(x)} \frac{1}{2} h^{-\frac{3}{2}} = 0. \qquad (7)$$

Multiplying Equation (7) by $h^{\frac{3}{2}}$, and solving for h, we obtain

$$h = \left(\frac{1}{2\sqrt{\pi}} \frac{r(x)}{(f''(x))^2} \right)^{\frac{2}{5}} x^{-\frac{3}{5}} n^{-\frac{2}{5}}. \qquad (8)$$

5 Applications

In this section, the performance of the proposed estimator in estimating the pdf and hazard rate function is tested upon two applications. For comparison purposes we also estimate the two functions using the Gaussian kernel estimator. For the practical implementation of the RIG estimator, we used the bandwidth selection procedure described in Section 4 and for the Gaussian estimator, we used Equation (3.28) in Silverman [6].

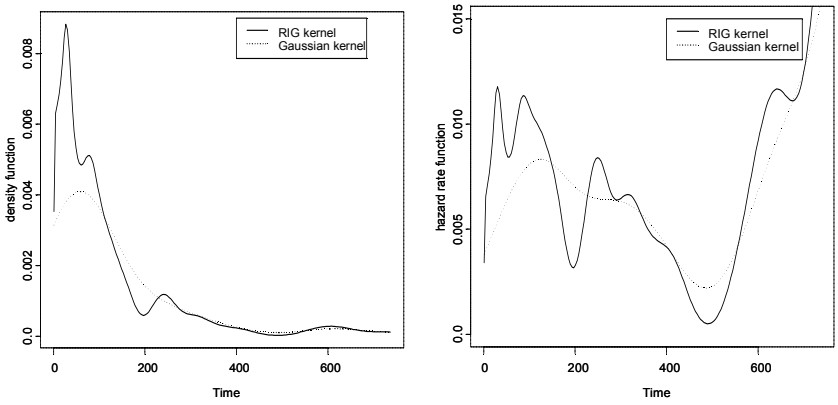
5.1 Real Data

In this subsection, we use the suicide data given in Silverman [6], to exhibit the practical performance of the RIG estimator. The data gives the lengths of the treatment spells (in days) of control patients in suicide study. Figures 1(a) and 1(b) show the two estimators of the probability density and hazard rate functions, respectively. Although the suggested values of the density and hazard rate functions from the two estimators are different, they both suggest a similar structure for the two estimated functions. As we see, the divergence of the two estimators gets large at the boundary near the zero and becomes smaller in the interior especially from approximately $t \geq 250$.

5.2 A Simulation Study

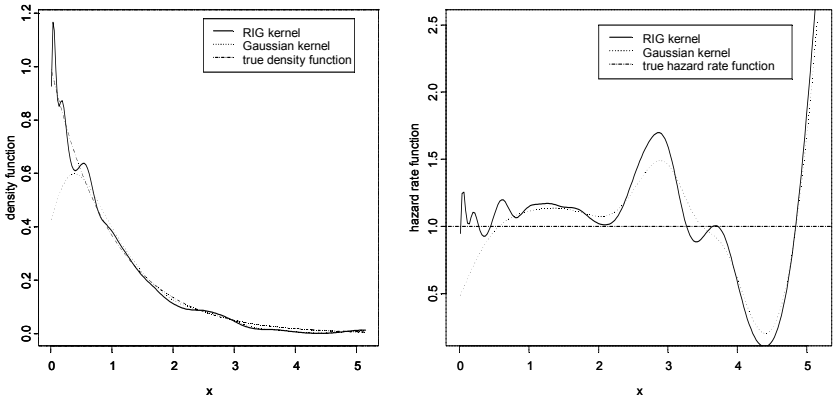
A sample of size 200 from the exponential distribution with pdf $f(x) = e^{-x}$ is simulated. After that the density function and the hazard rate functions were estimated using the RIG and the Gaussian estimators. The estimated values and

the true functions are plotted in Figures 2(a) and 2(b), respectively. The two figures show that the performance of the RIG estimator is better than that of the Gaussian estimator at the boundary near the zero. In the interior the behavior of the two estimators becomes more similar as we get away from the zero.



(a) (b)

Fig.1: The RIG and Gaussian kernel estimators of (a) the density function (b) the hazard rate function for the suicide data.



(a) (b)

Fig. 2: The RIG and Gaussian kernel estimators of (a) the density function (b) the hazard rate function for the simulated data of the exponential distribution

6. Comments and Conclusion

In this paper, we have proposed a new kernel estimator of the hazard rate function for (iid) data based on the RIG with nonnegative support which was proposed by Scaillet [3]. The proposed estimator overcomes the bias problem when the hazard rate function is estimated at the boundary region near the zero.

The asymptotic normality, the strong consistency and the AMSE of the proposed estimator were obtained. The AMSE of the new estimator is smaller than that of the Gaussian kernel near the zero.

Two applications show that the performance of the proposed estimator is better than that of the Gaussian kernel estimator at the boundary region near the zero. This is due to weight allocation by the Gaussian kernel outside the density support when smoothing is carried out at the boundary near the zero.

The new estimator can be modified by considering a new bandwidth selection technique that uses a variable bandwidth that depends on the points at which the hazard rate function is estimated rather than a constant bandwidth.

References

1. Chen, S. X. (2000). Probability density function estimation using gamma kernels. Ann. Inst. Stat. Math., 52, 471- 480.

2. J. Rice and M. Rosenblatt (1976). Estimation of the log survivor function and hazard function. Sankhya Series A 38, 60-78.

3. O. Scaillet (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels. Nonparametric Statistics, 16, 217-226.

4. Salha, R. (2012). Hazard rate function estimation using inverse Gaussian kernel. The Islamic university of Gaza journal of Natural and Engineering studies, 20(1), 73-84.

5. Salha, R (2009). Adaptive kernel estimation of the hazard rate function. The Islamic university of Gaza Journal of Natural and Engineering studies,17(1), 71-81.

6. Silverman, B. W. (1986). Density estimation for statistics and data analysis. Chapman and Hall, London.

7. Watson, G. and Leadbetter, M. (1964). Hazard analysis I, Biometrika, 51, 175-184.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Incorporating the Stochastic Process Setup in Parameter Estimation

Lino Sant[1] and Mark Anthony Caruana[2]

[1] Department of Statistics and Operations Research, Faculty of Science, University of Malta, Msida, Malta
(E-mail: lino.sant@um.edu.mt)
[2] Department of Statistics and Operations Research, Faculty of Science, University of Malta, Msida, Malta
(E-mail: mark.caruana@um.edu.mt)

Abstract. Estimation problems within the context of stochastic processes are usually studied with the help of statistical asymptotic theory and proposed estimators are tested with the use of simulated data. For processes with stationary increments it is customary to use differenced time series, treating them as selections from the increments' distribution. Though distributionally correct, this approach throws away most information related to the stochastic process setup. In this paper we consider the above problems with reference to parameter estimation of a gamma process. Using the derived bridge processes we propose estimators whose properties we investigate in contrast to the gamma-increments MLE. The proposed estimators have a smaller bias, comparable variance and offers a look at the time-evolution of the parameter estimation. Empirical results are presented.
Keywords: Lévy processes, gamma process, bridge process, Dirichlet distribution.

1 Introduction

The estimation of stochastic models to fit to data obtained from real systems borrows a lot from statistical estimation theory, but arguably not enough from the theory of stochastic processes. In many papers dedicated to the estimation of Lévy processes, the estimation is understandably restricted to the estimation of either the parameters of the infinitely divisible distribution of the increments or else to the Lévy measure. Clearly both characterize completely the distributional framework of the corresponding processes. However there are other static and dynamical statistical properties which are of interest to look into in practical applications.
In this paper we study the gamma process with the aim of tackling estimation issues using stochastic properties other than the stationary gamma increments. In particular, we consider the problems of estimator bias and minimization of estimator variance.
The gamma process has been much studied because of its common use in climate and hydrology related modelling exercises Thom [10] and more recently also in finance Avramidis et al. [1].
Let us denote the gamma process by $(G_t)_{t\in\mathbb{R}_+}$. The overall distributional structure of this process is underpinned by the independently gamma distributed increments with parameters α, λ. For $G_{t_1}, G_{t_2} - G_{t_1}, \ldots, G_{t_n} - G_{t_n}$ we have

the joint density function:

$$f(x_1, x_2, \ldots, x_n) = \frac{x_1^{\alpha t_1 - 1} x_2^{\alpha(t_2 - t_1)-1} \ldots x_n^{\alpha(t_n - t_{n-1})-1} \exp\left(\frac{-(x_n - x_1)}{\lambda}\right)}{\Gamma(at_1)\Gamma(a(t_2 - t_1)) \ldots \Gamma(\alpha(t_n - t_{n-1})\lambda^{\alpha(t_n - t_1)})} \quad (1)$$

So letting $Z_i = G_{t_i} - G_{t_{i-1}}$ the form above is exploited in maximum likelihood estimation applied on the increments $Z_1, Z_2, \ldots, Z_n$ which are treated as independent gamma random variables. This approach blurs out the stochastic process context.
The maximum likelihood estimator is not just a venerable tool for estimation, but a superior one on many counts. Its more important virtues are of course asymptotic normality, consistency, asymptotic efficiency and functional invariance. Nevertheless it is also known to exhibit defects, one of which is biasedness in the case of the gamma distribution. We would like to tackle this problem through the use of other estimators which make use of intrinsic properties of the gamma process.

2 The gamma bridges and their derived estimators

The gamma bridge process $(G_{tT})_{0 \leq t \leq T}$ on the time interval $[0, T]$ is derived from the gamma process by :

$$G_{tT} = \frac{G_t}{G_T} \quad (2)$$

The gamma bridge has some nice independence properties as discussed by Emery and Yor [9]. We propose the following construction of bridges. Given two time points t_1 and t_n, the time interval in between is partitioned as: $\{t_1, t_2, \ldots, t_n\}$. Now for any two intermediate time points, $0 \leq t_i < t_j \leq t_n$ the random variables $(G_{t_j} - G_{t_i})/(G_{t_n} - G_{t_1})$ and $G_{t_n} - G_{t_1}$ are independent. Furthermore standard theory as in Ferguson [8] tells us that the following random vector

$$\left(\frac{G_{t_2} - G_{t_1}}{G_{t_n} - G_{t_1}}, \frac{G_{t_3} - G_{t_2}}{G_{t_n} - G_{t_1}}, \ldots, \frac{G_{t_n} - G_{t_{n-1}}}{G_{t_n} - G_{t_1}}\right) \quad (3)$$

conditioned on the values of the bridge end-point, has a Dirichlet distribution: $f(u_1, u_2, \ldots, u_n | G_{t_1}, G_{t_n}) =$

$$\frac{\Gamma[\alpha(t_n - t_1)] u_1^{\alpha(t_2 - t_1)-1} u_2^{\alpha(t_3 - t_2)-1} \ldots u_n^{\alpha(t_n - t_{n-1})-1}}{\Gamma[\alpha(t_2 - t_1)]\Gamma[\alpha(t_3 - t_2)] \ldots \Gamma[\alpha(t_n - t_{n-1})]} \quad (4)$$

Given integer k such that $mk + 1$ takes us to the closest value to n, we take the following bridge end-points $G_{t_1}, G_{t_{k+1}}, \ldots, G_{mk+1}$. The corresponding increments are denote by $X_{(j-1)k+i}$ which is the i^{th} increment within the j^{th} bridge. We then define m Dirichlet distributed independent random vectors

$$(U_1^{k,1}, \ldots, U_k^{k,1}), \ldots, (U_1^{k,m}, \ldots, U_k^{k,m}), \quad (5)$$

defined similarly as above by

$$U_i^{k,j} = X_{(j-1)k+i} / \left(\sum_{i=1}^{k} X_{(j-1)k+i} \right)$$

These k gamma bridges thus give us an intricate construction, woven over the span of time and built from the intrinsic structure of the gamma process. It provides us with new estimation tools which we proceed to describe.
We shall assume that time increments are equal in size, say δt, and to simplify notation we drop δt which we pack in as part of α so that the joint density function will be given by $\Gamma[k\alpha](u_1 u_2 \dots u_k)^{\alpha-1})/(\Gamma[\alpha]^k)$. The log likelihood function for the j^{th} bridge is given by:

$$l_j(\alpha) = (\alpha - 1) \sum_{i=1}^{k} u_i^{k,j} + \log(\Gamma[k\alpha]) - k \log(\Gamma[\alpha]) \tag{6}$$

We quote the standard results about the Dirichlet distribution:

$$\mathbb{E}[U_j^{k,j}] = \frac{1}{k} \quad , Var[U_i^{k,j}] = \frac{(k-1)}{k^2(\alpha k + 1)}, \quad Cov[U_i^{k,j}, U_l^{k,j}] = \frac{-1}{k(\alpha k + 1)},$$

Also, $\mathbb{E}[\log U_i^{k,j}] = \psi(\alpha) - \psi(k\alpha)$, $Var[\log U_i^{k,j}] = \psi_{(1)}(\alpha) - \psi_{(1)}(k\alpha)$, and $Cov[\log U_i^{k,j}] = -\psi_{(1)}(k\alpha)$, where $\psi_{(j)}(\theta) = \partial^j \psi(\theta)/\partial\theta^j$.
The maximum likelihood equation for the j^{th} set of k readings is given by

$$\frac{1}{k} \sum_{i=1}^{k} log(U_i^{k,j}) = \psi(\hat{\alpha}_{D,j,k}) - \psi(k\hat{\alpha}_{D,j,k}), \tag{7}$$

where $\hat{\alpha}_{D,j,k}$ denotes the estimator from the j^{th} bridge using the Dirichlet model .

3 The pooled estimator $\hat{\alpha}_{D,k,n}$

Next we compare the ML estimates for α from a set of k original gamma increments with those obtained by using the Dirichlet model for the corresponding bridge increments. For the j^{th} set of k gamma distributed increments the ML equation is given by:

$$\sum_{i=1}^{k} \log(X_{(j-1)k+i}) = \psi(\hat{\alpha}_{G,j,k}) + \log \left(\frac{\sum_{i=1}^{k} X_{(j-1)k+i}}{k\hat{\alpha}_{G,j,k}} \right) \tag{8}$$

where $\hat{\alpha}_{G,j,k}$ is the corresponding ML estimator. Using equations (7) and (8) we obtain the following,

$$\begin{aligned} \psi(\hat{\alpha}_{D,j,k}) - \psi(k\hat{\alpha}_{D,j,k}) &= \sum_{i=1}^{k} \log \left(\frac{X_{(j-1)k+i}}{\sum_{l=1}^{k} X_{(j-1)k+l}} \right) \\ &= \psi(\hat{\alpha}_{G,j,k}) - \log(k\hat{\alpha}_{G,j,k}) \end{aligned} \tag{9}$$

Hence $\psi(\hat{\alpha}_{D,j,k}) - \psi(k\hat{\alpha}_{D,j,k}) < \psi(\hat{\alpha}_{G,j,k}) - \psi(k\hat{\alpha}_{G,j,k})$, and since the function $\Psi(k,x) = \psi(x) - \psi(kx)$ is a smooth, increasing function, we can conclude that $\hat{\alpha}_{D,j,k} < \hat{\alpha}_{G,j,k}$.
We are using ideas due to Berman [2] where it is also proved directly that these two estimators are both positively biased.
Using standard techniques as in Cox and Snell [6], Bowman and Shenton [3] and Cordeiro and McCullagh [5] we can obtain bias estimates for both estimators. The bias of $\hat{\alpha}_{D,j,k}$ is given by:

$$\mathbb{E}\left[\hat{\alpha}_{D,j,k}\right] - \alpha = \frac{k^2\psi_{(2)}(k\alpha) - \psi_{(2)}(\alpha)}{2k(k\psi_{(1)}(k\alpha) - \psi_{(1)}(\alpha))^2} + O\left(\frac{1}{k^2}\right), \tag{10}$$

For the gamma-based estimator, the minimal bias is attained over the whole range of readings, $k = n$, for which the bias estimate is

$$\mathbb{E}[\hat{\alpha}_{G,n}] - \alpha = \frac{\alpha\psi_{(1)}(\alpha) - \alpha^2\psi_{(2)}(\alpha) - 2}{2n(\alpha\psi_{(1)}(\alpha) - 1)^2} + O\left(\frac{1}{n^2}\right) \tag{11}$$

The results above tells us that if we increase k we will eventually have a bias which is smaller than that for $\hat{\alpha}_{G,n}$. Thus if we take $k = n-1$ so that we have one bridge with the end-points being the first and last data points, we have a guarantee that the bias of the Dirichlet estimator is smaller than the bias of the gamma estimator which uses all the data points. But we can do better since we do not have to take k too large to obtain a smaller bias. For moderate values of k , m will not be so small. So we have m Dirichlet-based estimators $\hat{\alpha}_{D,j,k}$ which we can pool together. In fact, since they are independent random variables the best thing to do is to average them as follows:

$$\hat{\alpha}_{D,k,n} = \frac{1}{m}\sum_{j=1}^{m}\hat{\alpha}_{D,j,k}. \tag{12}$$

The double-sequence of estimators $\hat{\alpha}_{D,j,k}$ offers a lot of information about the behaviour of the underlying stochastic process through its realized path. Besides testing whether the time evolution of these parameter estimators does occur in an independently random manner, we can also investigate changes as we vary the value of k. These checks would help reveal internal probabilistic structure which would go beyond the simple mechanism for the gamma process.

We now examine how the biases of the estimators $\hat{\alpha}_{D,k,n}$ and $\hat{\alpha}_{G,j,n}$ vary as we change the values of n, k, and α. Using equations 10 and 11 we first define the function $g(k,n)$ which computes their difference:

$$g(k,n) = \frac{k^2\psi_{(2)}(k\alpha) - \psi_2(\alpha)}{2k[k\psi_{(1)}(k\alpha) - \psi_{(1)}(\alpha)]^2} - \frac{\alpha[\psi_{(1)}(\alpha) - \alpha\psi_{(2)}(\alpha)] - 2}{2n(\alpha\psi_{(1)}(\alpha) - 1)^2} \tag{13}$$

Figures 1 and 2 illustrate $g(k)$ for different values of α and n. From these graphs one notices that the difference in the bias gets larger as the value of α increases for any value of k. However, as n gets larger, the difference gets smaller for any value of k. Furthermore, from some value of k, depending on n and α, onwards the bias of $\hat{\alpha}_{D,k,n}$ will be less than that of $\hat{\alpha}_{G,n}$.

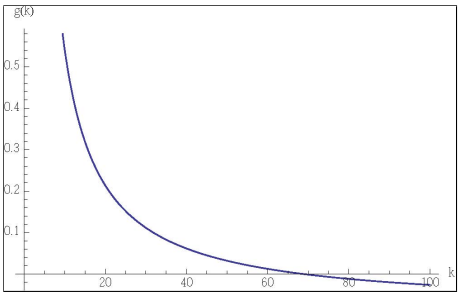


Fig. 1. $n = 100$, $\alpha = 3$

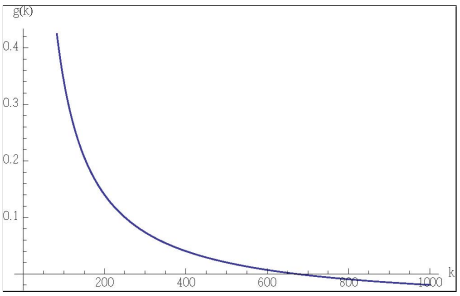


Fig. 2. $n = 1000$, $\alpha = 20$

This result is confirmed from a number of simulations the output of which is displayed in the table 1. In particular this table gives estimates of the bias of $\hat{\alpha}_{D,k,n}$ and $\hat{\alpha}_{G,n}$ as a percentage of the actual value of α. Furthermore the table gives the estimated variances of the said estimators.

k	$\widehat{\%\text{Bias}}(\hat{\alpha}_{D,k,n})$	$\widehat{\%\text{Bias}}(\hat{\alpha}_{G,n})$	$\widehat{\text{Var}}(\hat{\alpha}_{D,k,n})$	$\widehat{\text{Var}}(\hat{\alpha}_{G,n})$
50	3.9889	0.9898	0.7967	0.6650
100	1.8752	1.0049	0.6941	0.6598
250	0.7897	0.9354	0.7606	0.6507
290	0.6601	1.0468	0.7070	0.6873
n	0.5981	0.8826	0.6888	0.6934

Table 1. Simulation results using $n = 300$ and $\alpha = 10$

So the estimator we propose is one which corrects for the bias by using its estimate in the spirit of Giles and Feng [7]. This bias-corrected pooled MLE of α, which we shall denote by $\hat{\alpha}^*_{D,k,n}$, is defined by:

$$\hat{\alpha}^*_{D,k,n} = \hat{\alpha}_{D,k,n} - \frac{k^2\psi_{(2)}(k\hat{\alpha}_{D,k,n}) - \psi_{(2)}(\hat{\alpha}_{D,k,n})}{2k(k\psi_{(1)}(k\hat{\alpha}_{D,k,n}) - \psi_{(1)}(\hat{\alpha}_{D,k,n}))^2} \tag{14}$$

Note that the bias-correction procedure leaves the variance of the estimator unaltered. Table 2 illustrates the results obtained from a number of simulations during which the above bias correction technique was implemented. The said table gives the bias of $\hat{\alpha}_{D,k,n}$ and $\hat{\alpha}^*_{D,k}$ as a percentage of the true value α. Furthermore one can also observe that the estimated variances of the two estimators are comparable.

k	$\widehat{\%\text{Bias}}(\hat{\alpha}^*_{D,k,n})$	$\widehat{\%\text{Bias}}(\hat{\alpha}_{D,k,n})$	$\widehat{\text{Var}}(\hat{\alpha}^*_{D,k,n})$	$\widehat{\text{Var}}(\hat{\alpha}_{D,k,n})$
75	0.1987	2.5267	0.6781	0.7064
150	0.0872	1.2493	0.7905	0.7838
250	0.0785	0.7674	0.6435	0.6522

Table 2. Simulation results using $n = 300$ and $\alpha = 10$

4 Comparison of variances for $\hat{\alpha}_{D,k,n}$ and $\hat{\alpha}_{G,n}$

Finally we consider the variances of the above mentioned estimators. The estimator using ML on independent gamma distributed increments looks in a strong position. We know that it tends to the Cramer-Rao lower bound asymptotically. Using the usual Fisher matrix calculations an estimate for this variance is given by:

$$Var[\hat{\alpha}_{G,n}] = \frac{a}{n(a\psi_{(1)} - 1)} + O\left(\frac{1}{n^2}\right) \tag{15}$$

In the case of $\hat{\alpha}_{D,j,k}$, making suitable modifications to take the dependence between the $U_i^{k,j}$ into consideration, we apply again information matrix methodology with the Dirichlet distribution-based estimation to obtain $\mathbb{E}\left[\frac{\partial^2 l_j}{\partial \alpha^2}\right] = l_j'' = k^2\psi_{(1)} - k\psi_{(1)}(\alpha)$.
So MLE theory allows us to conclude that:

$$Var(l_j') = \mathbb{E}\left[\left(\sum_{i=1}^{k} \log\left(U_i^{k,j}\right) - k[\psi(\alpha) - \psi(k\alpha)]\right)^2\right] =$$
$$kVar\left[\log\left(U_i^{k,j}\right)\right] + k(k-1)Cov\left[U_i^{k,j}, U_l^{k,j}\right] = k\psi_{(1)}(\alpha) - k^2\psi_{(1)}(k\alpha) \tag{16}$$

Thus,

$$Var[\sqrt{k}\hat{\alpha}_{D,j,k}] = \frac{1}{\psi_{(1)} - k\psi_{(1)}(k\alpha)} + O\left(\frac{1}{k}\right).$$

It can be shown that as n and k tend to infinity, the formula for the variance of $\hat{\alpha}^*_{D,k,n}$ which is given in (17) tends from above to that of $\hat{\alpha}_{G,n}$ (15). We cannot improve on the variance-related performance of $\hat{\alpha}_{G,n}$ but we approach to it very closely with significant improvements in the bias.
Figure 3 and figure 4 shown below illustrate equations (15) and (17) with the variance of $\hat{\alpha}_{G,n}$ being indeed slightly less than that of $\hat{\alpha}_{D,k}$. As k increases the difference decreases fast to 0.

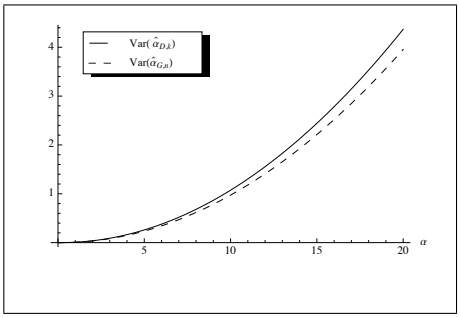


Fig. 3. $n = 200$, $k = 20$

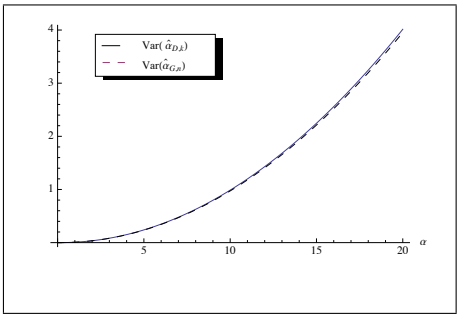


Fig. 4. $n = 200$, $k = 100$

5 Results about $\hat{\alpha}^*_{D,k,n}$

Having given all proofs and derivations above, the results are captured in the following theorem:

Theorem 1. *Estimator* $\hat{\alpha}^*_{D,k,n}$ *as defined in (14) is consistent, asymptotically unbiased with* $\mathbb{E}\left[\hat{\alpha}^*_{D,j,k}\right] - \alpha = O\left(\frac{1}{k^2}\right)$ *and*

$$Var[\hat{\alpha}^*_{D,k,n}] = \frac{1}{mk}\frac{1}{\psi_{(1)}(\alpha) - k\psi_{(1)}(k\alpha)} + O\left(\frac{1}{n^2}\right) \tag{17}$$

6 Conclusion

We started with a critique of the classical ML estimator widely used to identify gamma processes as models for given data. Arguing that the stationary increments property uses too little information about the process, we used results about the distributional evolution of the gamma process to construct a composite estimator. Over non-overlapping stretches of the data we fitted gamma-bridge-derived Dirichlet distributions to obtain a number of estimators for the same parameter. These were compared to the original one with reference to bias and variance. Using estimates for the asymptotic bias and variance we proposed the pooled estimator $\hat{\alpha}^*_{D,k,n}$ and showed it guarantees better statistical performance. It also allows modellers to see how different sections of the data behave by comparing the $\hat{\alpha}$'s from different bridges. Furthermore diagnostics can be developed to identify better anomalous sections within the data as estimates are compared with varying lengths k of the bridges. This helps build a dynamic picture of the data, which in practical applications like climate science and financial time series, can yield useful interpretations.

References

1.A. N. Avramidis, P. Lecuyer and P.A. Tremblay, Efficient Simulation of Gamma and Variance-Gamma Processes, *Proceedings of the 2003 Winter Simulation Conference*, 2003.

2.M. Berman. The maximum likelihood estimators of the parameters of the gamma distribution are always positively biased, *Communications in Statistics Theory and Methods*, 10, 7, 693-697, 1981.

3.K. O. Bowman and L.R. Shenton. Properties of estimators for the gamma distribution. *Communications in Statistics Simulation and Computation*, 11, 4, 377 519, 1982.

4.S. C. Choi and R. Wette. Maximum likelihood Estimation of the parameters of the Gamma distribution and their bias, *Technometrics*, 11, 4, 683-690, 1969

5.G. M. Cordeiro, P. McCullagh. Bias Correction in Generalized Linear Models, *Journal of the Royal Statistical Society. Series B (Methodological)*, 53, 3, 629-643.

6.D. R. Cox and E.J. Snell. A General Definition of Residuals, *Journal of the Royal Statistical Society*, 30, 2, 248-275, 1968.

7.T. S. Ferguson. A Bayesian Analysis of some nonparametric problems,*Annals of Statistics*,1, 2, 209-230, 1973.
8.D. E. Giles and H. Feng. Bias of the Maximum likelihood of the Two-Parameter Gamma Distribution Revisited, 2009.
9.M. Emery and M. Yor. A Parallel between Brownian Bridges and Gamma Bridges, *Publ. RIMS, Kyoto University*, 40, 669 688, 2004.
10.H. C. S. Thom. A Note on the Gamma Distribution, *Monthly Weather Review*, 86, 4, 117 119, 1958.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Identification of a Simple Homeostasis Stochastic Model Based on Active Principle of Adaptation

Innokentiy V. Semushin[1], Julia V. Tsyganova[2], and Anatoli G. Skovikov[3]

[1] Ulyanovsk State University, 42 Leo Tolstoy Str., Ulyanovsk, Russia
(E-mail: i2nokvsem@gmail.com)
[2] Ulyanovsk State University, 42 Leo Tolstoy Str., Ulyanovsk, Russia
(E-mail: tsyganovajv@gmail.com)
[3] Ulyanovsk State University, 42 Leo Tolstoy Str., Ulyanovsk, Russia
(E-mail: Transferfaculty@gmail.com)

Abstract. The Active Principle of Adaptation for linear time-invariant state-space stochastic MIMO filter systems is applied to human body temperature daily variation adaptive stochastic modeling.
Keywords: Adaptation, active principle, homeostasis, parameter estimation, stochastic modeling, thermoregulation.

1 Introduction

Stochasticity is a form of uncertainty among those discussed in general by Wolkenhauer[1]. As stated by Liao *et al.*[2] and many others in recent times, stochasticity is increasingly appreciated to play fundamental roles in systemic biology and bioinformatics especially when biological and clinical processes are studied at the cellular level. However, the reality in this life-critical area of research is such that the uncertainty of models is twofold: (1) models are stochastic in nature, id est, they are represented by stochastic differential equations, and (2) they are full of unknown parameters. Even if the stochastic models are linear in state, they are complex in the sense that many parameters need to be estimated from the data that are usually noisy and incomplete.

This is a traditional situation of uncertainty in many engineering problems where also emerges the question as to adaptive system modeling. Adaptation of models treated as *fitting models for data* is intended to eliminate or, at the very least, to reduce uncertainty. In Gibson's view, as cited in Semushin[3] the following three functions are considered as the determinant attributes of each adaptive system: (1) quickest Change Point Detection or more generally, Model Classification, (2) reliable Model Identification, and (3) adequate System Modification (or Change Parring). For engineering applications, there exist well-established mathematical methods for solving above three problems.

In systemic biology or medicine, systems are much more complicated than engineering ones because they include a living being, a human. This notwithstanding, it is yet very interesting to extend the approaches developed for engineering systems to biological or clinical processes.

Our work is aimed at such an extension. As a trial, benchmark task, we take a stochastic model of human body temperature. In so doing, we keep in mind three goals of research:

1. The *basic goal* is to extend the range of applicability of our active principle of adaptation (APA, Semushin[3]) to systemic biology, bioinformatics and medicine.
2. The *applied goal* is to implement the APA in some biomedical monitoring.
3. The *basic and applied goal* is to establish an all-purpose Computational Lab for Adaptive Stochastic Systems Modeling (CLASS-M).

The paper paves the way to the first goal only. Its outline is as follows. In Section 2, we start from a baseline model (BM) written as a system of stochastic differential equations for human body temperature daily variation. Section 3 shows what transformations are to be made towards the Discrete-time Standard Observable Model in order to build the Discrete Time Adaptive Kalman Filter in Section 4. Computational experiments with the filter are made in Section 5. We conclude the paper in Section 6.

2 Baseline model

Human body temperature regulation is a great example of how the homeostatic mechanism works. Let us consider human body temperature daily variation as it can be seen in many sources, for instance, in ANTRANIK.org (Fig. 1).

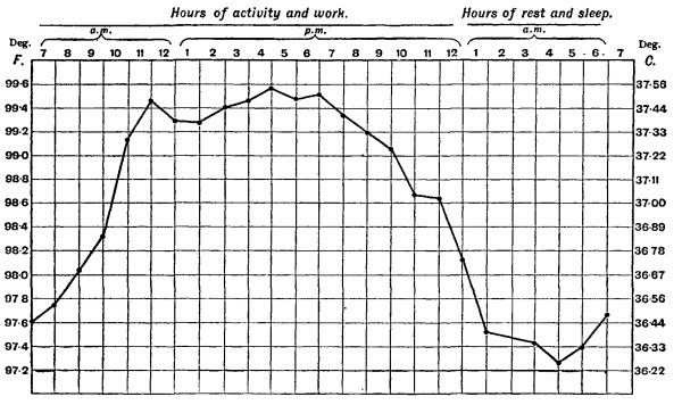


Fig. 1. Human body temperature daily variation (HBTDV). (Courtesy of ANTRANIK.org http://antranik.org/regulation-of-body-temperature/)

Consider the experimental data similar to Fig. 1 as a sample from a continuous-time stochastic process. Decompose it into the following additive components:

- $\bar{\theta}_t$, a mathematical expectation of temperature variation relative to daily mean temperature $\theta^\star$, for example, $\theta^\star = 36.7\ ^\circ C$,

- $\mathring{\theta}_t \triangleq \{\mathring{\theta}_t(\omega)\}$ represents a zero-mean stochastic process with $\omega \in \Omega$ being a point of a fundamental sample space Ω,
- $\tilde{\theta}_t = \mathring{\theta}_t + \theta^\star$, for which $\mathrm{d}\tilde{\theta}_t = \mathrm{d}\mathring{\theta}_t$ because $\theta^\star = \text{const}$.
- $\theta_t \triangleq \bar{\theta}_t + \tilde{\theta}_t$, the sum process modeling experimental data similar to those of Fig. 1.

For simplicity, in what follows we assume that $\bar{\theta}(t)$ is modeled by a A° C harmonic oscillator whose equation is well known:

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix}_t = \begin{bmatrix}0 & 1\\ -\omega_\mathrm{n}^2 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}_t, \qquad \begin{bmatrix}x_1\\ x_2\end{bmatrix}_0 = \begin{bmatrix}\theta_0{}^\circ\, C\\ \omega_0\,\mathrm{min}^{-1}\end{bmatrix} \tag{1}$$

$$y_t = \begin{bmatrix}1 & 0\end{bmatrix} x_t, \qquad t \in [0,\infty) \tag{2}$$

The solution to (1) is

$$\bar{\theta}_t \triangleq x_{1t} = A\sin(\omega_\mathrm{n} t + \varphi)\,, \quad A = \sqrt{\theta_0^2 + \left(\frac{\omega_0}{\omega_\mathrm{n}}\right)^2}\,, \quad \bar{\omega}_t \triangleq x_{2t}$$

$$\sin\varphi = \theta_0/A\,, \quad \cos\varphi = \left(\frac{\omega_0}{\omega_\mathrm{n}}\right)\Big/ A\,, \quad \tan\varphi = \theta_0 \Big/ \left(\frac{\omega_0}{\omega_\mathrm{n}}\right)$$

To obtain $A = 1^\circ\, C$ and $\varphi = 0$, the following initial condition must be assigned: $\begin{bmatrix}x_1\\ x_2\end{bmatrix}_0 = \begin{bmatrix}\theta_0 := 0^\circ\, C\\ \omega_0 := \omega_\mathrm{n}\,\mathrm{min}^{-1}\end{bmatrix}$. Obviously, to obtain $A = 0.65^\circ\, C$ and $\varphi = 0$, the initial condition are to be: $\begin{bmatrix}x_1\\ x_2\end{bmatrix}_0 = \begin{bmatrix}\theta_0 := 0^\circ\, C\\ \omega_0 := 0.65\,\omega_\mathrm{n}\,\mathrm{min}^{-1}\end{bmatrix}$, and so on.

To represent $\mathring{\theta}_t$, we introduce a Gaussian first order Markov process so as to write

$$\mathrm{d}\mathring{\theta}_t = -(1/T)\mathring{\theta}_t\,\mathrm{d}t + \mathrm{d}\beta_t\,, \qquad \lim_{t_0\to-\infty}\beta_{t_0} = 0 \text{ (a.s.)} \tag{3}$$

where β_t represents the scalar-valued Brownian motion (Wiener process, WP) with its constant diffusion $Q = 2\sigma^2/T$, or, in other words, the zero-mean process $\mathring{\theta}_t$ with mean squared value $\sigma^2 \triangleq \mathbf{E}\left\{\mathring{\theta}_t^2\right\} = QT/2$ and correlation time T. Here $\mathbf{E}\{\cdot\}$ denotes the expectation operator on Ω and $t_0 \to -\infty$ to provide wide-sense stationarity for $\mathring{\theta}_t$. Equation (3) is assumed to obtain the process $\mathring{\theta}_t$ with autocorrelation $\Psi_{\mathring{\theta}\mathring{\theta}}(\tau) \triangleq \mathbf{E}\left\{\mathring{\theta}_t\mathring{\theta}_{t+\tau}\right\} = \sigma^2\mathrm{e}^{-|\tau|/T}$. As a result, $\tilde{\theta}_t$ satisfies the following equation

$$\mathrm{d}\tilde{\theta}_t = -(1/T)(\tilde{\theta}_t - \theta^\star)\,\mathrm{d}t + \sigma\sqrt{2/T}\,\mathrm{d}\mathring{\beta}_t \tag{4}$$

where $\mathring{\beta}_t$ is the *standard* WP, that is, the unit diffusion Wiener process defined from $\beta_t \triangleq \sigma\sqrt{2/T}\mathring{\beta}_t = \eta\mathring{\beta}_t$, $\eta \triangleq \sigma\sqrt{2\lambda}$, $\lambda \triangleq 1/T$.

Having introduced $\tilde{\theta}_t$ as the third component into (1)–(2), we obtain

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\end{bmatrix}_t = \begin{bmatrix}0 & 1 & 0\\ -\omega_\mathrm{n}^2 & 0 & 0\\ 0 & 0 & -\lambda\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}_t + \begin{bmatrix}0\\ 0\\ \lambda\end{bmatrix}u_t + \begin{bmatrix}0\\ 0\\ \eta\end{bmatrix}\mathring{w}_t \tag{5}$$

$$y_t = \begin{bmatrix}1 & 0 & 1\end{bmatrix}x_t + v_t, \quad \begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}_0 = \begin{bmatrix}\theta_0{}^\circ\, C\\ \omega_0\,\mathrm{min}^{-1}\\ 0\end{bmatrix}, \; t \in [0,\infty) \tag{6}$$

There in the above equations $u_t \triangleq \theta^\star$, $\theta^\star$ is considered known, and $\mathring{w}_t$ is the *standard* Gaussian white noise defined by formally writing $\mathring{w}_t \triangleq \mathrm{d}\mathring{\beta}_t / \mathrm{d}t$, as it is usual in control literature. Thus, the baseline model (5), (6) has been built. In (6), a random measure error v_t has been introduced.

3 Towards the discrete-time standard observable model

Our sequence of model transformations is as follows.

3dCRPhM = 3-dimension Continuous-time Real-valued "Physical" Model is given by (5), (6). Assuming there $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_0 = \begin{bmatrix} 0^\circ C \\ \omega_\mathrm{n} \\ 0 \end{bmatrix}$, we have a $1^\circ C$ harmonic oscillator with $x_{1,t} = (1^\circ C)\sin\omega_\mathrm{n} t$, $x_{2,t} = (1^\circ C/\mathrm{min})\cos\omega_\mathrm{n} t$ with $\omega_\mathrm{n} = 2\pi/1440\ \mathrm{min}^{-1}$. Considering transformation $x = T_1 x^\star$ with

$$T_1 = \left[\begin{array}{c|c|c} 1 & 1 & 0 \\ \omega_\mathrm{n} & -\omega_\mathrm{n} & 0 \\ 0 & 0 & 1 \end{array}\right], \quad T_1^{-1} = \frac{1}{2}\left[\begin{array}{c|c|c} 1 & \omega_\mathrm{n}^{-1} & 0 \\ 1 & -\omega_\mathrm{n}^{-1} & 0 \\ 0 & 0 & 2 \end{array}\right]$$

we change to the next model.

3dCRCM = 3-dimension Continuous-time Real-valued Canonical Model

$$\begin{bmatrix} \dot{x}_1^\star \\ \dot{x}_2^\star \\ \dot{x}_3^\star \end{bmatrix}_t = \begin{bmatrix} 0 & -\omega_\mathrm{n} & 0 \\ \omega_\mathrm{n} & 0 & 0 \\ 0 & 0 & -\lambda \end{bmatrix} \begin{bmatrix} x_1^\star \\ x_2^\star \\ x_3^\star \end{bmatrix}_t + \begin{bmatrix} 0 \\ 0 \\ \lambda \end{bmatrix} \theta^\star + \begin{bmatrix} 0 \\ 0 \\ \sigma\sqrt{2\lambda} \end{bmatrix} \mathring{w}_t \tag{7}$$

$$y_t = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} x_t^\star + v_t, \quad \begin{bmatrix} x_1^\star \\ x_2^\star \\ x_3^\star \end{bmatrix}_0 = \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix} \tag{8}$$

To make possible further using the active principle of filter adaptation in order to estimate the unknown parameters λ and σ, we need to have the discrete-time standard observable model. Getting ready to this final step, we construct two more models as follows.

3dDRCM = 3-dimension Discrete-time Real-valued Canonical Model. Here we omit the $_\star$s and *s for variables and matrices and denote the sampling interval $\tau \triangleq \Delta t \triangleq t_{i+1} - t_i = \mathrm{const}$:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{t+1} = \underbrace{\begin{bmatrix} c & -s & 0 \\ s & c & 0 \\ 0 & 0 & d \end{bmatrix}}_{\Phi} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_t + \underbrace{\begin{bmatrix} 0 \\ 0 \\ a \end{bmatrix}}_{\Psi} u_t + \underbrace{\begin{bmatrix} 0 \\ 0 \\ b \end{bmatrix}}_{\Gamma} \mathring{w}_{\mathrm{d}t} \tag{9}$$

$$y_t = \underbrace{\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}}_{H} x_t + v_t, \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_0 = \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix} \tag{10}$$

$$c \triangleq \cos\omega_\mathrm{n}\tau,\ s \triangleq \sin\omega_\mathrm{n}\tau,\ d \triangleq e^{-\lambda\tau}$$

$$a \triangleq 1 - d,\ b \triangleq \sigma\sqrt{1 - d^2}$$

where $\mathring{w}_{\mathrm{d}t}$ is a *discrete* standard (that is, with unit covariance $\mathring{Q}_{\mathrm{d}} = 1$) Gaussian white noise.

3dDSOM = 3-dimension Discrete-time Standard Observable Model. Entering upon the construction of this final model, we determine the observability matrix

$$W_\star \triangleq \left[\, H^T \,\middle|\, (H\Phi)^T \,\middle|\, (H\Phi^2)^T \,\right]^T \tag{11}$$

This matrix can be calculated either by hand or with Maple. Below are the results of our by-hand calculations:

$$W_\star = \left[\begin{array}{c|c|c} 1 & 1 & 1 \\ c+s & c-s & d \\ f+g & f-g & d^2 \end{array}\right], \quad f \triangleq \cos 2\omega_{\mathrm{n}}\tau,\ g \triangleq \sin 2\omega_{\mathrm{n}}\tau, \tag{12}$$

$$\det W_\star = 2(-d^2 s + dg + sf - cg) = 2(dg - s(1+d^2))$$

The sought 3dDSOM is obtained as a result of transform $x^\star = W_\star x$ in the following equations:

$$\begin{bmatrix} x_1^\star \\ x_2^\star \\ x_3^\star \end{bmatrix}_{t+1} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 \end{bmatrix}}_{\Phi_\star} \begin{bmatrix} x_1^\star \\ x_2^\star \\ x_3^\star \end{bmatrix}_t + \underbrace{\begin{bmatrix} 1 \\ d \\ d^2 \end{bmatrix}}_{\Psi_\star} (1-d)\,\theta^\star + \underbrace{\begin{bmatrix} 1 \\ d \\ d^2 \end{bmatrix}}_{\Gamma_\star} \sigma\sqrt{1-d^2}\,\mathring{w}_{\mathrm{d}t} \tag{13}$$

$$y_t = \underbrace{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}}_{H_\star} x_t + v_t, \quad \begin{bmatrix} x_1^\star \\ x_2^\star \\ x_3^\star \end{bmatrix}_0 = \begin{bmatrix} 0 \\ s \\ g \end{bmatrix} \tag{14}$$

$$\begin{aligned} -a_3 &= d\,, \\ -a_2 &= -1 - 2d\cos\omega_{\mathrm{n}}\tau\,, \\ -a_1 &= d + 2\cos\omega_{\mathrm{n}}\tau \end{aligned}$$

An effective way to check the constructed 3dDSOM is to compare it with the same result produced by Maple™.

In the case when the daily average temperature $\theta^\star$ is considered unknown and so is to be also estimated, we infer usage of Maple the only reasonable way to change from the following 4dDRCM

$$\begin{aligned} x_{t+1} &= \begin{bmatrix} \cos\omega_{\mathrm{n}}\tau & -\sin\omega_{\mathrm{n}}\tau & 0 & 0 \\ \sin\omega_{\mathrm{n}}\tau & \cos\omega_{\mathrm{n}}\tau & 0 & 0 \\ 0 & 0 & d & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x_t + \begin{bmatrix} 0 \\ 0 \\ \sigma\sqrt{1-d^2} \\ 0 \end{bmatrix} \mathring{w}_{\mathrm{d}t} \\ y_t &= \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} x_t + v_t \end{aligned}$$

to the 4dDSOM (this case is omitted here for saving room).

4 Discrete-time adaptive Kalman filter

Our benchmark point for constructing the APA-b-AKF should be the Discrete-time Standard Observable Model (DSOM) written, according to Semushin[3], in the general case as follows:

$$\mathcal{D}^{\star}(\theta):\ \begin{array}{l} x^{\star}_{t+1} = \Phi_{\star}x^{\star}_{t} + \Psi_{\star}u_t + \Gamma_{\star}\mathring{w}_{\mathrm{d}t}\,,\ t\in\mathbb{Z}_{+}\,,\ x^{\star}\in\mathbb{R}^{n} \\ y_t = H_{\star}x^{\star}_{t} + v_t\,,\ t\in\mathbb{Z}_{1}\,,\ y\in\mathbb{R}^{m} \end{array} \tag{15}$$

Omitting details from the general theory by Semushin[3], recall that $W_{\star}$, the observability matrix, in our case is given by (11) and (12).

If we work in the context of DSOM, the set

$$\mathcal{A}^{\star} = \left\{ \mathfrak{M}^{\star}(\widehat{\theta}\,) \,\middle|\, \widehat{\theta}\in\Theta\subset\mathbb{R}^{l} \right\} \tag{16}$$

of adaptive models $\mathfrak{M}^{\star}(\widehat{\theta}\,)$ should be used with $\widehat{\theta}$ denoting an estimate of $\theta\in\mathbb{R}^{l}$. Reasoning from the Kalman (optimal) filter $\mathfrak{M}^{\star}(\theta\,)$, we build the adaptive model

$$\mathfrak{M}^{\star}(\widehat{\theta}\,):\ \begin{array}{l} \widehat{g}_{t+1|t} = A\widehat{g}_{t|t-1} + \Psi_{*}u_t + C\eta_{t|t-1} \\ y_t = H_{\star}\widehat{g}_{t|t-1} + \eta_{t|t-1} \end{array} \tag{17}$$

or equivalently (due to $C=AD$) the model

$$\mathfrak{M}^{\star}(\widehat{\theta}\,):\ \begin{array}{l} \widehat{g}_{t+1|t} = A\widehat{g}_{t|t} + \Psi_{*}u_t \\ \widehat{g}_{t|t} = \widehat{g}_{t|t-1} + D\eta_{t|t-1} \\ y_t = H_{\star}\widehat{g}_{t|t-1} + \eta_{t|t-1} \end{array} \tag{18}$$

with $H_{\star}$ and $A=A_{\star}$ taken in the form of (13) and (14).

With the understanding that errors

$$\begin{array}{ll} e_{t+1|t} \triangleq x^{\star}_{t+1} - \widehat{g}_{t+1|t}\,, & e_{t|t} \triangleq x^{\star}_{t} - \widehat{g}_{t|t} \\ r_{t+1|t} \triangleq x^{\star}_{t+1|t} - \widehat{g}_{t+1|t}\,, & r_{t|t} \triangleq x^{\star}_{t|t} - \widehat{g}_{t|t} \end{array} \tag{19}$$

are fundamentally unmeasurable, we construct the auxiliary performance index $\mathcal{J}^{\mathrm{a}}_{t}(\widehat{\theta}\,)$ which guarantees (proofs are done in Semushin[3]):

True (Unbiased) System Identifiability

$$\min_{\widehat{\theta}} \mathcal{J}^{\mathrm{a}}_{t}(\widehat{\theta}\,) \iff \mathfrak{M}^{\star}(\widehat{\theta}\,) \equiv \mathfrak{M}^{\star}(\theta^{\dagger})$$

5 Computational experiments made with MATLAB®

From now on, let $\rho\in\mathbb{R}^{l}$ be the vector of unknown parameters. Consider cases:

1. Parameter λ is unknown, parameter σ is known, i. e. $\rho=\lambda$.
2. Parameter σ is unknown, parameter λ is known, i. e. $\rho=\sigma$.
3. Parameters λ and σ are unknown, i. e. $\rho=[\lambda\,\big|\,\sigma]^{T}$.

To measure the quality of estimates compute three values (Table 1). Experimental conditions are shown in Table 2. The APA-b-AKF has been implemented in a robust square-root form developed by Tsyganova[4]. Results obtained are as shown in Tables 3 through 5.

1. The estimation sample mean	$\text{MEAN} = \frac{1}{N_{EXP}} \sum_{j=1}^{N_{EXP}} \hat{\rho}_j$
2. The root-mean-square error	$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{N_{EXP}} \|\|\hat{\rho}_j - \rho_*\|\|^2}{N_{EXP}}}$
3. The mean absolute percentage error	$\text{MAPE} = \frac{100\%}{N_{EXP}} \sum_{j=1}^{N_{EXP}} \frac{\|\|\hat{\rho}_j - \rho_*\|\|}{\|\|\rho_*\|\|},$
$\hat{\rho}_j$ is the parameter estimation obtained in the j-th experiment run, ρ_* is the true value of model parameter.	

Table 1. Values to measure the quality of parameter estimates

Number of experiments	$N_{EXP} = 100$
Sampling interval (min)	$\tau = 5$
True model noise parameter	$\sigma_* = 0.3$
Model noise correlation time (min)	$T_{\text{n}} = 24 \cdot 60$, $\omega_{\text{n}} = 2\pi/T_{\text{n}}$
Average daily temperature	$\theta^* = 36.85\ ^\circ C$
Covariance of measurement noise	$R = 0.1$
Initial values	$x_0^* = [0,\ 0.65\sin(\omega_{\text{n}}\tau),\ 0.65\sin(2\omega_{\text{n}}\tau)]$
True parameter values λ	$\lambda_* = 1/T = 0.01(6)$, $T = 60$
Initial value for estimate of λ	$\lambda_0 = 1/(2T)$
Initial value for estimate of σ	$\sigma_0 = 1.$

Table 2. Experimental conditions for estimating parameters λ and σ.

N	MEAN	RMSE	MAPE
0.5	0.01673842	0.00176175	8.53495848
1	0.01668899	0.00128099	6.00706384
3	0.01677946	0.00007005	3.24514550
6	0.01669467	0.00005602	2.80900894
12	0.01665221	0.00004990	2.37861756
24	0.01670608	0.00004936	2.40873864
48	0.01664612	0.00004899	2.43864040
72	0.01669707	0.00004534	2.20453053

Table 3. Experimental results for estimating parameter λ (N stands for hours of measurements collection).

6 Conclusion

In this paper, The Active Principle of Adaptation for linear time-invariant state-space stochastic MIMO systems is applied to human body temperature daily variation adaptive stochastic modeling.

The baseline HBTDV model has been patterned after the physical data available. The adaptive model $\mathfrak{M}^\star(\widehat{\theta})$, a replica of the Kalman filter for the standard observable data model, has been specified.

Computational experiments have been made to demonstrate the applicability of our Active Principle of Adaptation to bioinformatics problems.

N	MEAN	RMSE	MAPE
0.5	0.32169022	0.34800311	96.09089193
1	0.31693915	0.30205901	81.97886696
3	0.30406180	0.18993025	49.42388770
6	0.27961800	0.12641243	32.91595681
12	0.29865439	0.08187854	22.40155093
24	0.29767603	0.06173083	16.00851023
48	0.29723777	0.03914056	10.35061636
72	0.29601063	0.03366058	9.09541878

Table 4. Experimental results for estimating parameter σ (N stands for hours of measurements collection).

N	MEAN	RMSE	MAPE
0.5	$\bar{\rho} = \begin{bmatrix} 0.01659436 \\ 0.10575119 \end{bmatrix}$	0.28056287	85.14674225
1	$\bar{\rho} = \begin{bmatrix} 0.01669789 \\ 0.11527021 \end{bmatrix}$	0.27826336	84.28790200
3	$\bar{\rho} = \begin{bmatrix} 0.01658014 \\ 0.16829851 \end{bmatrix}$	0.22572891	65.65149641
6	$\bar{\rho} = \begin{bmatrix} 0.01669664 \\ 0.23287266 \end{bmatrix}$	0.16211670	44.23780054
12	$\bar{\rho} = \begin{bmatrix} 0.01666193 \\ 0.27265297 \end{bmatrix}$	0.08541397	22.48302905
24	$\bar{\rho} = \begin{bmatrix} 0.01672629 \\ 0.28222717 \end{bmatrix}$	0.06480563	17.75995499
48	$\bar{\rho} = \begin{bmatrix} 0.01668300 \\ 0.29025029 \end{bmatrix}$	0.03909148	10.45862558
72	$\bar{\rho} = \begin{bmatrix} 0.01666895 \\ 0.29820202 \end{bmatrix}$	0.03539374	9.21590355

Table 5. Experimental results for estimating parameters ρ (N stands for hours of measurements collection).

7 Acknowledgments

This paper was partly supported by the RFBR grant No. 13-01-97035.

References

1.O. Wolkenhauer. *Data engineering: Fuzzy mathematics in system theory and data analysis*, John Wiley & Sons, Inc., New York, 2001.

2.D. Liao, L. Estévez-Salmerón and Th. D. Tlsty. Generalized principles of stochasticity can be used to control dynamic heterogeneity. *Physical Biology*, 9, 6, 065006, 2012. http://iopscience.iop.org/1478-3975/9/6/065006/. doi:10.1088/1478-3975/9/6/065006.

3.I. V. Semushin, Adaptation in stochastic dynamic systems—survey and new results II, *Int. J. Communications, Network and System Sciences*, 4, 4, 266–285, 2011. http://www.SciRP.org/journal/ijcns. doi:10.4236/ijcns.2011.44032.

4.Yu. V. Tsyganova, Computing the gradient of the auxiliary quality functional in the parametric identification problem for stochastic systems, *Automation and Remote Control*, 72, 9, 1925–1940, 2011.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Optimal control of systems with several replenishment sources

Ksenia Shakhgildyan

Department of Mathematics and Mechanics, Lomonosov Moscow State University,
GSP-1, Leninskie Gory, Moscow, 119991, Russia
(E-mail: ksy-shakhgildyan@yandex.ru)

Abstract The aim of the paper is to investigate a model of inventory management with several sources of replenishment. There is a possibility of sending orders to either of two suppliers or both of them. It is supposed that the first supplier delivers orders immediately, while the other one is unreliable delivering the orders immediately only with probability $p \in (0,1)$. The optimal strategy of the company can be determined, namely, the values of orders at any step of the multi-step model for various values of the parameters are obtained. However, the information about the parameters as well as about distribution functions is often incomplete. This is the reason why we examine the sensitivity of solutions to small changes in parameters.
Keywords: Periodic-review inventory system, Optimal control, Unreliable suppliers, Sensitivity.

1 Introduction and model description

The development of a vast body of knowledge known as modern inventory theory is rapid nowadays and has a range of applications to practical situations. Modern information technology has created new possibilities for more sophisticated and efficient control of supply chains. Most organizations can substantially reduce their costs associated with the flow of materials. Inventory control techniques are crucial components in this development process.

A model of inventory management with the participation of several sources of replenishment is examined.

Let c_1 be the unit price with the first supplier and we assume that delivery is made immediately. In its turn, the second supplier makes delivery immediately with probability p and at the beginning of the next period with probability $q = 1 - p$. Denote by c_2 the corresponding unit price. At the beginning of each period (a day, a week, a month, etc.) the decision to order a certain amount of goods from the first and second supplier is made, namely $z_1 \geq 0$ and $z_2 \geq 0$ respectively. We also consider the storage cost h and deficiency payment for unit price r. Suppose x is the initial stock, from this moment onwards claims are received periodically, namely the amount ξ_i is demanded during the i-th period, $i \geq 1$. We assume that $\{\xi_i\}_{i\geq 0}$ form a sequence of mutually independent random variables with a common distribution function $F(\cdot)$ having density $\varphi(s) > 0$ for $s \in [a,b]$, where $a \geq 0$.

Denote by $f_n(x)$ minimum average discounted costs over n periods.
Estimated costs for one period are equal to

$$L(v) = E[r(\xi_1 - v)^+ + h(v - \xi_1)^+], \quad v = x + z_1, \; u = v + z_2.$$

By definition, put

$$G_n(u,v) = (c_1 - c_2)v + c_2 u + pL(u) + qL(v) + \alpha E f_{n-1}(u - \xi_1),$$

where α is the discount factor.
In this case, according to the Bellman's Principle of Optimality for $n \geq 1$ we obtain the following recurrence relations:

$$f_n(x) = -c_1 x + \min_{x \leq v \leq u} G_n(u,v), \quad f_0(x) = 0.$$

Note that the parameters of the model u and v corresponding to the minimum costs mean that it is optimal to order the amount $v - x$ of goods from the first supplier and the amount $u - v$ from the second.

2 Notations and preliminary results

Let us introduce the partial derivative of a function $G_n(u,v)$ with respect to the variables v and u:

$$A(v) = \frac{\partial G_1(u,v)}{\partial v} = c_1 - c_2 + qL'(v),$$

$$B_n(u) = \frac{\partial G_1(u,v)}{\partial u} = c_2 + pL'(u) + \alpha \int_0^\infty f'_{n-1}(u-t)\varphi(t)dt,$$

As well as a function

$$C_n(v) = \frac{\partial G_1(v,v)}{\partial v} = A(v) + B_n(v) = c_1 + L'(v) + \alpha \int_0^\infty f'_{n-1}(v-t)\varphi(t)dt.$$

It can be shown in the usual way that all the functions above are nondecreasing.

Critical levels v^*, u_n и v_n, in case of the existence, are defined as the solutions of the equations

$A(v^*) = 0$, $B_n(u_n) = 0$ and $C_n(v_n) = 0$ respectively.

In case of the nonexistence set them equal to $-\infty$ by definition. That is natural by the monotonicity of functions.
At each step we need to find the value of $\min\limits_{x \leq v \leq u} G_n(u,v)$

First let us obtain the values of $u_n(x)$ and $v_n(x)$ corresponding to the minimum costs depending on the relation of critical levels.

Lemma 1. *Let us consider 2 cases:*
1) If $u_n \geq v^*$, *the argument of* $\min\limits_{x \leq v \leq u} G_n(u,v)$ *is* $\begin{cases} v_n(x) = max(x, v^*) \\ u_n(x) = max(x, u_n). \end{cases}$
2) If $u_n < v^*$, *the argument of* $\min\limits_{x \leq v \leq u} G_n(u,v)$ *is* $v_n(x) = u_n(x) = max(x, v_n)$.

The proof follows directly from the analysis of partial derivatives of the function $G_n(u,v)$.
The next Lemma is needed for the sequel.

Lemma 2. *Suppose* $B_n(v^*) \geq 0$ *then* $u_n \leq v_n \leq v^*$, *whereas for* $B_n(v^*) < 0$ *we have* $v^* < v_n < u_n$.

Proof. Since $A(v^*) = C(v^*) - B(v^*) = 0$, we see that $C(v^*) = B(v^*)$.
It now follows from the monotonicity of functions that for $v < v^*$ we have $A(v) = C(v) - B(v) < 0 \Leftrightarrow C(v) < B(v)$, whereas for $v \geq v^*$
$A(v) = C(v) - B(v) \geq 0 \Leftrightarrow C(v) \geq B(v)$. Clearly, the Lemma statement is true.□

3 Optimal control for $n = 1$

In this case $A(v) = c_1 - c_2 + qL'(v) = 0$, $B_1(u) = c_2 + pL'(u) = 0$,
$C_1(v) = c_1 + L'(v) = 0$. From the explicit form of functions, taking into account $0 \leq F(x) \leq 1$, it can easily be checked that:

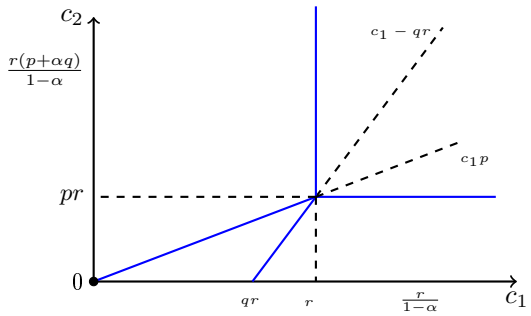
$$\begin{cases} v^* = F^{-1}\left(\frac{qr+c_2-c_1}{q(h+r)}\right) & \text{for } (c_1, c_2) \text{ such that } c_2 \geq c_1 - qr; \\ v^* = -\infty & \text{in the converse case;} \end{cases}$$

$$\begin{cases} u_1 = F^{-1}\left(\frac{pr-c_2}{p(h+r)}\right) & \text{for } (c_1, c_2) \text{ such that } c_2 \geq pr; \\ u_1 = -\infty & \text{in the converse case;} \end{cases}$$

$$\begin{cases} v_1 = F^{-1}\left(\frac{r-c_1}{h+r}\right) & \text{for } (c_1, c_2) \text{ such that } c_1 \leq r; \\ v_1 = -\infty & \text{in the converse case.} \end{cases}$$

According to Lemma 1, we need to compare u_1 and v^*, in order to find the optimal control. First note that $F(u) \geq F(v) \Leftrightarrow u \geq v$. Consequently,
$u_1 \geq v^* \Leftrightarrow \frac{pr-c_2}{p(h+r)} \geq \frac{qr+c_2-c_1}{q(h+r)} \Leftrightarrow prq - c_2 q \geq prq + pc_2 - pc_1$
$\Leftrightarrow -c_2(1-p) \geq pc_2 - pc_1 \Leftrightarrow c_2 \leq pc_1$. Therefore, we have

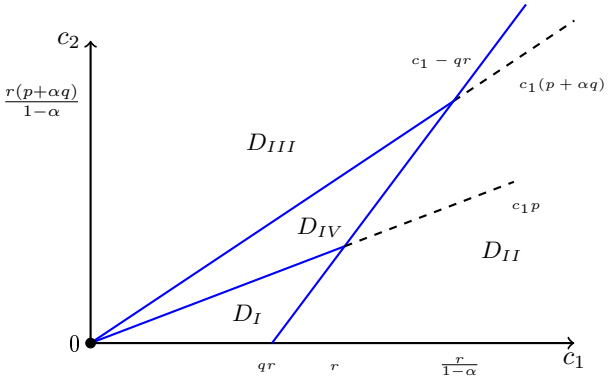
Theorem 1. *The optimal values of orders are the following:*
For $\{(c_1, c_2) : \ c_2 \geq pr, \ c_1 \geq r\}$ *we obtain that* $v_1(x) = u_1(x) = x$.
Whereas for $\{(c_1, c_2) : \ c_2 \leq c_1 - qr, \ c_2 \leq pr\}$ *the optimal orders are* $v_1(x) = x$, $u_1(x) = \max(x, u_1)$.
In turn, for $\{(c_1, c_2) : \ c_1 - qr \leq c_2 \leq pc_1\}$ *we have* $v_1(x) = \max(x, v^*)$, $u_1(x) = \max(x, u_1)$.
The remaining $\{(c_1, c_2) : \ c_2 \geq c_1 - qr, \ c_1 \leq r\}$ *coincides with* $v_1(x) = u_1(x) = = \max(x, v_1)$.



Proof. 1) Let us consider $\{(c_1, c_2) : \ c_2 \geq pr, \ c_1 \geq r\}$.
By Lemma 1 we have that $v_1(x) = u_1(x) = max(x, v_1)$ in the area above the line $c_2 = c_1 p$. Moreover, $v_1 = -\infty$ in the given area, hence $v_1(x) = u_1(x) = x$.
On the other hand $v_1(x) = max(x, v^*), \ u_1(x) = max(x, u_1)$ in the area below the line $c_2 = c_1 p$. However $v^* = u_1 = -\infty$ in this case, thus $v_1(x) = u_1(x) = x$.
2) As far as $\{(c_1, c_2) : \ c_2 \leq c_1 - qr, \ c_2 \leq pr\}$ is concerned
$v_1(x) = max(x, v^*), \ u_1(x) = max(x, u_1)$ by Lemma 1. Note that $v^* = -\infty$ in the considered area, so we obtain $v_1(x) = x, \ u_1(x) = max(x, u_1)$.
3) When it comes to $\{(c_1, c_2) : \ c_1 - qr \leq {}_2 \leq pc_1\}$ neither v^* nor u_1 are equal to $-\infty$. Consequently, $v_1(x) = \max(x, v^*)$, $u_1(x) = \max(x, u_1)$.
4) In the latter case, namely $\{(c_1, c_2) : \ c_2 \geq c_1 - qr, \ c_1 \leq r\}$, v_1 is not equal to $-\infty$. It now follows from Lemma 1 that $v_1(x) = u_1(x) = \max(x, v_1)$.□

4 Optimal control for $n = 2$

In order to discover optimal orders it is necessary to examine ranges of values of (c_1, c_2): D_I, D_{II}, D_{III} and D_{IV}.



1) Let us look first of all at D_I.
First note that $v^* \leq v_1 \leq u_1$ in this area. What we need is to get an equation for $u_2(x)$.

Lemma 3.

$$f_1'(x) = \begin{cases} -c_1, \ x < v^* \\ -c_2 + qL'(x), \ v^* \leq x < u_1 \\ L'(x), \ x \geq u_1 \end{cases} = -c_1 + \begin{cases} 0, \ x < v^*; \\ A(x), \ v^* \leq x < u_1; \\ A(x) + B_1(x), \ x \geq u_1. \end{cases}$$

Proof. This proposition can be proved by direct calculations. It now follows that $B_2(u) = B_1(u) + \alpha E f_1'(u - \xi_1) = c_2 - \alpha c_1 - pr + p(h + r)F(u) + \alpha F(u - - u_1)(c_2 - rp) + \alpha F(u - v^*)(c_1 - c_2 - qr) + \alpha(h + r)(p \int_0^{u-u_1} F(u - t)\varphi(t)dt + + q \int_0^{u-v^*} F(u - t)\varphi(t)dt)$, where u_2 is the root of the equation.
It can be easily seen that $B_2(v^*) = \frac{c_2 - c_1(p+\alpha q)}{q} < 0$ since $c_2 < c_1(p + \alpha q)$ in the considered area. So we can conclude that $v^* \leq u_2$ by Lemma2.
Finally, by Lemma 1 we obtain $v_2(x) = max(x, v^*), \ u_2(x) = max(x, u_2)$.□

Lemma 4. $u_2 \geq u_1$.

Proof. Consider the difference $B_2(u)-B_1(u) = -c_1\alpha+\alpha\int_0^{u-u_1} B_1(u-t)\varphi(t)dt+ \alpha\int_0^{u-v^*} A(u-t)\varphi(t)dt$. Substituting u_1 in the equation, we get:
$B_2(u_1) = -c_1\alpha + \alpha\int_0^{u_1-v^*} A(u_1-t)\varphi(t)dt =$
$= -c_1\alpha + c_1\alpha F(u_1-v^*) - c_2\alpha F(u_1-v^*) + q\alpha\int_0^{u_1-v^*} L'(u_1-t)\varphi(t)dt \leq$
$-c_2\alpha F(u_1-v^*) + q\alpha\int_0^{u_1-v^*} L'(u_1-t)\varphi(t)dt < 0$, since $L'(u_1-t) < 0$ over the region of integration. Thus we have $u_1 \leq u_2$ by the monotonicity of $B_2(u)$.□

2) Now let us turn to D_{II}.
To begin with, $v^* = -\infty$ in this area, therefore $u_2 \geq v^*$. Now if we recall Lemma 1, we get $v_2(x) = max(x, v^*) = x,\ u_2(x) = max(x, u_2)$.
As before, the aim is to obtain an equation for $u_2(x)$.

Lemma 5.
$$f_1'(x) = \begin{cases} -c_2 + qL'(x), & x < u_1 \\ L'(x), & x \geq u_1 \end{cases} = -c_1 + \begin{cases} A(x), & x < u_1; \\ A(x) + B_1(x), & x \geq u_1. \end{cases}$$

Consequently, $B_2(u) = B_1(u) + \alpha E f_1'(u-\xi_1) =$
$= c_2(1-\alpha) - r(p+\alpha q) + p(h+r)F(u) + \alpha F(u-u_1)(c_2-rp) +$
$+ \alpha(h+r)(\int_0^{u-u_1} F(u-t)\varphi(t)dt + q\int_{u-u_1}^{\infty} F(u-t)\varphi(t)dt)$,
where u_2 is the root.

Interestingly, $f_1' = L'(x)$ for $c_2 \geq pr$. In its turn,
$B_2(u) = c_2 - r(p+\alpha) + (h+r)(pF(u) + \alpha F^{*2}(u))$. Therefore
$F(u_2) + \frac{\alpha}{p}F^{*2}(u_2) = \frac{r(p+\alpha)-c_2}{p(h+r)}$. Since $F(u) + \frac{\alpha}{p}F^{*2}(u) \geq 0$, we see that
$u_2 = -\infty$ for $c_2 \geq r(p+\alpha)$.

Lemma 6. $u_2 \geq u_1$.

Proof. Consider the difference $B_2(u) - B_1(u) =$
$= -c_1\alpha + \alpha\int_0^{u-u_1} B_1(u-t)\varphi(t)dt + \alpha\int_0^{\infty} A(u-t)\varphi(t)dt$
Substituting u_1 in the equation, we get: $B_2(u_1) = -c_1\alpha+\alpha\int_0^{\infty} A(u-t)\varphi(t)dt =$
$-c_2\alpha - qr\alpha + \alpha q(h+r)F^{*2}(u_1) \leq -_2\alpha - qr\alpha + q\alpha(h+r)F(u_1) = -\frac{c_2\alpha}{p} < 0$
As a result we have $u_1 \leq u_2$ by the monotonicity of $B_2(u)$.□

The results for D_{III} and D_{IV} can be acquired in the same way.

It should be mentioned that optimal orders are $v_2(x) = u_2(x) = max(x, v_2)$ for D_{III}, $v_2 \geq v_1$. Whenever $c_1 \geq r$, it follows that $f_1' = L'(x)$ and
$C_2(v) = c_1 - r(1+\alpha) + (h+r)(F(v) + \alpha F^{*2}(v))$. Arguing as above, we see that: $F(v_2) + \alpha F^{*2}(v_2) = \frac{r(1+\alpha)-c_1}{h+r}$.
Let us also remark that $F(v) + \alpha F^{*2}(v) \geq 0$, hence $v_2 = -\infty$ for $c_2 \geq r(1+\alpha)$.

5 Optimal control for any step n.

Denote by $v_n(x)$ and $u_n(x)$ optimal values of orders at the $n-$th step, meaning that we need to purchase the amount $v_n(x) - x$ of goods from the first supplier

and the amount $u_n(x) - v_n(x)$ from the second.

Our main results are the following:

Theorem 2. *For $(c_1, c_2) \in D_I$ we have:*
$$f_n'(x) = -c_1 + \begin{cases} 0, & x < v^*; \\ A(x), & v^* \le x < u_n; \\ A(x) + B_n(x), & x \ge u_n. \end{cases}$$
The corresponding values of orders are $v_n(x) = \max(x, v^*)$ and $u_n(x) = \max(x, u_n)$. Moreover, the sequence u_n is nondecreasing.

Proof. The proof is by induction on n from $n = 1$. The basis has already been proved in the 3th and 4th sections. Now let us assume that conditions of the Theorem are true for $f_m'(x)$, $m \le n-1$.

By direct calculations, using $u_{n-2} \le u_{n-1}$, we obtain: $B_n(u) - B_{n-1}(u) =$
$= \alpha Q_{n-1}(u)$, where $Q_{n-1}(u) = \int\limits_0^{u-u_{n-1}} (B_{n-1}(u-t) - B_{n-2}(u-t))\varphi(t)dt -$
$- \int\limits_{u-u_{n-1}}^{u-u_{n-2}} B_{n-2}(u-t)\varphi(t)dt$. Substituting u_{n-1}, we get:
$$B_n(u_{n-1}) = -\alpha \int\limits_0^{u_{n-1}-u_{n-2}} B_{n-2}(u_{n-1} - t)\varphi(t)dt < 0.$$

Therefore, by the monotonicity of B_n and by Lemma 2, it is true that $u_{n-1} \le u_n$ and $v^* < v_n < u_n$.□

The following theorems can be proved in the same way.

Theorem 3. *For $(c_1, c_2) \in D_{II}$ it is true that:*
$$f_n'(x) = -c_1 + \begin{cases} A(x), & x < u_n; \\ A(x) + B_n(x), & x \ge u_n. \end{cases}$$
*We obtain $v_n(x) = x$ u $u_n(x) = \max(x, u_n)$. In addition, the sequence u_n nondecreasing. At the same time for all (c_1, c_2) such that $(c_1, c_2) \in \{r(p +$
$+ \sum\limits_{i=1}^{m-1} \alpha^i) \le c_2 \le r(p + \sum\limits_{i=1}^{m} \alpha^i)\}$ it follows that $u_n = -\infty \; \forall n \le m$, whereas u_{m+1} is defined by the equation $\alpha F(u_{m+1}) + \sum\limits_{i=1}^{m} \alpha^i F^{*(i+1)}(u_{m+1}) = \frac{r(p+\sum_{i=1}^{m}\alpha^i) - c_2}{h+r}$.*

Theorem 4. *For $(c_1, c_2) \in D_{III}$ we have:*
$$f_n'(x) = -c_1 + \begin{cases} 0, & x < v_n; \\ C_n(x), & x \ge v_n. \end{cases}$$
*Optimal values of orders are $v_n(x) = u_n(x) = \max(x, v_n)$. Besides, the sequence v_n is nondecreasing. Furthermore, assume that $(c_1, c_2) \in \{r(1 + \sum\limits_{i=1}^{m-1} \alpha^i) \le$
$c_1 \le r(1 + \sum\limits_{i=1}^{m} \alpha^i)\}$. In this case, $v_n = -\infty \; \forall n \le m$, and for v_{m+1} we get an equation $F(v_{m+1}) + \sum\limits_{i=1}^{m} \alpha^i F^{*(i+1)}(v_{m+1}) = \frac{r(1+\sum_{i=1}^{m}\alpha^i) - c_1}{h+r}$.*

Theorem 5. *For* $(c_1, c_2) \in D_{IV}$ *it is true that:*
$f_n'(x) = -c_1 + \begin{cases} 0, & x < v_n; \\ C_n(x), & x \geq v_n. \end{cases}$
There exists $k(c_1, c_2)$ *such that for any* $n < k$ *we obtain* $v_n(x) = u_n(x) = \max(x, v_n)$, *whereas for any* $n \geq k$ *we get* $v_n(x) = \max(x, v^*)$ *and* $u_n(x) = \max(x, u_n)$.

In order to analyze the sensitivity of the obtained results let us introduce Sobol's decomposition. Assume that $A = (A_1, ..., A_n)$ is uniformly distributed in $K^n = [0,1]^n$ and the function $g(a), a \in K^n$, is integrable. Put $g_0 = \mathbb{E}R = \int\limits_{K^n} g(a)da$, $g_i(a_i) = \int\limits_0^1 ... \int\limits_0^1 g(a) \prod\limits_{k \neq i} da_k - g_0$, $g_{i,j}(a_i, a_j) = \int\limits_0^1 ... \int\limits_0^1 g(a) \prod\limits_{k \neq i,j} da_k - (g_0 + g_i(a_i) + g_j(a_j))$,...
Then the following decomposition of variance holds for a square integrable random variable $R = g(A)$:

$$V[R] = \sum_{i=1}^{n} V_i + \sum_{i<j} V_{i,j} + \sum_{i<j<k} V_{i,j,k} + ... + V_{1,2,...,n},$$

where $V[R] = \int\limits_{K^n} g^2(a)da - g_0^2$ and partial variances are calculated by way of $V_{i_1,...,i_s} = \int\limits_0^1 ... \int\limits_0^1 g_{i_1,...,i_s}^2(a_{i_1}, ..., a_{i_s}) \prod\limits_{k=i_1,...,i_s} da_k$. Assuming $V[R] \neq 0$ we can formulate the following definition:
Sensitivity index $S_{i_1,...,i_s}$ for a group of parameters $(a_{i_1}, ..., a_{i_s}), 1 \leq i_1 < ... < i_s \leq n$, is given by $V_{i_1,...,i_s}/V[R]$, whereas the sensitivity index of order s is $\sum_{1 \leq i_1 < ... < i_s \leq n} S_{i_1,...,i_s}$. Moreover, global sensitivity index $GI(a_i)$ of parameter a_i is the sum os all indices $S_{i_1,...,i_s}, s \geq 1$, containing i

$$GI(a_i) = (V_i + \sum_{i \neq j} V_{i,j} + ... + V_{1,2,...,n})/V[R]$$

Thus, $GI(a_i)$ represents the total contribution of parameter a_i to the variance of output.
Applying this approach in Wolfram Mathematica we examined the sensitivity of the obtained solutions to small changes in parameters.

References

1.Bellman, R. (1957). *Dynamic Programming*, Princeton, New York: Princeton University Press.
2.Bulinskaya, E. (2007). Some aspects of decision making under uncertainty. Journal of Statistical Planning and Inference 137:2613–2632.
3.Bulinskaya, E. (2012). Optimal and asymptotically optimal control for some inventory models. In: A.N.Shiryaev et al. eds. *Prokhorov and Contemporary Probability Theory*, Springer Proceedings in Mathematics and Statistics, 33, chapter 8, Berlin: Springer-Verlag.

4.Caliskan-Demirag, O., Chen, Y. and Yang, Yi. (2012). Ordering policies for periodic-review inventory systems with quantity-dependent fixed costs. Operations Research 60:785–796.
5.Fox, E.J., Metters, R. and Semple, J. (2006). Optimal inventory policy with two suppliers. Operations Research 54:389–393.
6.Huggins, E.L., Olsen, T-L. (2010). Inventory control with generalized expediting. Operations Research 58:1414–1426.
7.Li, Q., Wu, X., Cheung, K.L. (2009). Optimal policies for inventory systems with separate delivery request and order-quantity decision. Operations Research 57:626–636.
8.Papachristos, S., Katsaros, A. (2008). A periodic-review inventory model in a fluctuating environment. IIE Transactions 40:356–366.
9.Sobol',I., (1990). Sensitivity estimayes for nonlinear mathematical models. Matematicheskoe Modelirovanie 2:112-118(in Russian).

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Stability Selection and Randomization in L_1 Quantile Regression

Sidi Zakari Ibrahim[1], Mkhadri Abdallah[2], and N'Guessan Assi[3]

[1] Department of Mathematics, Faculty of Sciences Semlalia ,Cadi Ayyad University, 40000, Marrakech, Morocco
(E-mail: i.sidizakari@uca.ma)
[2] Department of Mathematics, Faculty of Sciences Semlalia ,Cadi Ayyad University, 40000, Marrakech, Morocco
(E-mail: mkhadri@uca.ma)
[3] Department of Mathematics
University of Sciences and Technologies of Lille, 59650, Villeneuve d'Ascq, France
(E-mail: assi.nguessan@polytech-lille.fr)

Abstract Statistical models in linear regression generally focus on estimation and interpretation of conditional mean effects. However, in some situations considering mean effects could be not appropriate when for example we have great variations in response variable percentiles or when we have outliers. We here propose the Stability Selection method for variables selection in high dimension penalized linear Quantile Regression. This approach combines subsampling and variable selection algorithms adapted to the case of high dimension. Particularly, we apply Stability Selection with Lasso and Randomized Lasso Quantile Regression. Finally, the proposed method is compared with its competitors on simulated and real data sets.
Keywords: Quantile Regression, High Dimension, Resampling, Stability Selection.

1 Introduction

Meinshausen and Bühlmann[7] advocate that subsampling can be used for Stability Selection in penalized linear regression models to determine the amount of regularization such that a certain family type I error rate in multiple testing can be conservatively controlled for finite sample size. Particularly for complex and high dimensional problems, a finite sample control is much more valuable than an asymptotic statement with the number of observations tending to ∞. Moreover, the previous authors also prove that subsampling in conjunction with L_1-penalized estimation requires much weaker assumptions on the design matrix for asymptotically consistent variable selection than what is needed for the non-subsampled L_1-penalty scheme. Furthermore they show that additional improvements can be achieved by randomizing not only via subsampling but also in the selection process for the variables. Recently a variant approach called Complementary Pairs Stability Selection (CPSS) has been also proposed by Shah and Sameworth[2]. Beinrucker et al.[1] also propose a simple extension of the original stability feature selection approach used in Meinshausen and Bühlmann[7]. We can mention here that variable selection approaches based on bootstrap, Random Lasso[5] and Bolasso[10], have been proposed in linear regression case.

In the lines below, we will focus on the adaptation of Stability Selection[7] and bootstrap based approaches to Quantile Regression[17]. As a motivation of this work we have a dataset on volatile compounds previously used in Duflos et al.[6]. The classical approach based on penalized mean regression is not adapted for this setting since the response variable presents many variations. On the other hand, we want to select the set of stable variables among the volatile compounds which represent the predictors.

The second section of this paper is devoted to the adaptation of Stability Selection to Quantile Regression(QR). Moreover, an illustration on a real dataset is presented and the selection of the tuning parameter based on Belloni and Chernozhukov[3] idea is discussed in the same section. Finally, last section is devoted to numerical experiments including simulations and real data set applications.

2 Linear Quantile Regression Stability Selection

2.1 Variable selection in Quantile Regression

We consider a size n i.i.d sample $\{(\mathbf{x}_i, y_i), i = 1, ..., n\}$ from some unknown population, where $\mathbf{x}_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$. Linear quantile regression solves the following optimization problem for $0 < \tau < 1$:

$$(\hat{\beta}_0(\tau), \hat{\beta}(\tau)) = \text{argmin}_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} \left\{ \sum_{i=1}^{n} \rho_\tau(y_i - \beta_0 - \mathbf{x}_i^t\beta) \right\}. \quad (1)$$

The function $\rho_\tau(.)$ is called the check function and is defined by $\rho_\tau(u) = \tau u I_{u>0} + (\tau - 1)u I_{u\leq 0}$, where $I(.)$ is the indicator function which takes value 0 or 1. Since we are interested on variable selection, which is a common practice because in major studies, the main objective is to have a set of relevant variables in a set of predictors used to explain the response variable y.

The following penalized problem is considered:

$$(\hat{\beta}_0(\tau), \hat{\beta}(\tau)) = \text{argmin}_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} \left\{ \sum_{i=1}^{n} \rho_\tau(y_i - \beta_0 - \mathbf{x}_i^t\beta) + \lambda P(\beta) \right\}, \quad (2)$$

where $P(.)$ is the penalty function and the tuning parameter $\lambda > 0$ controls the sparsity of the model. As a survey on frequently used penalty functions in the field of quantile regression we can cite the excellent references of Zou and Yuan [11], Wu and Liu [9] and Slawski [8]. Since the idea of our proposed methods are based on Lasso penalty[14], in all the lines below, we will only focus on the following problem defined by:

$$\min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} \left\{ \sum_{i=1}^{n} \rho_\tau(y_i - \beta_0 - \mathbf{x}_i^t\beta) + \lambda \parallel \beta \parallel_1 \right\}. \quad (3)$$

All of the methods in this paper are based on the previous formulation except the Randomized Lasso[7] for quantile regression with weakness $\alpha \in (0, 1]$ which

takes the following form:

$$\min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} \left\{ \sum_{i=1}^{n} \rho_\tau(y_i - \beta_0 - \mathbf{x}_i^t\beta) + \lambda \sum_{k=1}^{p} \frac{\mid \beta_k \mid}{W_k} \right\}, \tag{4}$$

where W_k are iid random variables in $[\alpha, 1]$ for $k = 1, ...p$.

2.2 Stability Selection and pointwise control

This part is a slightly modified part of Meinshausen and Bühlmann[7] approach. Since stability paths are derived from the concept of regularization paths, we recall that for each quantile τ, $0 < \tau < 1$ a regularization path is given by the coefficient value of each variable over all regularization parameters

$$\{\hat{\beta}_k^\lambda(\tau); \lambda \in \Lambda, k = 1, ..., p\}.$$

Stability paths are the probability for each variable to be selected when randomly resampling from the data. For any given regularization parameter $\lambda \in \Lambda$, the selected set $\hat{S}_\tau^\lambda$ is implicitly a function of the samples $I = \{1, ..., n\}$.

Definition 1(selection probabilities): *Let I be a random subsample of $\{1, ..., n\}$ of size $\lfloor n/2 \rfloor$, drawn without replacement. For every set $K \subseteq \{1, ..., p\}$, the probability of being in the selected set $\hat{S}_\tau^\lambda(I)$ is*

$$\hat{\Pi}_K^\lambda(\tau) = P^*\{K \subseteq \hat{S}_\tau^\lambda(I)\}. \tag{5}$$

For every variable $k = 1, ..., p$, the stability path is given by the selection probabilities $\hat{\Pi}_K^\lambda(\tau)$, $\lambda \in \Lambda$.

In the remainder of the paper, we look at the selection probabilities of individual variables. Generally, for each quantile of interest variable selection is concerned by the choice of one element in the set of models

$$\{\hat{S}_\tau^\lambda; \lambda \in \Lambda\}, \tag{6}$$

where Λ is again the set of regularization parameters considered, which can be either continuous or discrete. There are typically two problems: first, the correct model S_τ might not be a member of set (6). Second, even if it is a member it is typically very difficult for high dimensional data to determine the right amount of regularization λ to select exactly S_τ, or at least a close approximation. With Stability Selection, we do not simply select one model in the list (6), instead the data are perturbed (e.g by subsampling) many times and we choose all variables that occur in a large fraction of the resulting selection sets.

Definition 2(stable variables). *For a cut-off π_{thr} with $0 < \pi_{thr} < 1$ and a set of regularization parameters Λ, the set of stable variables for a quantile τ is defined as*

$$\hat{S}_\tau^{stable} = \{k : max_{\lambda\in\Lambda}(\hat{\Pi}_k^\lambda(\tau)) \geq \pi_{thr}\}. \tag{7}$$

We keep variables with a high selection probability and discard those with low selection probabilities.

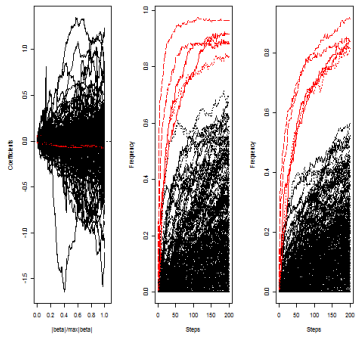


Figure1. From left to right: comparison between L_1 median regression regularization paths, QR Stability Selection without randomization and QR Randomized Stability Selection with $\alpha = 0.1$ on PAC dataset for log(y).

2.3 Illustration on PAC dataset

In Figure1, we illustrate the advantage of using Stability Selection with or without Randomization. We use PAC data available under R software about GC-retention indices of polycyclic aromatic compounds(y) which have been modeled by molecular descriptors(X). The data set contains n=209 observations and p=467 predictors. We first take the 50 variables with the highest marginal correlation with log(y) and randomly select five predictors. These five predictors are kept unpermuted and the remaining 462 are permuted across the samples, using the same permutation that keeps the dependence structure between the permuted observations intact. The left plot, corresponding to L_1 penalized median regression shows that it is very difficult to isolate the five unpermuted variables paths with noise variables paths. For Stability Selection, a threshold of $\pi_{thr} = 0.6$ includes all of the five unpermuted variables with some noise variables. When using Stability Selection with randomization parameter $\alpha = 0.1$, we can see that with the same threshold $\pi_{thr} = 0.6$ all of the five unpermuted variables are selected without any noise variable.

2.4 Tuning parameter selection

Among many alternatives on the choice of the tuning parameter we can cite (see Li and Zhu[12]) the Schwarz Information Criterion (Schwarz[16]; Koenker, Ng, and Portnoy[15]) (SIC) and the generalized approximate cross-validation criterion (Yuan[13]). The SIC is defined by

$$SIC(\lambda) = log(\frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y_i - f(\mathbf{x_i}))) + \frac{log(n)}{2n}df,$$

where df is a measure of the effective dimensionality of the fitted model. In our case, $f(\mathbf{x_i}) = \beta_0 + \mathbf{x}_i^t\beta$. However, recently Koenker[4] claims that using the SIC optimization method often produced insufficient shrinkage and the optimization

process was quite slow. He also claims that when simulating realizations of the random vector $S_n = \sum_{i=1}^{n}(\tau - I(U_i \leq \tau))\mathbf{x_i}$, one can assert that the event $\|S_n\|_\infty \leq \lambda$ should hold with high probability, provided of course that λ is chosen sensibly so that $\hat{\beta}$ is close to the true parameter $\beta(\tau)$. Following this idea, Belloni and Chernozukov[3] suggested choosing $\hat{\lambda}$ as a $(1-\alpha)$ quantile of the simulated distribution of $\|S_n\|_\infty$, or perhaps a constant multiple of such a quantile for some $c \in (1, 2]$. This extremely simple approach was used in our simulations for $\alpha = 0.1$ and $c = 1$.

3 Numerical results

3.1 Simulations settings

For our simulations, we consider the following model $Y = X\beta(0.5) + \epsilon$, where Y is the n dimensional response vector, X is the $n \times p$ predictors matrix, $\beta(\tau)$ is the true p dimensional parameter vector and ϵ is the n dimensional vector of errors. In the lines below, we consider $p = 200$-dimensional predictor variables follow an $N(0, \Sigma)$ distribution, where $\Sigma_{ij} = \rho^{|i-j|}$ and $\rho \in \{0, 0.5, 0.75, 0.9\}$. The sample size is fixed to $n = 100$ and the Signal to Noise Ratio $SNR = Var(X\beta(\tau))/Var(\epsilon)$ considered takes values in {0.5,2,4}. The error vector $\epsilon \sim N(0, 1)$ and $\beta(\tau)$ vector has s nonzero components chosen as uniforms on $[0, 1]$. The s value considered here is $s \in \{4, 8, 12, 20\}$.

Simulations are performed 100 times and median of False positive, False negative and the mean probability to select $0.1s$ and $0.4s$ correct variables without any noise variable are given on Figure 2 and Figure 3.
Each plot gives the performances of the following five methods "QR Stability Selection","QR Randomized Stability Selection", "QR Lasso", "QR Bolasso" and "QR Random Lasso".

Results for n=50 seems to be similar. For Stability Selection based approaches, the threshold is fixed to $\pi_{thr} = 0.6$ and $\alpha = 0.5$ for Randomized Stability Selection . We use fixed value of the tuning parameter(pointwise control) as previously advocated. Nevertheless we can use R software "lpRegPath" package which generates the entire solution path as a function of the tuning parameter. Taking into account the limited size for papers, we only present results for $s \in \{4, 8\}$.

For $s = 4$ and $\rho = 0$, stability selection based methods and QR Bolasso do not introduce FP ($SNR \in \{0.5, 2, 4\}$). QR Lasso FP selection is around 14 FP which represents the median about 100 bootstrap ($SNR \in \{0.5, 2, 4\}$). QR Random Lasso seems to introduce more FP variables for $SNR = 0.5$ than for $SNR \in \{2, 4\}$. In terms of FN, QR Randomized Stability Selection and QR Bolasso seems to delete more true regression coefficients than other methods with decreasing number of median FN when SNR increases. In terms of P($0.1s$ correct), QR Randomized Stability Selection and QR Bolasso give very good result (probability=1), QR Stability Selection corresponding probability increases when the SNR increases (0.73,0.77,0.78). Since QR Lasso introduces too many variables, the corresponding probability is zero. QR Random Lasso also have increasing probability with increasing SNR (0.02,0.27,0.41).

For P($0.4s$ correct) and $SNR = 0.5$, QR Stability Selection gives great probability(0.68) followed by QR Randomized Stability Selection (probability=0.46) and QR Bolasso (0.18). QR Lasso and QR Random Lasso give zero probability. For $SNR \in \{2, 4\}$, QR Randomized Lasso and QR Bolasso give great similar probability (0.99 and 1 respectively) and we have probability values of 0.01 and 0.1 for the QR Random Lasso. QR Stability selection gives probability of 0.89 and 0.86. The previous comments and for other values of ρ can be seen on Figure 2 and Figure 3.
When $s = 8$ and $\rho \in \{0, 0.5\}$ we have the same remarks as previous (s=4) for the median number of FP and FN. For P($0.1s$ correct), QR Randomized Stability and QR Bolasso have higher performances followed by QR Stability Selection and QR Random Lasso. For P($0.4s$ correct), QR Stability Selection has the highest performances due to the fact that it has the best trade off between FP and FN, followed by Randomized Stability Selection. We remark that QR Bolasso has the higher performance for P($0.1s$ correct) when $\rho = 0.75$ followed by QR Randomized Stability Selection and QR Stability Selection without Randomization. QR Random Lasso has very low selection probability. For P($0.4s$ correct) and SNR=0.5, all methods fail to select 40% of correct variables without introducing any noise variables, due to the fact that QR Lasso and QR Random Lasso introduce noise variables, and other methods delete some correct variables. Results for other values of SNR and for $\rho = 0.9$ can be seen on Figure2 and Figure3.

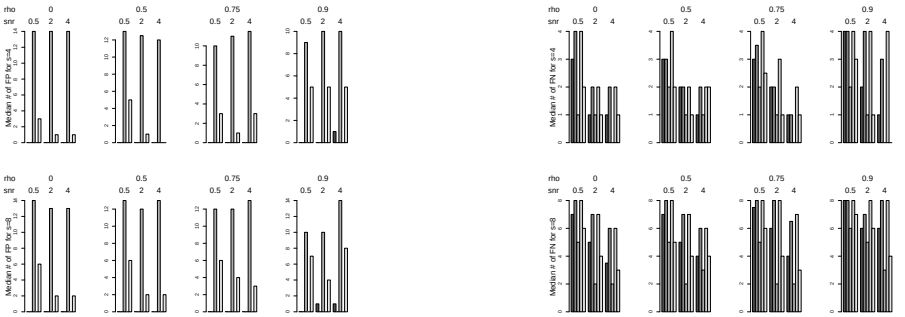


Figure2. Median number of False Positive and False Negative selection among 100 replications for s=4(top row) and s=8(bottom row). For each SNR value we have from left to right "QR Stability Selection","QR Randomized Stability Selection", "QR Lasso", "QR Bolasso" and "QR Random Lasso".

3.2 Real data application

We consider the data set on volatile compounds previously used in Duflos et al.[6]. The sample considered consists of $n = 37$ observations and $p = 49$ predictors(X) used to model Freshness index and Quality scores (y). A direct use of methods on the full data gives results only for L_1 median regression and QR Bolasso. For Stability Selection and bootstrap based methods, we have no results due to the fact that subsampling of size $n/2$ leads in some cases

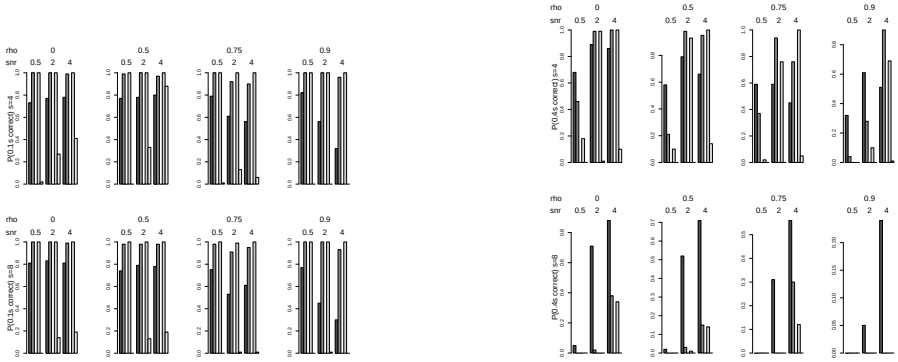


Figure3. Probability of selection of 0.1s and 0.4s of relevant variables without selection any noise variables among 100 replications for s=4(top row) and s=8(bottom row). For each SNR value we have from left to right "QR Stability Selection","QR Randomized Stability Selection", "QR Lasso", "QR Bolasso" and "QR Random Lasso".

to singular design sub matrixes and we cannot compute the tuning parameter $\lambda = c\Lambda(1-\alpha|X)$ since the matrix of predictors X is very sparse with many zero entries. On another hand, QR Bolasso with 100 bootstrap selects no variables for many values of λ, so we consider QR Soft Bolasso (QR SBolasso) which selects at least 60% of variables which are selected for all bootstrap samples. In the line below we only consider compounds with at least $n/2$ nonzero observations which lead to a predictors matrix with sizes $n = 37$ and $p = 37$. Selection results are summarized below where for the tuning parameter λ, $c = 2$ and $\alpha = 0.1$. For stability selection methods, the threshold is fixed to $\pi_{thr} = 0.6$. According to selection results in Table1, we see that for $\lambda = 25.253$ no variable is selected by all of the methods. Selection results for $\lambda/5$ and $\lambda/10$ are given in Table1 with selected volatiles names given in Table2. As expected, subsampling improves important variables selection and additional randomization prevents against noise variables selection. The selected compounds by Stability Selection approach also have been found by Duflos et al.[6] to be related to fish spoilage during storage. This important compounds are Ethanol(d), Ethyl acetate(o), 3-Methyl butanal(r) and 3-Methyl-1-butanol(ag) where more stable compounds seem to be Ethanol(d) and 3-Methyl-1-butanol(ag).

Methods	Selected Compounds	
λ=25.253	Freshness index	Quality scores
All methods	-	-
$\lambda/5$		
QR Lasso	a,d,l,o,q,r,aa,ag	b,d,r,aa,ag,ah
QR SBolasso	d,r,ag	d,ag,ah
QR Stability Selection	d,r,ag	d,ag
QR Randomized Stability Selection $\alpha = 0.2$	-	-
QR Randomized Stability Selection $\alpha = 0.5$	d,ag	ag
QR Randomized Stability Selection $\alpha = 0.8$	d,r,ag	d,ag
$\lambda/10$		
QR Lasso	b,d,j,l,o,r,aa,ag,aq,ar,bh	b,d,l,aa,ag,ah,aq,bh
QR SBolasso	d,o,r,ag,bh	d,r,ag
QR Stability Selection	d,o,r,ag	d,ag
QR Randomized Stability Selection $\alpha = 0.2$	d	d,ag
QR Randomized Stability Selection $\alpha = 0.5$	d,r,ag	d,ag
QR Randomized Stability Selection $\alpha = 0.8$	d,r,ag	d,ag

Table1. Selected volatiles for penalized median regression.

Compounds	Names	Compounds	Names
a	Acetaldehyde	r	3-Methyl butanal
b	Methanethiol	aa	2,3-Pentanedione
d	Ethanol	ag	3-Methyl-1-butanol
j	2,3-Butanedione	ah	2-Methyl-1-butanol
l	2-Butanone	aq	1-Hexanol
o	Ethyl acetate	ar	3-Heptanone
q	2-Methyl-1-propanol	bh	Nonanal

Table2. Selected volatiles names.

References

1.A. Beinrucker, U. Dogan, and G. Blanchard. A Simple Extension of Stability Feature Selection.*Pattern Recognition Lecture Notes in Computer Science*, 7476, 256-265, 2012.

2.R. Shah and R. Samworth. Variable selection with error control: Another look at stability selection. *J. Roy. Statist. Soc., Ser. B* (to appear), 2012.

3.A. Belloni and V. Chernozhukov. L_1-penalized quantile regression in high-dimensional sparse models. *Annals of Statistics*, 39, 1, 82-130, 2011.

4.R. Koenker. Additive models for quantile regression: Model selection and confidence bandaids. *Brazilian Journal of Probability and Statistics*, 25, 3, 239-262, 2011.

5.S. Wang, B. Nan, S. Rosset and J. Zhu. Random Lasso. *Annals of Applied Statistics*, 5, 1, 468-485, 2011.

6.G. Duflos, F. Leduc, A. N'Guessan, F. Krezewinski, K. Ossarath and P. Malle. Freshness characterisation of whiting (Merlangius merlangus) using an SPME/GC/PM method and a statistical multivariate approach. *J Sci Food Agric*, 90, 2568-2575, 2010.

7.N. Meinshausen and P. Buhlmann. Stability selection. *Journal of the Royal statistical Society, B.*, 72, 4, 417-473, 2010.

8.M. Slawski. The structured elastic net for quantile regression and support vector classification. *Statistics and Computing*, (to appear), 2010.

9.Y. Wu and Y. Liu. Variable selection in quantile regression. *Statistica Sinica*, 19, 801-817, 2009.

10.F. R. Bach. Bolasso: Model Consistent Lasso Estimation through the Bootstrap. *Proceedings of the 25 th International Conference on Machine Learning, Helsinki, Finland*, 2008.

11.H. Zou and M. Yuan. Regularized simultaneous model selection in multiple quantiles regression. *Computational Statistics and Data Analysis*, 52, 5296-5304, 2008.

12.Y. Li and J. Zhu. L_1-Norm Quantile Regression. Journal of Computational and Graphical Statistics, 17, 1, 1-23, 2008.

13.M. Yuan. GACV for Quantile Smoothing Splines. *Computational Statistics and Data Analysis*, 5, 813-829, 2006.

14.R. Tibshirani. Regression shrinkage and selection via the Lasso. *journal of the Royal statistical Society, B.*, 58, 267-288, 1996.

15.R. Koenker, P. Ng and S. Portnoy. Quantiles smoothing splines.*Biometrika*, 81, 4, 673-80, 1994.

16.G.Schwarz. Estimating the Dimension of a Model. *The annals of Statistics*, 6, 461-464, 1978.

17.R. Koenker and G. Bassett. Regression Quantiles. *Econometrica*, Econometric Society, 46, 1, 33-50, 1978.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Alternative Assessments of the Probability of Death with a Case Study for Persons with Celiac Disease in Selected East European Countries

Simpach Ondrej

Department of Demography, Faculty of Informatics and Statistics, University of Economics in Prague, Winston Churchill sq. 4, 130 67, Prague, Czech Republic (E-mail: ondrej.simpach@vse.cz)

Abstract. The probability of death depended in the past to a considerable extent on the level of advancement of the health service, the medical findings acquired and knowledge of the appropriate treatment processes. In the case of persons with Celiac Disease, which is a disease involving gluten intolerance, the hope of survival in the majority of countries was slim until the eighties of last century. These people died at a very young age thanks to ignorance of the diagnosis of their disease. However, as soon as it was possible to determine the diagnosis of Celiac Disease correctly there was a considerable breakthrough and progress rapidly changed the hope of survival for these people. This breakthrough occurred earlier in some countries and later in others. In this way treatment procedures were found for hitherto unknown diseases, or at least there was information on reducing the consequences of these diseases. The submitted study will provide a look at the alternative assessment of the probability of death of persons with Celiac Disease and the probability of death in general. The modelling of the probability of death is possible with the use of the LOGIT model. On the basis of supplementary information about the population it is then possible to construct various probability scenarios with the utilisation of alternative variables.
Keywords: Probability of Death, Celiac Disease, LOGIT, Alternative Assessment.

1 Introduction

In spite of the fact that medicine is constantly bringing people new information and the diagnosis of new diseases, there were and still are diseases for which the existing diagnosis is only partial and thus insufficient for the complete cure of the patients (Logan et al. [4]). In the second half of last century the diagnosis began to appear in some countries of a disease involving gluten intolerance, later described as Celiac Disease. This diagnosis, however, only spread to certain countries. There were countries not only in Europe, but also throughout the world, which did not have all the necessary information from science and research published abroad (Rubio-Tapia et al. [5]). This caused various information delays and the consequences of insufficient information about the diagnoses of certain diseases had an impact on the life expectancy of these people. The probability of the death of persons with specific diseases was thus raised in comparison with the probability of death of persons in the general population not burdened by any of the diseases with a still insufficient diagnosis.

2 Methodology and Model

The dependence on age of the probability of death of a person x years old can be explained with the use of further variables, both discrete and also categorical (Freese and Long [1]). The LOGIT models are capable of estimating, with the use of the distribution function of logistic distribution, the value of the probability of death of a person x years old, where further supplementary information may create various forms of the probability function. In the presented model the probability of death of a person x years old will be estimated for the course of the next k years after the medical examination (where k is any whole number) in the case where the person has some diagnosed disease, or in the case where the person is completely healthy. In the illustration case study the analysis will be used of the probability of death of persons with Celiac Disease in comparison with the probability of death for the population as a whole. The explained variables of Y will be alternative. When the value of variable Y equals 1, then the person will die within k years, and on the contrary when the value of variable Y equals 0, the person will survive k years. So that it would be possible also to determine the values of the probability of the occurrence of this phenomenon between the two extremes, the LOGIT model of discrete selection will be applied, when the explained variables acquire values from the interval <0 ; 1> (Hoyos et al. [2]). The following variables may be used for the model:
- AGE is the precise current age of the person invited for a health check,
- CIRD is the Constant of Increased Risk of Death, which acquires values from the interval $< l ; h >$, where l and h are whole numbers. The calculation of this constant arises for the i^{th} patient from Table 1, which is created during the medical examination and where instead of the verbal replies given there were recorded $w_{i,j}$, acquiring the values 0 and 1, where 0 = patient's reply does not coincide with the word given in the appropriate square and 1 = patient's reply coincides with the word given in the appropriate square.

	$\mathbf{V}_1$	$\mathbf{V}_2$	$\mathbf{V}_3$	$\mathbf{V}_1$	$\mathbf{V}_2$	$\mathbf{V}_3$
Smoker	no	occasionally	regularly	$v_{i,j} = 0/1$	...	...
Black Coffee	no	occasionally	regularly	...		
Alcohol	no	occasionally	regularly	...		
Sleep	regular	irregular	poor	...		
Nutrition	regular	irregular	poor	...		

Table 1. Replies of patients to doctor's questions during general examination (left) and 0/1 matrix replies (right)

From Table 1, in which the replies are recorded in the 0/1 format, emerges the CIRD for the i^{th} patient from the formula (1),

$$CIRD = (w_1 \times \sum_{i=1}^{5} v_{i,1}) + (w_2 \times \sum_{i=1}^{5} v_{i,2}) + (w_3 \times \sum_{i=1}^{5} v_{i,3}) \qquad (1)$$

where w_1, w_2 and w_3 are the weights recommended on the basis of the doctor's opinion, who provided data matrices for analysis. (We can use $w_1 = 1$, $w_2 = 3.5$ and $w_3 = 7$). The general rule, arising from the literature, is not here. This Constant can take values from interval <5 ; 35>, where the extreme value of 5 means, that the patient does not increase the risk of death because of its poor lifestyle and extreme value of 35 means, that the patient increases the risk of death in the worst way possible.
- ILL is a binary variable, acquiring the values 0 = the person does not have a diagnosed illness, or 1 = the person has a diagnosed illness. In the model with Celiac Disease the variable CEL will be used.
- DEATH_K is a binary variable, acquiring the values 0 = the person did not die within k years after the medical examination, or 1 = the person died within k years after the medical examination.

The probability function for the LOGIT model (Christensen [3]) is

$$P_i = E(Y = 1|\mathbf{X}_i) = \frac{1}{1 + e^{-(b_0 + \mathbf{b}^{'}\mathbf{x}_i)}} \tag{2}$$

modified for this study in the form

$$P_i = E(Y = 1|\mathbf{X}_i) = \frac{1}{1 + e^{-(b_0 + b_1 AGE_i + b_2 CIRD_i + b_3 CEL_i)}}, \tag{3}$$

where i is the i^{th} patient. Let us set

$$Z_i = b_0 + \mathbf{b}^{'}\mathbf{x}_i \tag{4}$$

and let us insert it for the purposes of this study

$$Z_i = b_0 + b_1 AGE_i + b_2 CIRD_i + b_3 CEL_i. \tag{5}$$

The subsequent expression

$$P_i = \frac{1}{1 + e^{-Z_i}} = \frac{e^{Z_i}}{1 + e^{Z_i}} = F(Z_i) \tag{6}$$

is the distribution function of the logistic distribution. The probability that a person aged x-years will not die within k years after the moment of the medical examination is

$$1 - P_i = \frac{1}{1 + e^{Z_i}} \tag{7}$$

and therefore

$$\frac{P_i}{1 - P_i} = e^{Z_i}. \tag{8}$$

By calculating the logarithm we obtain LOGIT

$$ln(\frac{P_i}{1 - P_i}) = Z_i = b_0 + \mathbf{b}^{'}\mathbf{x}_i, \tag{9}$$

which is transferred for the purposes of this study into the form

$$ln(\frac{P_i}{1 - P_i}) = Z_i = b_0 + b_1 AGE_i + b_2 CIRD_i + b_3 CEL_i. \tag{10}$$

From the general assumptions, the logarithm of the credibility function

$$lnL(b_0, \mathbf{b}) = \sum_{i=1}^{N} [Y_i ln(\frac{e^{Z_i}}{1+e^{Z_i}}) + (1-Y_i) ln(1 - \frac{e^{Z_i}}{1+e^{Z_i}})] \tag{11}$$

there arises after the substitution

$$lnL(b_0, \mathbf{b}) = \sum_{i=1}^{N} [Y_i ln(\frac{e^{b_0+\mathbf{b}'\mathbf{x}_i}}{1+e^{b_0+\mathbf{b}'\mathbf{x}_i}}) + (1-Y_i) ln(1 - \frac{e^{b_0+\mathbf{b}'\mathbf{x}_i}}{1+e^{b_0+\mathbf{b}'\mathbf{x}_i}})] \tag{12}$$

and for the purposes of this study is

$$lnL(b_0, b_1, b_2, b_3,) = \sum_{i=1}^{N} [Y_i ln(\frac{e^{b_0+b_1 AGE_i+b_2 CIRD_i+b_3 CEL_i}}{1+e^{b_0+b_1 AGE_i+b_2 CIRD_i+b_3 CEL_i}}) +$$

$$+(1-Y_i) ln(1 - \frac{e^{b_0+b_1 AGE_i+b_2 CIRD_i+b_3 CEL_i}}{1+e^{b_0+b_1 AGE_i+b_2 CIRD_i+b_3 CEL_i}})]. \tag{13}$$

3 Data, Material and Case Study

For the study mentioned it is possible to use data from the databases of health insurance companies and medical statistics. There are few health insurance companies which record events to the necessary extent. Practical analysis will be carried out for selected periods of the nineties in the Czech Republic, Slovakia and Poland. The analysis will be restricted to persons with Celiac Disease and persons with no health complications and the results will be published separately for the male and the female gender. For the experiment of non-linear regression (Spector and Mazzeo [6] or Yang and Raehsler [7]), applied in the first part of this study about 200 observations of variables consisting of two samples were obtained for each country - Czech Rep., Slovakia and Poland. It is important to note, that this is not a representative selection for the application of standard methods of mathematical statistics. The selection was not taken at random. This is the data matrix, obtained by tentative minor research. Selection consists all individual invited in 1990 to general medical examination and their health status was checked in the future. For consecutive experiment of non-linear regression, applied in the second part of the study, approximately other 200 observations of patients, consisting of two samples were obtained for each country. It is a selection of patients invited in 1995 to the overall medical examination and their health status was checked in the future (but obtained from other sources than the first selection). We hope that there is minimum probability that some patients from the first sample are contained in the second sample. Estimating the unknown parameters of non-linear regression models is no problem today. To estimate the parameters of LOGIT model Statgraphics Centurion XVI version 16.1.11 and Gretl 1.8.7 build 2010-01-24 were used.
Based on the methodology showed above the estimates of unknown parameters of LOGIT models for males in 1990 and 1995 as well as for females in 1990 and

1995 were calculated for Czech Rep., Slovakia and Poland. Table 2 shows the results for the Czech Republic (top), for Slovakia (middle) and Poland (bottom). The first model for each country is always for males in 1990, the second model always for females in 1990, a third model always for males in 1995 and the fourth model always for females in 1995. Of the estimated models were constructed graphs showing the development of the probability of death of x-years old person. (See Figure 1 for Czech Rep., Figure 2 for Slovakia and finally Figure 3 for Poland.

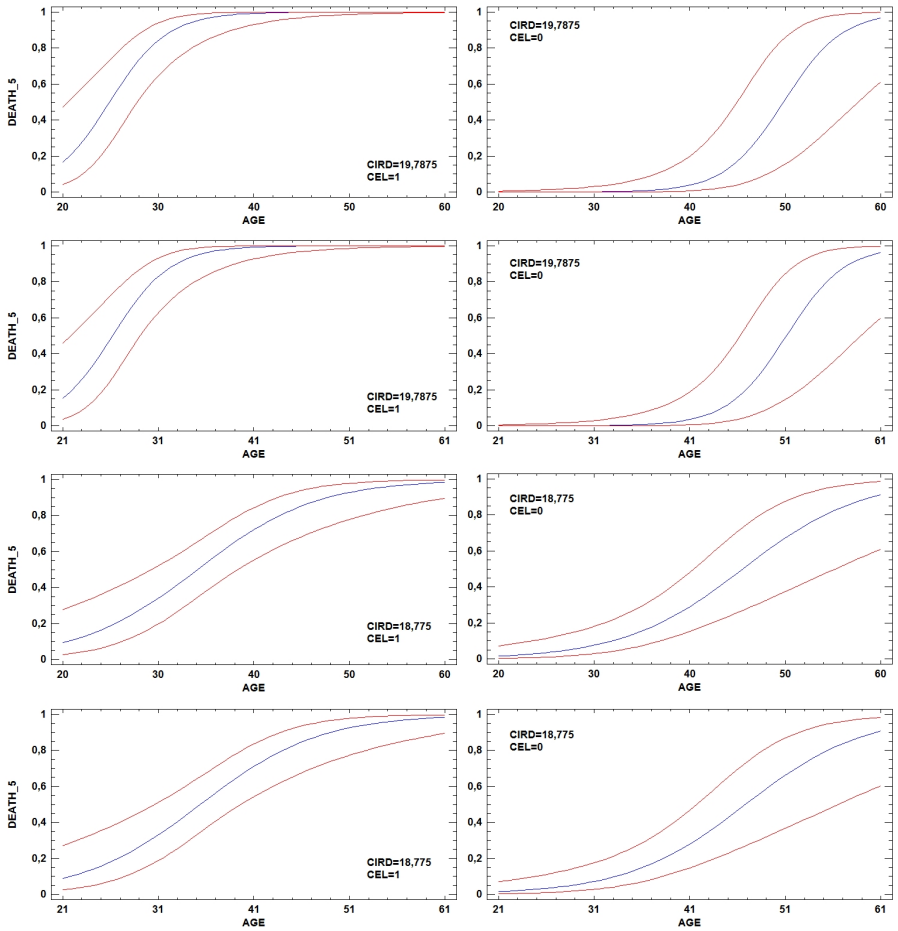


Fig. 1. Probability of death of x-years old person in the Czech Republic (males 1990 with CEL - 1^{st} row left, males 1990 without CEL - 1^{st} row right, females 1990 with CEL - 2^{nd} row left, females 1990 without CEL - 2^{nd} row right, males 1995 with CEL - 3^{rd} row left, males 1995 without CEL - 3^{rd} row right, females 1995 with CEL - 4^{th} row left and females 1995 without CEL - 4^{th} row right.)

Parameter	Estimate	St. Error	Odds Ratio	Factor	Chi-Sq.	DF	P-Value
Constant	-15,6423	3,71635					
AGE	0,333211	0,09654	1,40052	AGE	31,3889	1	0,0000
CIRD	0,370041	0,08211	1,40563	CIRD	39,0002	1	0,0000
CEL-0	-8,23231	1,90601	0,00029	CEL	78,8484	1	0,0000
Constant	-15,7816	3,71599					
AGE	0,329121	0,08534	1,38975	AGE	31,2349	1	0,0000
CIRD	0,361625	0,08578	1,43566	CIRD	38,8651	1	0,0000
CEL-0	-8,19201	1,89216	0,00027	CEL	78,8339	1	0,0000
Constant	-9,17489	1,90414					
AGE	0,161341	0,03913	1,17509	AGE	21,591	1	0,0000
CIRD	0,195672	0,04715	1,21613	CIRD	22,1363	1	0,0000
CEL-0	-1,84446	0,52223	0,15811	CEL	14,2464	1	0,0002
Constant	-9,26566	1,91888					
AGE	0,189633	0,04001	1,18655	AGE	22,0001	1	0,0000
CIRD	0,201122	0,04023	1,22366	CIRD	22,1963	1	0,0000
CEL-0	-1,83663	0,52889	0,16889	CEL	14,6398	1	0,0001
Constant	-14,771	3,61432					
AGE	0,319513	0,08661	1,37646	AGE	29,5255	1	0,0000
CIRD	0,337395	0,08146	1,40129	CIRD	36,9356	1	0,0000
CEL-0	-7,98723	1,89016	0,00033	CEL	76,0443	1	0,0000
Constant	-11,177	2,47924					
AGE	0,221478	0,05806	1,24792	AGE	22,5223	1	0,0000
CIRD	0,27303	0,06150	1,31394	CIRD	33,842	1	0,0000
CEL-0	-5,64065	1,16173	0,00355	CEL	64,0071	1	0,0000
Constant	-7,16711	1,60754					
AGE	0,133789	0,03569	1,14315	AGE	16,7686	1	0,0000
CIRD	0,146313	0,04042	1,15756	CIRD	15,711	1	0,0001
CEL-0	-1,90989	0,50245	0,14809	CEL	16,8345	1	0,0000
Constant	-7,05805	1,62331					
AGE	0,119955	0,03438	1,12745	AGE	14,1924	1	0,0002
CIRD	0,15557	0,04191	1,16832	CIRD	16,4657	1	0,0000
CEL-0	-1,87092	0,49519	0,15398	CEL	16,4205	1	0,0001
Constant	-10,565	2,35537					
AGE	0,23943	0,05954	1,27053	AGE	28,3616	1	0,0000
CIRD	0,2161	0,05382	1,24123	CIRD	24,6238	1	0,0000
CEL-0	-5,30962	1,12046	0,00494	CEL	59,8911	1	0,0000
Constant	-8,55789	1,92926					
AGE	0,163137	0,04635	1,1772	AGE	16,7961	1	0,0000
CIRD	0,217977	0,04990	1,24356	CIRD	28,4366	1	0,0000
CEL-0	-4,27303	0,84652	0,01393	CEL	51,4505	1	0,0000
Constant	-5,64467	1,42814					
AGE	0,103277	0,03250	1,1088	AGE	11,2856	1	0,0008
CIRD	0,122385	0,03745	1,13019	CIRD	12,1086	1	0,0005
CEL-0	-1,80914	0,47412	0,16379	CEL	16,6206	1	0,0000
Constant	-7,11323	1,63255					
AGE	0,123955	0,04111	1,13222	AGE	14,2396	1	0,0000
CIRD	0,23357	0,04263	1,17888	CIRD	16,4756	1	0,0000
CEL-0	-1,89992	0,50122	0,14536	CEL	16,4322	1	0,0000

Table 2. Estimations of unknown LOGIT models parameters

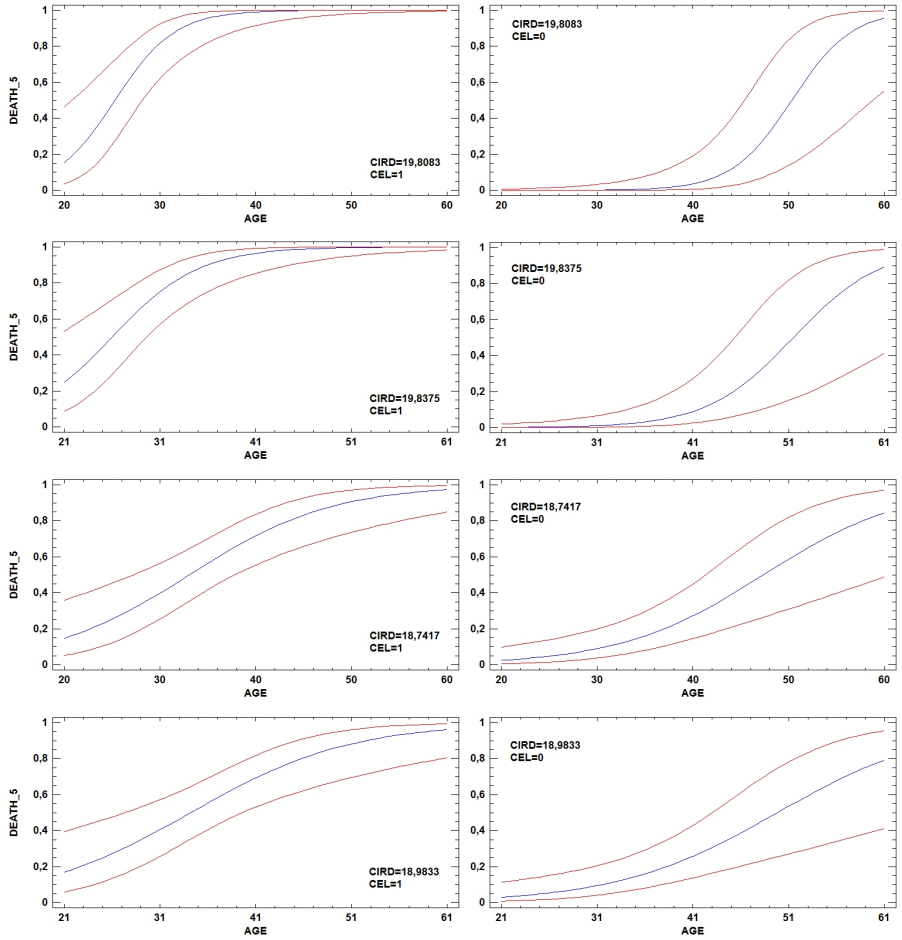


Fig. 2. Probability of death of x-years old person in Slovakia (see legend in FIG. 1.)

4 Conclusion

The aim of this study was to analyse the probability of death of x-year old persons in Czech Rep., Slovakia and Poland during next five years ($k = 5$) after the general medical examination in 1990 and 1995. The analyses were solved using LOGIT models and tried to confirm the hypothesis claiming, that the probability of death of x-year old person suffering from celiac disease decreased few years after the gaining of another medical knowledge from other countries. Even if some assumptions for the application of methods of mathematical statistics are broken, it is possible to say, that the key hypothesis was confirmed. Looking at Fig. 1, 2 and 3 we can see only slight differences between the presented countries. Their development of the compared statistics in the past should be similar.

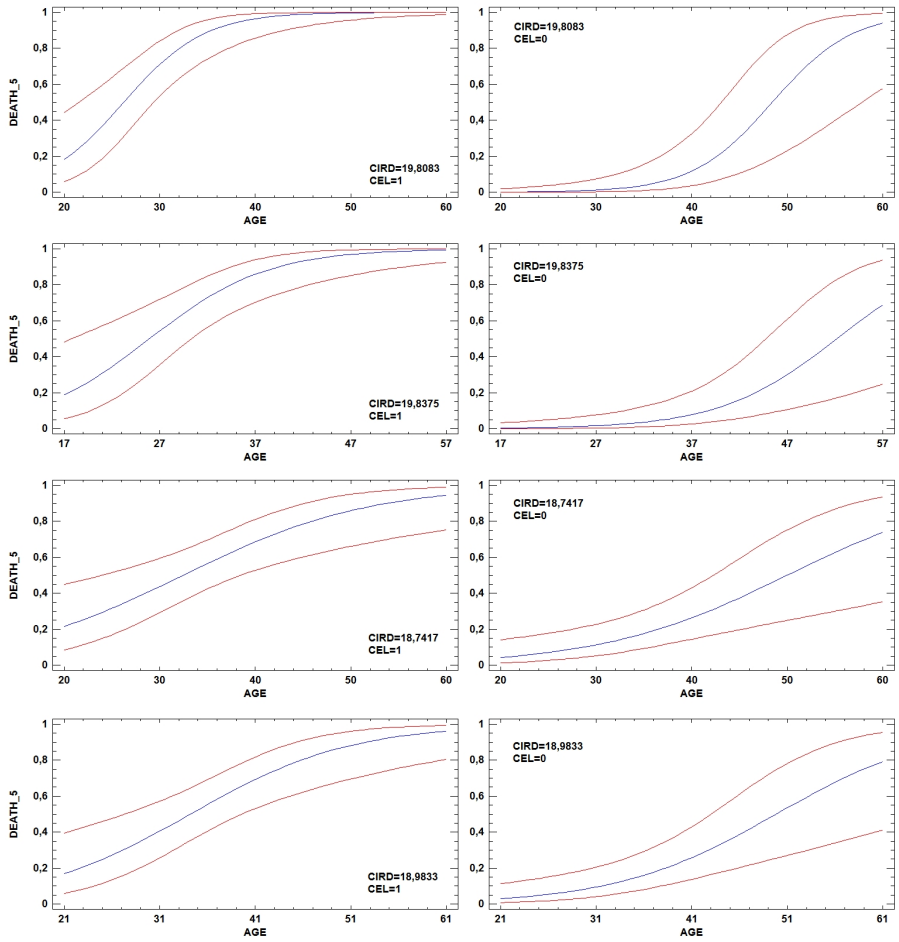


Fig. 3. Probability of death of x-years old person in Poland (see legend in FIG. 1.)

References

1.J. Freese and J. S. Long. Regression Models for Categorical Dependent Variables Using Stata, *College Station: Stata Press,* 2006.
2.D. Hoyos, P. Mariel and J. Meyerhoff. Comparing the performance of different approaches to deal with attribute non-attendance in discrete choice experiments: a simulation experiment, *BILTOKI 201001*, Universidad del Pais Vasco, 2010.
3.R. Christensen. Log-Linear Models *Springer-Verlag*, New York, 1990.
4.R. F. Logan, E. A. Rifkind, I.D. Turner and A. Ferguson. Mortality in celiac disease. *Gastroenterology*, 97(2), 265–271, 1989.
5.A. Rubio-Tapia et al. Increased Prevalence and Mortality in Undiagnosed Celiac Disease. *Gastroenterology*; 137 : 88–93. 2009.
6.L. C. Spector and M. Mazzeo. Probit Analysis and Economic Education, *Journal of Economic Education*. Spring, 11, pp. 37-44. 1980.
7.CH. W. Yang and R. D. Raehsler. An Economic Analysis on Intermediate Microeconomics: An Ordered PROBIT Model, *Journal for Economic Educators*, Volume 5, No. 3, Fall. 2005.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

High frequency trading with Hidden Markov Models by using clustering and PPCA algorithms

I. Róbert Sipos[1], János Levendovszky[2]

[1] Department of Networked Systems and Services, Budapest University of Technology and Economics, Budapest, Hungary (e-mail: siposr@hit.bme.hu)
[2] Department of Networked Systems and Services, Budapest University of Technology and Economics, Budapest, Hungary (e-mail: levendov@hit.bme.hu)

Abstract. In this paper a Hidden Markov Model (HMM) based prediction algorithm will be introduced for algorithmic trading, the performance of which is enhanced by clustering and PPCA algorithms. The performance of the new method is tested on different financial assets and instruments. We use various training methods (e.g. Baum-Welch expectation maximization, simulated annealing and hybrid methods) for optimizing the parameters of HMM in order to capture the underlying characteristics of the financial time series. The new hybrid algorithm combines simulated annealing (SA) with the Baum Welch algorithm (used as a local search after each step of SA) and can provide relatively fast and good quality solutions. The real time nature of learning can be guaranteed by running SA only for a limited number steps determined by a predefined time interval. To cope with the underlying complexity and perform high frequency trading we apply clustering for reducing the number of data and PPCA for dimension reduction as preprocessings. The algorithms are tested on US SWAP rates and FOREX series. The results demonstrate that a good average return can be obtained by HMM based prediction and the applied preprocessings. It is noteworthy that the continuous model gives better result, however it requires more complex training.

Keywords: Hidden Markov Models, financial time series, algorithmic trading

1 Introduction

In this paper we investigate the performance of prediction based trading on financial time series by Hidden Markov Models (HMM). HMMs are widely used for predictions (Mamon and Elliott[15], Durbin *et al.*[5], Jurafsky and Martin[13], Hassan and Nath[7]), however there are no efficient algorithms have yet been developed for real-time training of the free parameters for the purpose of high-frequency algorithmic trading (Hassan *et al.*[8]). The trading action is based on the probabilities of the predicted future values. For minimizing the risk, the highest probability future value is selected and depending its increasing (or decreasing) nature a buying (or short-selling) action is taken. In order to optimize the parameters of the HMM, different training methods are used: Baum-Welch expectation maximization (Baum *et al.*[3]), simulated annealing (Kirkpatrick *et al.*[14]) and hybrid methods proposed by the present paper. In the hybrid method, after accepting a new state by SA the Baum-Welch algorithm is used as a local search method. After the convergence of the Baum-Welch algorithm to a state in the parameter space, SA adopts it as the new state and calculates its next step from there. The real-time nature of this

algorithm can be ensured by limiting the number of steps of SA making it fit into a predefined time interval. In this way, good quality solutions can be achieved in relatively short time.
Unfortunately, the algorithms in the present form are very exhaustive computationally. Thus, to ease the complexity, we introduce clustering to limit the number of data and PPCA to reduce the dimension.
The HMM based predictor is tested on US SWAP rates and FOREX series. The results demonstrate that a good average return can be obtained. It is noteworthy that the continuous model gives better result, however it requires more complex training.
The material is treated in the following order:

- in section 2, the discrete and continuous HMM is briefly summarized;
- in section 3, the computational model is mapped out;
- in section 4, we delve into the different learning algorithms;
- in section 5 and 6 we explain the preprocessing with clustering and PPCA, respectively;
- in section 7 the trading method is outlined;
- in section 8, the numerical results are given;
- finally, in section 9, some conclusions are drawn.

2 The model

A Hidden Markov Model (HMM) (Baum and Petrie[2]) is a statistical model which is an extension of Markov chains. In this model, the current state is no longer directly visible to the observer, but each state emits an observable output quantity denoted by

$$\mathbf{X} = \{o_1, o_2, ..., o_T\}, \tag{2.1}$$

and each emission depends only on the hidden state

$$\mathbf{Q} = \{q_1, q_2, ..., q_T\}. \tag{2.2}$$

The probability of an emitting a specific output is determined by the conditional probabilities $P(o_t \mid \mathbf{Q}) = P(o_t \mid q(t) = q_t)$, while the transition probabilities of the underlying Markov chain, describing the jumping probabilities from one state to another is given by the transition probability matrix $A_{ij} = P\left(q(t+1) = q_i \,\middle|\, q(t) = q_j\right)$. $\boldsymbol{\pi}_N$ denotes the initial distribution vector.

HMMs are commonly used in various fields, for instance in bioinformatics (Durbin *et al.*[5]) or speech recognition (Jurafsky and Martin[13], Rabiner[17]). In our case we use them to predict future values of financial time series.

2.1 Description of a discrete HMM

If we treat the observations having discrete values over a given alphabet

$$o_t \in \{k_1, k_2, ..., k_M\}, \tag{2.3}$$

then the observation probability matrix, $\mathbf{B}_{NxM}$ describes the probability distribution over the possible values for each state. For continuous data, pre-processing is needed to quantize the data to the discrete alphabet. The easiest

alphabet has a three-element code indicating that the asset price is {increasing, stagnating, decreasing}. In the discrete case, HMM is described by

$$\Theta = \{\boldsymbol{\pi}, \mathbf{A}, \mathbf{B}\}. \quad (2.4)$$

Vector valued observations can be coded as n-tuples, note that in this case the alphabet size will grow according to $O(M^n)$.

2.2 Description of a continuous HMM

When the observable output is continuous, the observation probabilities are described by a probability density function, instead of a probability matrix. For this purpose a multivariate Gaussian mixture model (Dasgupta[4]) was used, in which the density function is composed as a weighted sum (according to the $\boldsymbol{W}$ weight matrix) of M independent Gaussian functions:

$$P(o_t \mid q(t) = q_j) = \sum_{k=1}^{M} w_{jk} b_{jk}(o_t), \quad (2.5)$$

where $b_{jk}(o_t) \sim N(\boldsymbol{\mu}_{jk}, \boldsymbol{\Sigma}_{jk})$. Note that this approach can handle multivariate observations as well. In the continuous case, HMM is then described as

$$\Theta = \{\boldsymbol{\pi}, \mathbf{A}, \mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}. \quad (2.6)$$

3 The computational approach to trading

The model parameters of HMM are identified during the training phase based on a set of observations in historical data sequences. After learning, the HMM can predict the values of the newly observed sequence, and then the predicted distribution is converted to a trading signal for taking the appropriate trading actions.

Our computational framework is shown on the following structural block diagram (Fig. 1) and detailed as follows:

- Training phase: HMM model fitting (by using one of the learning algorithms described in section 4) based on a training set;
- Prediction: using the identified model and a newly observed sequence of the time series, in this step the most likely prediction for the future asset prices is calculated (see chapter 7);
- Trading strategy forming the trading signal: mapping the prediction into a trading action (also detailed in chapter 7);
- Performance analysis: a framework for trading and testing and evaluating various numerical indicators for the sake of comparison of the profitability of different methods (chapter 8 contains further details).

This framework will be extended in chapter 6 with dimensional reduction.

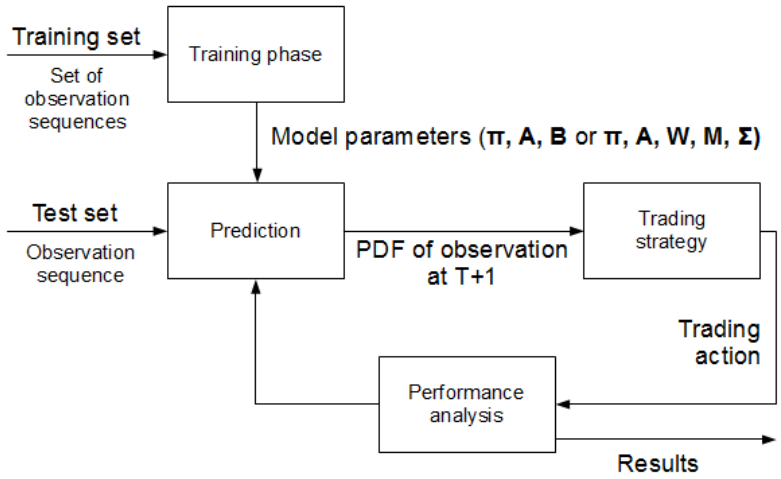


Fig. 1. Computational approach

4 Training for predicting future values of financial time series

The model parameter estimation (learning) is a key aspect when we are using HMMs for prediction. First, the construction of a training set is needed. This is carried out by sampling the input time series with a sliding window of length *T*:

$$\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_K\}, \tag{4.1}$$

where $\mathbf{X}_t = \{\mathbf{r}_{t-T+1}, ..., \mathbf{r}_t\}$ and $\mathbf{r}_t$ is the daily return at the *t*th time instance:

$$r_{t,i} = \frac{p_{t,i} - p_{t-1,i}}{p_{t-1,i}}; \mathbf{r}_t = \{r_{t,i}, i = 1, ..., n\}, \tag{4.2}$$

where $p_{t,i}$ denotes the price of the *i*th asset in the *t*th time instance.

In the discrete case, the daily returns were partitioned into *M* equal sized subsets. While handling the daily returns as a continuous value required a normalization step:

$$r^*_{t,i} = \frac{r_{t,i} - \bar{\mathbf{r}}^{(i)}}{std(\mathbf{r}^{(i)})}. \tag{4.3}$$

During training, the likelihood of the model is maximized based on the given observations (Rabiner and Juang[16]):

$$\Theta_{opt} = \arg\max_{\Theta} P(\mathbf{X} \mid \Theta). \tag{4.4}$$

To avoid numerical instabilities due to the small order of magnitude of such probabilities, it is better to calculate the logarithm of the likelihood:

$$L(\Theta) = \log P(\mathbf{X} \mid \Theta). \tag{4.5}$$

The optimization problem described in (4.4) can be solved in multiple ways.

4.1 Baum-Welch expectation maximization

The Baum-Welch expectation maximization (EM) algorithm (Baum *et al.*[3]) is a mechanism to iteratively update the model ((2.4) or (2.6)) starting from an arbitrary initial value and iterating until the likelihood of the model converges to a certain value. Since this is an iterative method, which can use the forward-backward algorithm, implemented in an efficient way by dynamic programming (Rabiner[17]), this algorithm is relatively fast.
On the other hand, it may get stuck in one of the local minima making the final result highly dependent on the initialization. In our implementation, these matrices were initialized randomly.

4.2 Simulated annealing

In the absence of analytical solutions for finding the global optimum of the likelihood of model parameters, one can use simulated annealing (SA) to obtain good quality heuristic solutions. Simulated annealing (Kirkpatrick *et al.*[14]) is a stochastic search method for finding the global optimum in a large search space. In this context the energy function $J(\Theta)$ is the log-likelihood of the selected model:

$$J(\Theta) = L(\Theta). \tag{4.6}$$

Let Θ be an arbitrarily initialized model, and then by calling random number generation a model Θ' is generated subject to its constraints. The neighbor function on each iteration modifies one of the probability vectors or matrices with an amount according the current temperature (*T*). Accept the new model if $J(\Theta') > J(\Theta)$, or otherwise with $e^{-\frac{J(\Theta)-J(\Theta')}{T}}$ probability. Continue the sampling while decreasing the *T* until zero. The last state is now describes the identified model.
This method, in theory, can provide us the best fitting model. It is a rather slow method due to the large dimension of the search space.

4.3 Hybrid solution

Having the Baum-Welch EM algorithm and the SA at hand, one can construct a hybrid solution. In this case, after a new state was accepted during the annealing process, we ran the Baum-Welch as a local optimization algorithm to speed up the convergence. After the Baum-Welch algorithm converged to a state in the parameter space, SA adopts it as the new initial state and calculates its next move from there. The real-time nature of this algorithm can be ensured by limiting the number of steps of SA. In this way, good quality solutions can be achieved in relatively short time. This approach can bring more reliable results, which are less dependent on the initialization.

5 Clustering algorithms

As we have seen, optimizing model likelihood in the full space of Θ including every available observation is computationally exhausting. To ease the underlying computational complexity, one can pre-process the training set and

form clusters out of the observations. Having the clusters at hand and assigning them to the corresponding hidden states, the p.d.f. describing the observation probabilities can be estimated for each cluster, separately. (Note that the obtained model parameters could also be used as an initial estimate for the previously shown methods.)
Rabiner addressed this topic in his work (Rabiner[17]) suggesting the following algorithm in continuous case:

Algorithm 1.:

1. Arbitrary initialization of model parameters.
2. Segment the observations in the training set based on the optimal state sequence (Viterbi path) calculated by the Viterbi algorithm (Viterbi[20], Forney[6]) and assign each data point to the corresponding maximum likelihood hidden states.
3. Perform k-means clustering over the observation vectors within each state resulting *M* clusters per state.
4. Re-estimate the model parameters in a standard manner (calculate the relative frequency of observations in each state and the relative frequency of transitions, estimate sample mean and covariance for each cluster).
5. Calculate statistical similarity with the previous model, if we can assume convergence then stop, continue from step 2 otherwise.

It has been demonstrated in (Rabiner[17]) that the algorithm above performs well in the field of speech recognition, giving essentially identical likelihood values in an order of magnitude faster than the model obtained by the Baum-Welch algorithm. However, attempts of application for financial time series shown in (Idvall and Jonsson[11]) has yielded much more moderate performance, although the importance of a proper clustering mechanism is emphasized. This gives the motivation for our new approach to obtain HMM model parameters by clustering.
The main idea was to avoid a separate and linear clustering phase if a non-linear GMM is used to describe the observation p.d.f. for the hidden states. This leads to the following algorithm:

Algorithm 2.:

1. Fit a Gaussian mixture distribution using every data point in the training set.
2. Assign each mixture component to a corresponding hidden state, index the price vectors accordingly.
3. Calculate the transition probabilities and the initial distribution vector from relative frequencies (note that the remaining model parameters, namely the p.d.f. of the observations, are already estimated during the first step).

One may note, that in the proposed algorithm only one single Gaussian component will belong to each state. This will not restrict generality, as a model having *N* hidden states with *M* mixture components can be rewritten in the form of *N* times *M* states with only one component. Clearly, a grouping step could take place in the algorithm forming two phase clustering similar to the previous one. However, if the model parameter estimation is not used for further iteration in the EM or SA algorithm, then the increased state space does not cause difficulties in terms of computational resources. Using only one component per state also gives us the advantage to control the model degree by tuning only one free parameter.

The performance of the algorithm can be further enhanced by also taking into consideration the price vector in the next time instance besides the current one during clustering, as data points having similar distribution can differ in this aspect. In this case we need to perform the Gaussian mixture estimation (step 1.) on the following data set:

$$\boldsymbol{\tau} = \left\{ \left\{ \mathbf{r}_i, \mathbf{r}_{i+1} \right\}, 0 \le i < K \right\}, \tag{5.1}$$

which embed the return vectors into a *2n* dimensional space. After the clustering is done, in step 2, we need to consider only the top-left quarter of the matrices describing the p.d.f., which belongs to the first *n* dimensions.

6 Dimension reduction

Transforming the data from a higher dimensional space into a space of fewer dimensions as a pre-processing step for the HMM can also yield certain advantages:

- a smaller dimensional space requires much less computational time, which enables us to do higher frequency trading, or to involve a larger number of assets, which was not feasible formerly;
- allows us to control the model degree in order to avoid overfitting.

Deploying a feature extraction algorithm into to computational framework introduced in chapter 3 can be done in a straightforward manner.

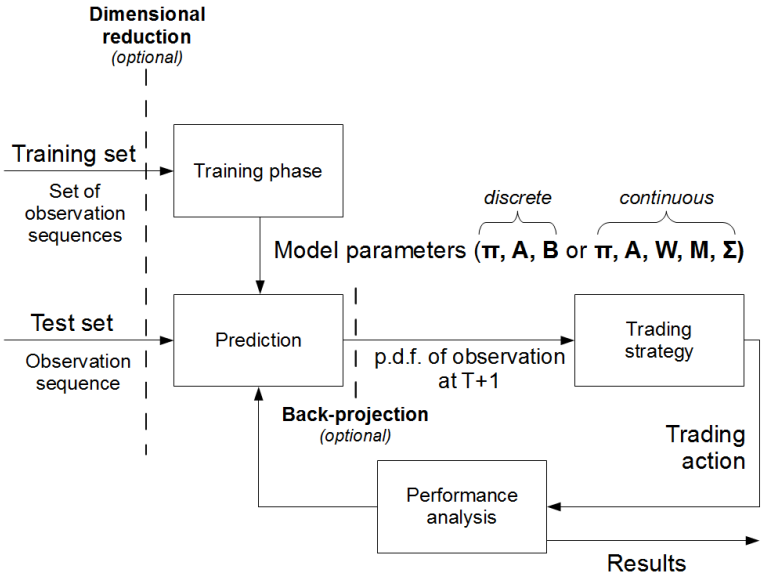


Fig. 2. Extended computational approach

As it is shown in figure 2, the dimensional reduction (as in (6.2)) and the back-projection to the original space (will be given in (6.1)) is simply precedes and follows the training and prediction modules, while the trading strategy is the same as before.

The goal of dimension reduction is to relate a *d*-dimensional return vector ($\mathbf{r}$) into a *q*-dimensional ($q < d$) latent vector denoted by $\mathbf{x}$. If we assume a linear relationship with non-zero mean, we get the basic model of factor analysis (Bartholomew[1]):

$$\mathbf{r} = \mathbf{W}\mathbf{x} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}, \tag{6.1}$$

where matrix $\mathbf{W}_{d\times q}$ describes the transformation between the two spaces, $\boldsymbol{\mu}$ is the mean and $\boldsymbol{\varepsilon}$ represents the noise in the model. Having the mapping matrix and the mean of the data, the reduction can be formulated as

$$\mathbf{x} = (\mathbf{r} - \boldsymbol{\mu})\left(\mathbf{W}^T\right)^{-1}. \tag{6.2}$$

6.1 Probabilistic principal component analysis

Principal component analysis (PCA) is a well-known technique for dimension reduction in various applications (Jolliffe[12]), however, the lack of associated probabilistic model called for a derivation, called probabilistic PCA (PPCA), having a proper density-estimation framework (Tipping and Bishop[18]).

If we assume isotropic Gaussian noise in (6.1) with $\boldsymbol{\varepsilon} \sim N\left(0, \sigma^2\mathbf{I}\right)$, that leads us to the following conditional probability distribution (Tipping and Bishop[18]):

$$\mathbf{r} \mid \mathbf{x} \sim N\left(\mathbf{W}\mathbf{x}+\boldsymbol{\mu}, \sigma^2\mathbf{I}\right). \tag{6.3}$$

By convention, the marginal distribution for the latent variables is Gaussian with

$$\mathbf{x} \sim N\left(0, \mathbf{I}\right), \tag{6.4}$$

causing the marginal distribution for the observed data to be

$$\mathbf{r} \sim N\left(\boldsymbol{\mu}, \mathbf{W}\mathbf{W}^T+\sigma^2\mathbf{I}\right). \tag{6.5}$$

To obtain the maximum-likelihood estimator for $\mathbf{W}$ and σ^2, in the lack of a closed-form solution, we used an iterative expectation-maximization (EM) algorithm to estimate their values (Tipping and Bishop[18]). The EM algorithm consists of the following two steps:

$$\tilde{\mathbf{W}} = \mathbf{S}\mathbf{W}\left(\sigma^2\mathbf{I}+\mathbf{M}^{-1}\mathbf{W}^T\mathbf{S}\mathbf{W}\right)^{-1} \tag{6.6}$$

and

$$\tilde{\sigma}^2 = \frac{1}{d} tr\left(\mathbf{S} - \mathbf{S}\mathbf{W}\mathbf{M}^{-1}\tilde{\mathbf{W}}^T\right), \tag{6.7}$$

where $\mathbf{M} = \mathbf{W}^T\mathbf{W}+\sigma^2\mathbf{I}$, $\mathbf{S}$ is the covariance matrix

$$\mathbf{S} = \frac{1}{K}\sum_{i=1}^{K}\left(\mathbf{r}_i-\boldsymbol{\mu}\right)\left(\mathbf{r}_i-\boldsymbol{\mu}\right)^T, \tag{6.8}$$

while $\tilde{\mathbf{W}}$ and $\tilde{\sigma}$ denote the new value of the related parameters. When the iteration judged to have converged, one should perform an orthogonalization on the matrix $\mathbf{W}$ as well.

7 Prediction based trading

After the training phase has been done, the model is able to predict future stock prices based on a window of observed previous data denoted by ***X***. The forward algorithm (Rabiner[17]) can be used to calculate the forward values, the conditional probabilities being in each hidden state:

$$\alpha_t\left(i\right) = P\left(\mathbf{X} \mid \Theta, q(t) = q_i\right). \tag{7.1}$$

The forward algorithm calculates these values in a memory- and running time efficient manner by using a dynamic programming table. In our case, we are interested only in the probability vector belonging to the last state: $\boldsymbol{\alpha} = \left\{\alpha_T\left(i\right), i = 1, ..., N\right\}$. Having this at hand, one can formulate the probability density functions belonging to observations in the next step as follows:

- discrete case: $\boldsymbol{\omega} = \boldsymbol{\alpha}\mathbf{A}\mathbf{B}$, where $\omega_i = P\left(o_{T+1} = k_i\right)$;
- continuous case:

$$\xi \sim N\left(\sum_{i=1}^{N}\left(\boldsymbol{\alpha}\mathbf{A}\right)_i\sum_{j=1}^{M}w_{ij}\boldsymbol{\mu}_{ij}, \sum_{i=1}^{N}\left(\boldsymbol{\alpha}\mathbf{A}\right)_i\sum_{j=1}^{M}w_{ij}\boldsymbol{\Sigma}_{ij}\right), \tag{7.2}$$

and $P\left(\xi = \mathbf{y}\right) = P\left(\mathbf{o}_{T+1} = \mathbf{y}\right)$.

These p.d.f.-s are our predictions for the future asset prices. In each step, the asset with the highest probability of increasing (or decreasing) is selected which determines the trading action, whether to buy or shortsell for the sake of maximizing the profit. In the discrete case, the trading action is determined by the movement coded by different states, we can de-quantize it as $E(\xi) = \boldsymbol{\omega}^T \mathbf{v}$ accordingly, where $\boldsymbol{v}$ denotes the quantization vector used to map the alphabet (in the case of vector data, represented as n-tuples, an additional decoding step is needed to sum up the probabilities belonging to each asset). While, in the continuous case, a de-normalization step is needed as per (4.3). Thus, the trading signal is described as:

$$i := \arg\max_i \left| E(\xi_i) \right|; \begin{matrix} E(\xi_i) > 0 \rightarrow buy \\ E(\xi_i) < 0 \rightarrow shortsell \end{matrix}. \qquad (7.3)$$

8 Simulation results and performance analysis

An extensive back-testing framework was created to handle trading actions on various input data sets and provide numerical results for comparison of different methods on different financial data series. In this section we show the numerical results obtained on SWAP and FOREX mid-prices.

For performance analysis we used the following data sets:

- U.S. SWAP rates (between August 2008 and August 2010 in daily resolution);
- FOREX rates (EUR/USD, GBP/USD, AUD/USD, NZD/USD, USD/CHF, USD/CAD between December 2009 and 2011 in daily resolution).

The training set consisted of the first one year long period of daily returns, while the tests were performed on the second year of data.

For the sake of comparison the following performance measures were calculated for each simulation, where c_t denotes the sum of owned cash and the market value of the owned portfolio at time instance t, while c_0 denotes the initial cash (in each case the agent started with \$10,000): (i) minimal value $G_{\min} = \frac{1}{c_0} \min_{0 \le t \le T} c_t$; (ii) final value $G_{final} = \frac{c_T}{c_0}$; (iii) maximal value $G_{\max} = \frac{1}{c_0} \max_{0 \le t \le T} c_t$; (iv) average value $G_{avg} = \frac{1}{c_0} \frac{1}{T} \sum_{t=1}^{T} c_t$.

8.1 Comparison of different training methods

Both in the training set and during the test phase a sliding window of 8 days were used (T=8). In the continuous case, 3 hidden states (N=3) were used with the mixture of 3 uncorrelated Gaussian functions (M=3). Unfortunately, there is no analytical method known to find the optimal value for these parameters (Hassan *et al.*[8]). During the simulations, using the discrete representation of

observations, it turned out that trading SWAP rates is more profitable with a larger number of hidden states and a smaller alphabet size (during the simulations N=4 and M=2 were used), while on FOREX data a longer alphabet size together with less hidden states proved to be more favorable (we used N=2 and M=5 respectively).
In this period, the U.S. SWAP rates had a decreasing tendency, at the end of the year they are worth only 60.73% of their initial prices on average. The bar charts (Fig. 3 and Fig. 4) show that all of the introduced methods beat this tendency, and in the scenario when SA was used to train the continuous mode HMM the trading was profitable with a 108.52% yearly profit.
The FOREX rates in this period showed only a slight increase during the year (0.45%), our methods achieved up to 26.62% profit (Fig. 4).

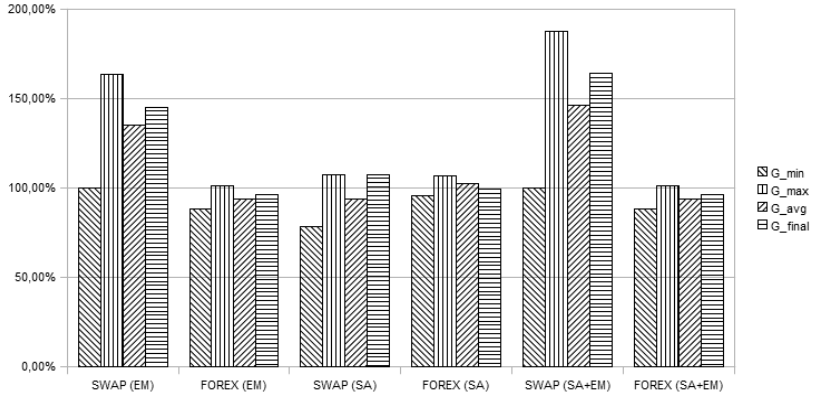


Fig. 3. Trading results in discrete case

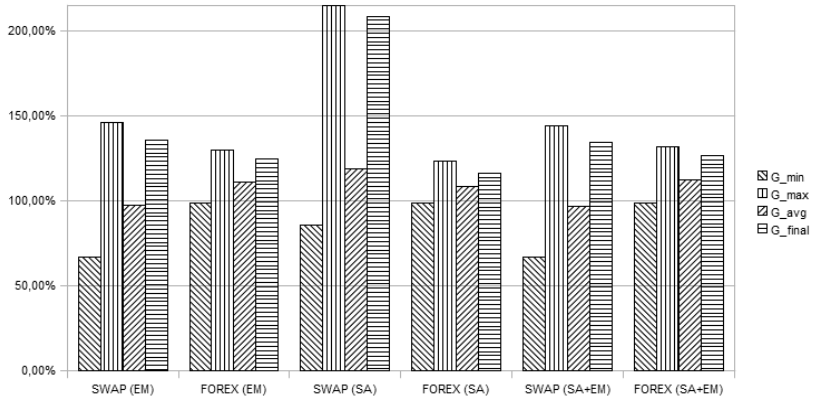


Fig. 4. Trading results in continuous case

As one can see, the novel HMM optimization methods outperform the traditional EM algorithm in most scenarios.

During the training phase, an EM step takes around *900ms*, while one step of SA consumes *250ms* using the continuous training sets on an Intel M330@2.13GHz processor. The prediction based trading step is quasi real-time. The hybrid approach speeds up the convergence with one order of magnitude.

8.2 Results obtained with clustering and PPCA

During clustering 8 clusters were formed in each case (N=8, M=1) with a shared covariance matrix. On the following bar charts `CLUST` denotes the case when clustering were applied as described in algorithm 2, while `CLUST (2)` refers to its enhanced version, where the next states are also taken into consideration. In the case of SWAP data, we investigated the event when the number of dimensions were reduced to 3 by PPCA from the original 8 assets. Similarly, the dimensions of FOREX data were reduced to 4 from 6.

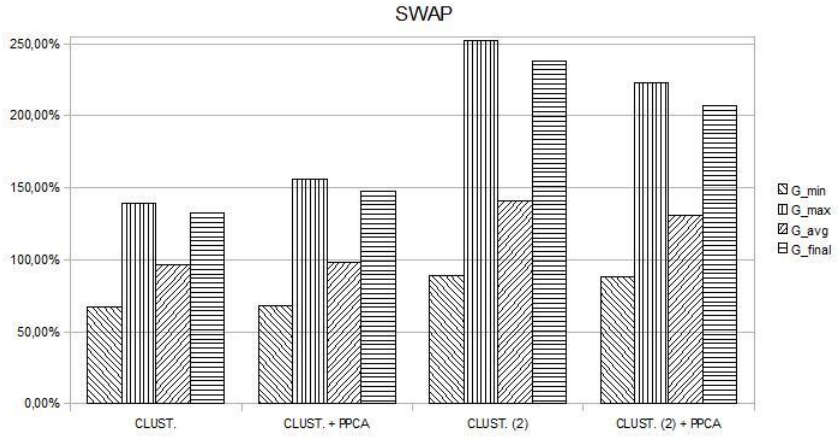


Fig. 5. Trading results on SWAP using clustering and PPCA

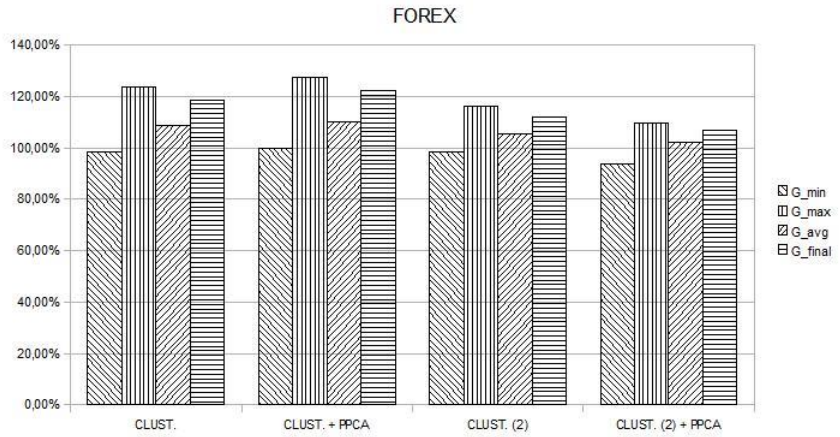


Fig. 6. Trading results on FOREX using clustering and PPCA

As it is shown in figure 5 and 6, these methods were also proven to be profitable on the tested time series, outperforming the market tendencies. The numerical results show that applying PPCA is beneficial with normal clustering, and less profitable with the enhanced version.
In comparison with the results shown in section 8.1, the range of the realized profits are comparable, reaching up to total 138.28% yearly profit on SWAP. The applied preprocessing steps yielded over one order of magnitude speed up compared to the previous methods.

9 Conclusion

In this paper we have proposed novel approaches for training HMMs and using them for predicting future values of financial time series. The proposed trading algorithms proved to be profitable in real scenarios. Using the continuous representation yields better trading results, however, its training time is substantially longer than the discrete case. We used three learning methods: (i) Baum-Welch EM; (ii) SA; and (iii) a hybrid solution. The performance analysis demonstrated that a better learning algorithm could increase trading efficiency and profit compared to the traditional learning strategies. The newly introduced clustering and PPCA methods can further speed up the algorithms giving rise to high frequency trading applications.
As a direction for future research, besides single asset price prediction, HMMs could be used for portfolios, obtained from an arbitrary portfolio optimization method, as well.

Acknowledgement

This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013).

References

1. D. J. Bartholomew. Latent variable models and factor analysis. Charles Griffin & Co. Ltd., London, 1987.
2. L. E. Baum and T. Petrie. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. The Annals of Mathematical Statistics 37 (6): 1554–1563, 1966.
3. L. E. Baum, T. Petrie, G. Soules and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann. Math. Statist., vol. 41, no. 1, pp. 164–171, 1970.
4. S. Dasgupta. Learning Mixtures of Gaussians. Proceedings of Symposium on Foundations of Computer Science (FOCS), 1999.
5. R. Durbin, S. R. Eddy, A. Krogh and G. Mitchison. Biological sequence analysis, Cambridge University Press, 1998.
6. G. D. Forney. The Viterbi algorithm. Proc. IEEE, vol. 61, pp. 268-278, Mar. 1973.
7. R. Hassan and B. Nath. Stock Market Forecasting Using Hidden Markov Model: A New Approach. Proceedings of the 2005 5th International Conference on Intelligent Systems Design and Applications.

8. R. Hassan, B. Nath and M. Kirley. A fusion model of HMM, ANN and GA for stock market forecasting. Expert Systems with Applications 33 (2007) 171–180.
9. M. Hassoun. Fundamentals of artificial neural networks. MIT Press, 1995.
10. S. Haykin. Neural Networks a comprehensive foundation. MacMillan 2004.
11. P. Idvall and C. Jonsson. Algorithmic Trading - Hidden Markov Models on Foreign Exchange Data. Master's Thesis, Department of Mathematics, Linköpings Universitet, 2008.
12. I. T. Jolliffe. Principal Component Analysis. Springer-Verlag, New York, 1986.
13. D. Jurafsky and J. H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing. Computational Linguistics, and Speech Recognition, 2nd Edition, Pearson Prentice Hall, London, 2006.
14. S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi. Optimization by Simulated Annealing. Science, 220, pp. 671-680, 1983.
15. R. S. Mamon and R. J. Elliott. Hidden Markov Models in Finance, Springer Science Business Media, 2007.
16. L. R. Rabiner and B. H. Juang. An introduction to hidden Markov models. IEEE ASSP Magazine 3:4-16, 1986.
17. L. R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE, 77 (2), p. 257–286, February 1989.
18. M. Tipping, and C. Bishop. Probabilistic principal component analysis. Journal of the Royal Statistical Society: Series B, 61, part 3, pp. 611-622, 1999.
19. T. L. Tri, H. N. Van, N. T. Thi, V. P. Quoc and K. D. Xuan. A Normal - Hidden Markov Model in forecasting stock prices, 2011.
20. A. J. Viterbi. Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. IEEE Trans. Informat. Theroy, vol. IT-13, pp. 260-269, Apr. 1967.
21. K. F. C. Yiu, Y. Liu and K. L. Teo. A Hybrid Descent Method for Global Optimization. Journal of Global Optimization 28: 229–238, 2004.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Numerical approximation of solutions of stochastic differential equations driven by multifractional Brownian motion

Anna Soós[1]

Faculty of Mathematics and Computer Science, Babes-Bolyai University, Cluj Napoca, Romania
(E-mail: asoos@math.ubbcluj.ro)

Abstract. We study the numerical approximation of Ito stochastic differential equations driven by multifractional Brownian motion.

We consider the following stochastic differential equation driven by multifractional Brownian motion

$$X(t) = X_0 + \int_0^t F(X(s), s)ds + \int_0^t G(X(s), s)dB(s),\ t \in [0, T]. \tag{1}$$

We assume that with probability 1 we have $F \in C(\mathbb{R}^n \times [0,T], \mathbb{R}^n), G \in C^1(\mathbb{R}^n \times [0,T], \mathbb{R}^n)$ and for each $t \in [0,T]$ the functions $F(\cdot, t), \frac{\partial G(\cdot,t)}{\partial x}, \frac{\partial G(\cdot,t)}{\partial t}$ are locally Lipschitz. B is a multifractional Brownian motion.

The equation (1) will be approximated for each $N \in \mathbb{N}$ through

$$X_N(t) = X_0 + \int_0^t F(X_N(s), s)ds + \int_0^t G(X_N(s), s)dB_N(s). \tag{2}$$

We will show that the equation (2) has a local solution, which converges in probability to the solution of (1) in the interval, where the solutions exist. We use power series expansions for multifractional Brownian motion.
Keywords: multifractional Brownian motion, stochastic differential equations, series expansion.

1 Introduction

The notion of fractional Brownian motion (fBm) was introduced by Kolmogorov in 1940. Self-similarity, long-range dependence, and smoothness of the sample paths make fBm a useful tool in modelling natural phenomena. A constant Hurst parameter is too rigid for some applications, for example in finance and turbulence. Therefore different generalizations of fBm have been introduced. The multifractional Brownian motion (mfBm) was proposed in [11] and [2] replacing the Hurst parameter H of fBm by a scaling function $t \to H(t)$.

The fractional Brownian motion (fBm) with Hurst index $H \in (0,1)$ is a zero mean Gaussian random process $\Big(B(t)\Big)_{t\geq 0}$ with continuous sample paths and with covariance function

$$E\Big(B(s)B(t)\Big) = \frac{1}{2}\Big(t^{2H} + s^{2H} - |s-t|^{2H}\Big).$$

For $H = \frac{1}{2}$ the fractional Brownian motion is the ordinary standard Brownian motion.

The fractional Brownian motion B has on any finite interval $[0, T]$ Hölder continuous paths with exponent $\gamma \in (0, H)$ (see [5]). Moreover, the quadratic variation on $[a, b] \subseteq [0, T]$ is

$$\lim_{|\Delta_n| \to 0} \sum_{i=1}^{n} \left(B(t_i^n) - B(t_{i-1}^n) \right)^2 = \begin{cases} \infty & \text{if } H < \frac{1}{2}, \\ b - a & \text{if } H = \frac{1}{2}, \\ 0 & \text{if } H > \frac{1}{2}, \end{cases} \tag{3}$$

where $\Delta_n = (a = t_0^n < \ldots < t_n^n = b)$ is a partition of $[a, b]$ with $|\Delta_n| = \max_{1 \le i \le n} (t_i^n - t_{i-1}^n)$.

The multifractional Brownian motion (mfBm) is obtained by replacing the constant parameter H of the fractional Brownian motion by a smooth enough functional parameter $H(\cdot)$. We denote by H a function defined on the real line and with values in a fixed interval $[a, b] \subset (0, 1)$. We assume that it is uniformly Hölder continuous of order $\beta > b$ on each compact subset of $\mathbb{R}$. For example let H piecewise constant function $H : R \to]0, 1[$

$$H(t) = \sum_{i=0}^{k} a_i \mathbf{1}_{[\tau_i, \tau_{i+1}[}(t)$$

where $\tau_0 = -\infty, \tau_{k+1} = \infty$ and $\tau_1, \tau_2, \ldots \tau_k$ is an increasing finite sequence of real numbers.

Investigations concerning stochastic differential equations driven by a fractional Brownian motion or a more general fractional process have been done by L. Coutin and L. Decreusefond [3], L. Coutin and Z. Qian [4], M.L. Kleptsyna, P.E. Kloeden and V.V. Anh [7], F. Klingenhöfer and M. Zähle [8], M. Zähle [16], [17] and many others. These studies were motivated by the problems in mathematical finance, internet traffic, biology, hydrology etc. The main difficulty raised by the fractional Brownian motion and the processes related to it, is that they are not Markovian, even more, they are not even semimartingales. Hence a new approach to stochastic fractional calculus was developed. There exist several ways to define the stochastic integral pathwise and related techniques, Dirichlet forms, anticipating techniques using Malliavin calculus and Skorohod integration (e.g. [15], [5]). In this paper we use the approach of M. Zähle [15], based on the ideas of Lebesgue-Stieltjes integrals and fractional calculus [12] for fractional Brownian motion.

The multifractional Brownian motion $\Big(B(t)\Big)_{t \in [0,1]}$ with Hurst index H can be approximated using the series expansion given in [13].

For $\nu \neq -1, -2, \ldots$ **the Bessel function J_ν of the first type of order** ν is defined on the region $\{z \in \mathbb{C} : |\arg z| < \pi\}$ as the absolutely convergent sum

$$J_\nu(z) = \sum_{k=0}^{\infty} \frac{(-1)^k}{\Gamma(k+1)\Gamma(\nu + k + 1)} \left(\frac{z}{2} \right)^{\nu + 2k}.$$

It is known that for $\nu > -1$ the function J_ν has a countable number of real, positive simple zeros (see [14]). Let $x_1(t) < x_2(t) < \dots$ be the positive, real zeros of $J_{-H(t)}$ and let $y_1(t) < y_2(t) < \dots$ be the positive, real zeros of $J_{1-H(t)}$.

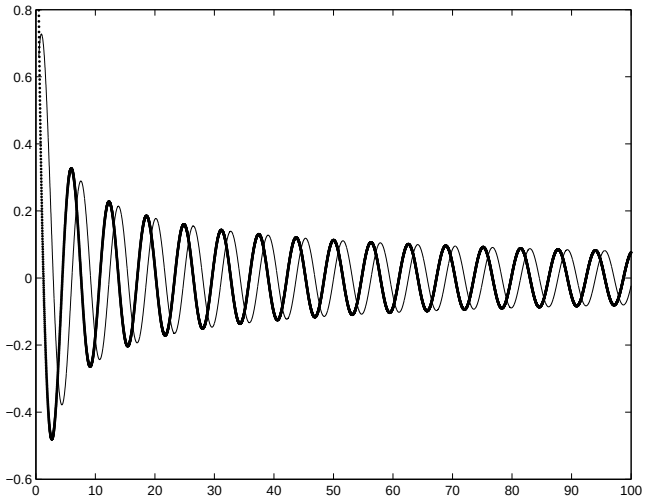


Fig. 1. Bessel functions: J_{-H} (with '·'), J_{1-H} (with '-'), $H = 0.65$

Let $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ be two independent sequences of independent Gaussian random variables such that for each $n \in \mathbb{N}$ we have

$$E(X_n) = E(Y_n) = 0$$

and

$$\mathrm{Var}X_n = \frac{2c^2_{H(t)}}{x_n^{2H(t)} J^2_{1-H(t)}(x_n)}, \quad \mathrm{Var}Y_n = \frac{2c^2_{H(t)}}{y_n^{2H(t)} J^2_{-H(t)}(y_n)},$$

where

$$c^2_{H(t)} = \frac{\sin(\pi H(t))}{\pi} \Gamma(1 + 2H(t)).$$

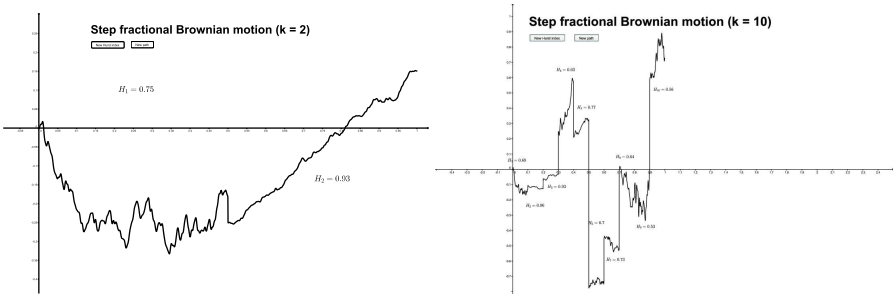
It is proved in [13] that the random process $\Big(B(t)\Big)_{t\in[0,1]}$ given by

$$B(t) = \sum_{n=1}^{\infty} \frac{\sin(x_n t)}{x_n} X_n + \sum_{n=1}^{\infty} \frac{1 - \cos(y_n t)}{y_n} Y_n, \quad t \in [0, 1]$$

is well defined and both series converge absolutely and uniformly in $t \in [0, 1]$with probability 1. The process B is a mfBm with Hurst index H.

For each $N \in \mathbb{N}$ we define the process

$$B_N(t) = \sum_{n=1}^{N} \frac{\sin(x_n t)}{x_n} X_n + \sum_{n=1}^{N} \frac{1 - \cos(y_n t)}{y_n} Y_n, \quad t \in [0, 1]. \tag{4}$$



a) $K = 2$ b) $K = 10$

Fig. 2. Approximation B_N of multifractional Brownian motion

Then using the above mentioned result from [6] we have

$$P(\lim_{N\to\infty} \sup_{t\in[0,1]} |B(t) - B_N(t)| = 0) = 1. \tag{5}$$

We will use the following result:

Theorem 1. *For all $N \in \mathbb{N}$ the approximating processes $(B_N(t))_{t\in[0,1]}$ are Lipschitz continuous with probability* 1 .

Proof. Let $N \in \mathbb{N}$ be fixed. We write

$$|B_N(t) - B_N(s)| \le \sum_{n=1}^{N} \left| \frac{\sin(x_n t) - \sin(x_n s)}{x_n} X_n \right| + \sum_{n=1}^{N} \left| \frac{\cos(y_n s) - \cos(y_n t)}{y_n} Y_n \right|.$$

But the functions sin and cos are Lipschitz continuous, therefore

$$|B_N(t) - B_N(s)| \le |t-s| \sum_{n=1}^{N} \Big(|X_n| + |Y_n| \Big) = C_N |t-s| \text{ for all } s,t \in [0,1],$$

where $C_N = \sum_{n=1}^{N} \Big(|X_n| + |Y_n| \Big) < \infty$ is a random variable.

2 Fractional Integrals and Derivatives

Let $a, b \in \mathbb{R}$, $a < b$ and $f, g : \mathbb{R} \to \mathbb{R}$. We use notions and results about fractional calculus from [12] and [15]:

$$f(a+) := \lim_{\delta \searrow 0} f(a+\delta), \quad f(b-) := \lim_{\delta \searrow 0} f(b-\delta),$$

$$f_{a+}(x) = \mathbb{I}_{(a,b)}(x)(f(x) - f(a+)), \quad g_{b-}(x) = \mathbb{I}_{(a,b)}(x)(g(x) - g(b-)).$$

Note that for $\alpha > 0$ we have $(-1)^\alpha = e^{i\pi\alpha}$.

For $f \in L_1(a,b)$ and $\alpha > 0$ the **left-** and **right-sided fractional Rieman-Liouville integral of f of order** α on (a,b) is given for a.e. x by

$$I_{a+}^{\alpha} f(x) = \frac{1}{\Gamma(\alpha)} \int_a^x (x-y)^{\alpha-1} f(y) dy$$

and

$$I_{b-}^{\alpha} f(x) = \frac{(-1)^{-\alpha}}{\Gamma(\alpha)} \int_x^b (y-x)^{\alpha-1} f(y) dy.$$

For $p > 1$ let $I_{a+}^{\alpha}(L_p(a,b))$, be the class of functions f which have the representation $f = I_{a+}^{\alpha} \Phi$, where $\Phi \in L_p(a,b)$, and let $I_{b-}^{\alpha}(L_p(a,b))$ be the class of functions g which have the representation $g = I_{b-}^{\alpha} \varphi$, where $\varphi \in L_p(a,b)$. If $0 < \alpha < 1$, then the functions Φ, respectively φ, in the above representations agree a.s. with the **left-sided** and respectively **right-sided fractional derivative of f of order** α (in the Weyl representation)

$$\Phi(x) = D_{a+}^{\alpha} f(x) = \frac{1}{\Gamma(1-\alpha)} \left(\frac{f(x)}{(x-a)^{\alpha}} + \alpha \int_a^x \frac{f(x)-f(y)}{(x-y)^{\alpha+1}} dy \right) \mathbb{I}_{(a,b)}(x)$$

and

$$\varphi(x) = D_{b-}^{\alpha} g(x) = \frac{(-1)^{\alpha}}{\Gamma(1-\alpha)} \left(\frac{g(x)}{(b-x)^{\alpha}} + \alpha \int_x^b \frac{g(x)-g(y)}{(y-x)^{\alpha+1}} dy \right) \mathbb{I}_{(a,b)}(x).$$

The convergence at the singularity $y = x$ holds in the L_p-sense. Recall that

$$I_{a+}^{\alpha}(D_{a+}^{\alpha} f) = f \text{ for } f \in I_{a+}^{\alpha}(L_p(a,b)), \quad I_{b-}^{\alpha}(D_{b-}^{\alpha} g) = g \text{ for } g \in I_{b-}^{\alpha}(L_p(a,b))$$

and

$$D_{a+}^{\alpha}(I_{a+}^{\alpha} f) = f, \quad D_{b-}^{\alpha}(I_{b-}^{\alpha} g) = g \text{ for } f, g \in L_1(a,b).$$

For completeness we denote

$$D_{a+}^{0} f(x) = f(x), D_{b-}^{0} g(x) = g(x), D_{a+}^{1} f(x) = f'(x), D_{b-}^{1} g(x) = g'(x).$$

Let $0 \le \alpha \le 1$. The **fractional integral** of f with respect to g is defined as

$$\int_a^b f(x) dg(x) = (-1)^{\alpha} \int_a^b D_{a+}^{\alpha} f_{a+}(x) D_{b-}^{1-\alpha} g_{b-}(x) dx + \\ + f(a+)(g(b-) - g(a+)) \tag{6}$$

if $f_{a+} \in I_{a+}^{\alpha}(L_p(a,b)), g_{b-} \in I_{b-}^{1-\alpha}(L_q(a,b))$ for $\frac{1}{p} + \frac{1}{q} \le 1$.

In our investigations we will take $p = q = 2$. If $0 \le \alpha < \frac{1}{2}$, then the integral in (6) can be written as

$$\int_a^b f(x) dg(x) = (-1)^{\alpha} \int_a^b D_{a+}^{\alpha} f(x) D_{b-}^{1-\alpha} g_{b-}(x) dx \tag{7}$$

if $f \in I_{a+}^{\alpha}(L_2(a,b))$, $f(a+)$ exists, $g_{b-} \in I_{b-}^{1-\alpha}(L_2(a,b))$ (see [15]).

3 The Stochastic Integral

Without loss of generality we consider $0 < T \leq 1$, because for arbitrary $T > 0$ we can rescale the time variable using the H-self-similar property of the mfBm meaning that $\Big(B(ct)\Big)_{t\geq 0}$ and $\Big(c^H(t)B(t)\Big)_{t\geq 0}$ are equal in distribution for every $c > 0$.

We will define the Itô integral $\int\limits_0^T G(u)dB(u)$ instead of $\int\limits_0^t G(u)dB(u)$ and use

$$\int\limits_0^t G(u)dB(u) = \int\limits_0^T \mathbb{I}_{[0,t]}(u)G(u)dB(u) \text{ for } t \in [0,T]$$

(see [15]).

We consider $\alpha > 1 - H$. It follows by (7) that

$$\int\limits_0^T G(u)dB(u) = (-1)^\alpha \int\limits_0^T D^\alpha_{0+}G(u)D^{1-\alpha}_{T-}B_{T-}(u)du \tag{8}$$

for $G \in I^\alpha_{0+}(L_2(0,T))$, where $G(0+)$ exists and $B_{T-} \in I^{1-\alpha}_{T-}(L_2(0,T))$.

The condition $G \in I^\alpha_{0+}(L_2(0,T))$ (with probability 1) means that $G \in L_2(0,T)$ and

$$\mathcal{I}_\varepsilon(x) = \int\limits_0^{x-\varepsilon} \frac{G(x) - G(y)}{(x-y)^{\alpha+1}}dy \text{ for } x \in (0,T)$$

converges in $L^2(0,T)$ as $\varepsilon \searrow 0$.

The condition $B_{T-} \in I^{1-\alpha}_{T-}(L_2(0,T))$ means $B_{T-} \in L_2(0,T)$ and

$$\mathcal{J}_\varepsilon(x) = \int\limits_{x+\varepsilon}^T \frac{B(x) - B(y)}{(y-x)^{2-\alpha}}dy \text{ for } x \in (0,T)$$

converges in $L_2(0,T)$ as $\varepsilon \searrow 0$ This condition for B is fulfilled for $\alpha > 1 - H$, since the fractional Brownian motion B is a.s. Hölder continuous with exponent $\gamma \in (0,H)$ (see [5]).

We will use (7) for the integrals with respect to the approximating processes $\Big(B_N(t)\Big)_{t\in[0,T]}$. Observe that $B_{N,T-} \in I^{1-\alpha}_{T-}(L_2(0,T))$, which follows from the Lipschitz continuity property in Theorem 1. We have

$$\int\limits_0^T G(u)dB_N(u) = (-1)^\alpha \int\limits_0^T D^\alpha_{0+}G(u)D^{1-\alpha}_{T-}B_{N,T-}(u)du \tag{9}$$

for $G \in I^\alpha_{0+}(L_2(0,T))$, where $G(0+)$ exists.

Let $\Big(Z(t)\Big)_{t\in[0,T]}$ be a cádlág process. Its **generalized quadratic variation process** $\Big([Z](t)\Big)_{t\in[0,T]}$ is defined as

$$[Z](t) = \lim_{\varepsilon\searrow 0} \varepsilon \int_0^1 u^{\varepsilon-1} \int_0^t \frac{1}{u}(Z_{t-}(s+u) - Z_{t-}(s))^2 dsdu + (Z(t) - Z(t-))^2,$$

if the limit exists uniformly in probability (see [17]).

In particular, if B is a fractional Brownian motion with Hurst index $H \in (\frac{1}{2}, 1)$ and B_N is an approximation of B as given in (4), it is easy to verify that

$$[B](t) = 0 \quad \text{and} \quad [B_N](t) = 0 \quad \text{for each} \ \ t \in [0,T], \tag{10}$$

because B is locally Hölder continuous with exponent $> \frac{1}{2}$ and B_N is Lipschitz continuous. The **Ito formula** for change of variable for fractional integrals is given in the next theorem.

Theorem 2 ([17]). *Let $\Big(Z(t)\Big)_{t\in[0,T]}$ be a continuous process with generalized quadratic variation $[Z]$. Let $Q : \mathbb{R}\times[0,T] \to \mathbb{R}$ be a random function such that a.s. we have $Q \in \mathcal{C}^1(\mathbb{R}\times[0,T])$ and $\frac{\partial^2 Q}{\partial x^2} \in \mathcal{C}(\mathbb{R}\times[0,T])$. Then, for $t_0, t \in [0,T]$ we have*

$$Q(Z(t),t) - Q(Z(t_0),t_0) = \int_{t_0}^t \frac{\partial Q}{\partial x}(Z(s),s)dZ(s) + \int_{t_0}^t \frac{\partial Q}{\partial t}(Z(s),s)ds + \frac{1}{2}\int_{t_0}^t \frac{\partial^2 Q}{\partial^2 x}(Z(s),s)d[Z]s.$$

Let $1 - H < \alpha < \frac{1}{2}$ and let $G \in I^{\alpha}_{0+}(L_2(0,T))$ such that $G(0+)$ exists. We define the processes

$$Z(t) = \int_0^t G(s)dB(s) \text{ and } Z_N(t) = \int_0^t G(s)dB_N(s), \quad t \in]0,T].$$

Then by [17] it follows that

$$[Z](t) = 0 \text{ and } [Z_N](t) = 0.$$

So,if $Q : \mathbb{R}\times[0,T] \to \mathbb{R}$ is a random function such that a.s. we have $Q \in \mathcal{C}^1(\mathbb{R}\times[0,T])$ and $\frac{\partial^2 Q}{\partial x^2} \in \mathcal{C}(\mathbb{R}\times[0,T])$, then for $t_0, t \in [0,T]$ we have

$$Q(Z(t),t) - Q(Z(t_0),t_0) = \int_{t_0}^t \frac{\partial Q}{\partial x}(Z(s),s)G(s)dB(s) + \int_{t_0}^t \frac{\partial Q}{\partial t}(Z(s),s)ds \tag{11}$$

and

$$Q(Z_N(t),t) - Q(Z_N(t_0),t_0) = \int_{t_0}^{t} \frac{\partial Q}{\partial x}(Z_N(s),s)G(s)dB_N(s) + \int_{t_0}^{t} \frac{\partial Q}{\partial t}(Z_N(s),s)ds. \tag{12}$$

4 Stochastic Differential Equations Driven by MultiFractional Brownian Motion

Let $\Big(B(t)\Big)_{t\geq 0}$ be a mfBm with Hurst parameter H such that $H(t) > \frac{1}{2}$. We investigate stochastic differential equations of the form

$$\begin{aligned} dX(t) &= F(X(t),t)dt + G(X(t),t)dB(t), \\ X(t_0) &= X_0, \end{aligned} \tag{13}$$

where $t_0 \in]0,T]$, X_0 is a random vector in $\mathbb{R}^n$ and the random functions F and G satisfy with probability 1 the following conditions:

(C1) $F \in C(\mathbb{R}^n \times [0,T], \mathbb{R}^n), G \in C^1(\mathbb{R}^n \times [0,T], \mathbb{R}^n)$;

(C2) for each $t \in [0,T]$ the functions $F(\cdot,t), \dfrac{\partial G(\cdot,t)}{\partial x^i}, \dfrac{\partial G(\cdot,t)}{\partial t}$ are locally Lipschitz for each $i \in \{1,\ldots,n\}$.

We consider the pathwise auxiliary partial differential equation on $\mathbb{R}^n \times \mathbb{R} \times [0,T]$

$$\begin{aligned} \frac{\partial K}{\partial z}(y,z,t) &= G(K(y,z,t),t), \\ K(Y_0,Z_0,t_0) &= X_0, \end{aligned} \tag{14}$$

where Y_0 is an arbitrary random vector in $\mathbb{R}^n$ and Z_0 an arbitrary random variable in $\mathbb{R}$. From the theory of differential equations it follows that with probability 1 there exists a local solution $K \in C^1(\mathbb{R}^n \times \mathbb{R} \times [0,T], \mathbb{R}^n)$ in a neighbourhood V of (Y_0,Z_0,t_0) with partial derivatives being Lipschitz in the variable y and

$$\det\left(\frac{\partial K^i}{\partial y^j}(y,z,t)\right)_{1\leq i,j\leq n} \neq 0.$$

For $(x,y,t) \in V$ we have

$$\frac{\partial^2 K}{\partial z^2}(y,z,t) = \sum_{j=1}^{n} \frac{\partial G}{\partial x^j}(K(y,z,t),t)G^j(K(y,z,t),t).$$

We also consider the pathwise differential equation (in matrix representation) on $[0,T]$

$$\frac{\partial K}{\partial y}(Y(t),B(t),t)dY(t)+\frac{\partial K}{\partial t}(Y(t),B(t),t)dt=F(K(Y(t),B(t),t),t)dt \quad (15)$$

$$Y(t_0)=Y_0, \quad (16)$$

or

$$dY(t)=\left(\frac{\partial K}{\partial y}(Y(t),B(t),t)\right)^{-1}\left[F(K(Y(t),B(t),t),t)-\frac{\partial K}{\partial t}(Y(t),B(t),t)\right]dt$$
$$Y(t_0)=Y_0,$$

which has a unique local solution on a maximal interval $]t_0^1,t_0^2[\subseteq[0,T]$ with $t_0\in]t_0^1,t_0^2[$ (see [9]).

Applying the Ito formula, and relation (11), to the random function $Q(z,t)=K(Y(t),z,t)$ (in fact, successively for $K^1,\ldots,K^n$) and the step fractional Brownian motion B we obtain

$$\begin{aligned}
&K(Y(t),B(t),t)-K(Y(t_0),B(t_0),t_0)\\
&=\sum_{j=1}^{n}\int_{t_0}^{t}\frac{\partial K}{\partial y^j}(Y(s),B(s),s)dY^j(s)+\int_{t_0}^{t}\frac{\partial K}{\partial z}(Y(s),B(s),s)dB(s)\\
&\quad+\int_{t_0}^{t}\frac{\partial K}{\partial t}(Y(s),B(s),s)ds\\
&=\sum_{j=1}^{n}\int_{t_0}^{t}\frac{\partial K}{\partial y^j}(Y(s),B(s),s)dY^j(s)+\int_{t_0}^{t}G(K(Y(s),B(s),s),s)dB(s)\\
&\quad+\int_{t_0}^{t}\frac{\partial K}{\partial t}(Y(s),B(s),s)ds\\
&=\int_{t_0}^{t}F(K(Y(s),B(s),s),s)ds+\int_{t_0}^{t}G(K(Y(s),B(s),s),s)dB(s).
\end{aligned}$$

Therefore,

$$X(t):=K(Y(t),B(t),t)$$

satisfies

$$X(t)=X_0+\int_{t_0}^{t}F(X(s),s)ds+\int_{t_0}^{t}G(X(s),s)dB(s).$$

Instead of the process $\left(B(t)\right)_{t\in[0,1]}$ we consider its approximations $\left(B_N(t)\right)_{t\in[0,1]}$ given in (4). For each $N\in\mathbb{N}$ we consider the pathwise differential equation (in

matrix representation)

$$\begin{aligned} dY_N(t) &= \left(\frac{\partial K}{\partial y}(Y_N(t), B_N(t), t)\right)^{-1} \Big[F(K(Y_N(t), B_N(t), t), t) \\ &\quad - \frac{\partial K}{\partial t}(Y_N(t), B_N(t), t)\Big] dt \\ Y_N(t_0) &= Y_0, \end{aligned}$$

which has a unique local solution Y_N on a maximal interval $(t^1, t^2) \subset (t_0^1, t_0^2)$ of existence which contains t_0 ([13]). Applying the Ito formula to the random function $Q(z,t) = K(Y_N(t), z, t)$ (in fact, successively for $K^1, \ldots, K^n$) and the process B_N we obtain

$$\begin{aligned} &K(Y_N(t), B_N(t), t) - K(Y_N(t_0), B_N(t_0), t_0) \\ &= \sum_{j=1}^{n} \int_{t_0}^{t} \frac{\partial K}{\partial y^j}(Y_N(s), B_N(s), s) dY_N^j(s) + \int_{t_0}^{t} \frac{\partial K}{\partial z}(Y_N(s), B_N(s), s) dB_N(s) \\ &\quad + \int_{t_0}^{t} \frac{\partial K}{\partial t}(Y_N(s), B_N(s), s) ds \\ &= \sum_{j=1}^{n} \int_{t_0}^{t} \frac{\partial K}{\partial y^j}(Y_N(s), B_N(s), s) dY_N^j(s) + \int_{t_0}^{t} G(K(Y_N(s), B_N(s), s), s) dB_N(s) \\ &\quad + \int_{t_0}^{t} \frac{\partial K}{\partial t}(Y_N(s), B_N(s), s) ds \\ &= \int_{t_0}^{t} F(K(Y_N(s), B_N(s), s), s) ds + \int_{t_0}^{t} G(K(Y_N(s), B_N(s), s), s) dB_N(s). \end{aligned}$$

Therefore,

$$X_N(t) := K(Y_N(t), B_N(t), t)$$

satisfies

$$X_N(t) = X_0 + \int_{t_0}^{t} F(X_N(s), s) ds + \int_{t_0}^{t} G(X_N(s), s) dB_N(s), \quad t \in]t_1, t_2[.$$

So we have the following pathwise property

$$\lim_{N \to \infty} \sup_{t \in]t_1, t_2[} \|Y_N(t) - Y(t)\| = 0.$$

Then the continuity properties of K and (5) imply that for a.e. $\omega \in \Omega$ it holds

$$\lim_{N \to \infty} \sup_{t \in]t_1, t_2[} \|X_N(t) - X(t)\| = 0.$$

By this we have proved the main result of our paper:

Theorem 3. *Let B be a mfBm approximated by the processes B_N given in* (4) *and* (5). *Let $F, G : \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$ be random functions satisfying conditions* (C1) *and* (C2) *with probability* 1. *Let $t_0 \in]0,T]$ be fixed. Then each of the stochastic equations*

$$X(t) = X_0 + \int_{t_0}^{t} F(X(s),s)ds + \int_{t_0}^{t} G(X(s),s)dB(s),$$

$$X_N(t) = X_0 + \int_{t_0}^{t} F(X_N(s),s)ds + \int_{t_0}^{t} G(X_N(s),s)dB_N(s), \quad N \in \mathbb{N}$$

admits almost surely a unique local solution on a common interval $]t_1,t_2)[$ (which is independent of N and contains t_0). Moreover, we have the following approximation result

$$P(\lim_{N\to\infty} \sup_{t\in]t_1,t_2[} \|X_N(t) - X(t)\| = 0) = 1.$$

5 Application

We consider the one dimensional stochastic linear equation from financial mathematics, modeling the price S of a stock

$$S(t) = S_0 + \int_0^t \mu(s)S(s)ds + \int_0^t \sigma(s)S(s)dB(s),$$

where $(B(t))_{t\in[0,T]}$ is a mfBm with Hurst index $H(t) > \frac{1}{2}$, μ is the interest rate and σ the volatility function.

It is known (see [8]) that this equation has the following unique solution

$$S(t) = S_0 \exp\left\{\int_0^t \mu(u)du + \int_0^t \sigma(u)dB(u)\right\} \text{ for all } t \in [0,T].$$

By the methods of the above section we approximate B through the processes B_N, via (4) and (5) and consider

$$S_N(t) = S_0 \exp\left\{\int_0^t \mu(u)du + \int_0^t \sigma(u)dB_N(u)\right\} \text{ for all } t \in [0,T].$$

Using Theorem 3 it follows that

$$P(\lim_{N\to\infty} \sup_{t\in[0,T]} \|S_N(t) - S(t)\| = 0) = 1.$$

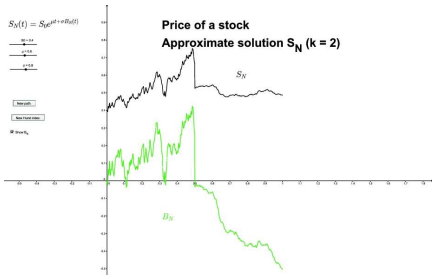
In the special case when μ and σ are constants, we obtain that the price of a stock is

$$S(t) = S_0 e^{\mu t + \sigma B(t)}$$

and we can simulate it by computer using

$$S_N(t) = S_0 e^{\mu t + \sigma B_N(t)}$$

as given in Figure 3, where K is the number of constant pieces of H.



$K = 2$

Fig. 3. Approximated solution S_N

References

1.A. Ayache and M. S. Taqqu. Rate optimality of wavelet series approximations of fractional Brownian motion. *Jour. Fourier Anal. Appl.* **9**, 451-471, 2003.
2.A. Benassi, S. Jaffard and D. Roux. Elliptic Gaussian random processes, *Rev. Mathematica Iberoamericana,* **13**, no.1, 19–90, 1997.
3.L. Coutin and L. *Decreusefond. Stochastic Differential Equations Driven by a Fractional Brownian Motion*, Tech. Report 97C004, ENST Paris 1997.
4.L. Coutin and Z. Qian. Stochastic Differential Equations for Fractional Brownian Motions, *C. R. Acad. SCI. Paris,* **331**, Srie I, 75–80, 2000.
5.L. Decreusefond and A.S. Üstünel. Stochastic Analysis of the Fractional Brownian Motion. *Potential Analysis,* **10**, 177–214, 1999.
6.D. Dzhaparidze and H. van Zanten. A Series Expansion of Fractional Brownian Motion, *Probab. Theory Relat. Fields,* **130**, 39–55, 2004.
7.M.L. Kleptsyna, P.E. Kloeden and V.V. Anh. Existence and Uniqueness Theorems for FBM Stochastic Differential Equations, *Information Transmission Problems,* **34**, 41–51, 1998.
8.F. Klingenhöfer and M. Zähle. Ordinary Differential Equations with Fractal Noise, *Proc. Amer. Math. Soc.* **127**, 1021–1028, 1999.
9.H. Lisei, and A. Soós. Approximation of Stochastic Differential Equations Driven by Fractional Brownian Motion, in *Seminar on stochastic analysis, random fields and applications V, Progress in Probability,* **59**, Birkhuser Verlag, 229-244, 2007.
10.B. Mandelbrot, and Van Ness. Fractional Brownian motion, fractional noises and applications, *SIAM Rev.,* **10**, 422-437, 1968.
11.R. Peltier and J. Lévy Vehel. *Multifractional Brownian Motion: definition and preliminary results*, Rapport technique INRIA, 1996.

12.S. G. Samko, A. A. Kilbas, and O. I. Marichev. *Fractional Integrals and Derivatives. Theory and Applications,* Gordon and Breach, 1993.
13.A. Soós. Series expansion of multifractional Brownian motion,(to appear).
14.G. N. Watson. *A Treatise on the Theory of Bessel Functions,* Cambridge University Press, 1944.
15.M. Zähle. Integration with respect to Fractal Functions and Stochastic Calculus I., *Probab.Theory Relat. Fields,* **111**, 333–374, 1998.
16.M. Zähle. Ordinary Differential Equations with Fractal Noise. *Proc. Am. Math. Soc.,* **127**, 333–374, 1999.
17.M. Zähle. Integration with respect to Fractal Functions and Stochastic Calculus II. *Math. Nachr.,* **225**, 145–183, 2001.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Interpolation methods for internet traffic

Soós Anna Somogyi Ildikó

Abstract

The classical methods of data interpolation can be generalized with fractal interpolation. Our aim is to maid some comparison of the fractal and numerical analysis interpolation methods. The experimental data regarding the internet traffic were processed using fractal interpolation and also spline and some Shepard type interpolation.

1 Spline interpolation

Let $H^{m,2}[a,b]$, $m \in \mathbb{N}^*$ be the set of functions $f \in C^{m-1}[a,b]$ with $f^{(m-1)}$ absolutely continuous on $[a,b]$ and $f^{(m)} \in L^2[a,b]$, $\Lambda = \{\lambda_i | \lambda_i : H^{m,2}[a,b] \to \mathbb{R}, i = 1...,n\}$ a set of linear functionals, $yin\mathbb{R}^n$ and

$$U_y = \{f \in H^{m,2}[a,b] | \lambda_i(f) = y_i, i = 1,...,n\}.$$

Definition 1 *The problem that consists of determining the elements $s \in U$ such that*

$$\left\| s^{(m)} \right\|_2 = \inf_{u \in U} \left\| u^{(m)} \right\|_2$$

is called polinomial spline interpolation problem.

For the solution of a spline interpolation problem we can give the following structural characterization theorem ([4]):

Theorem 2 *Let Λ be a set of Birkhoff type functionals and let U be the corresponding interpolatory set. The functions $s \in U$ is a solution of the spline interpolation problem if and only if:*

1. $s^{(2m)}(x) = 0, \quad x \in [x_1, x_k]$
   $\{x_1, ...x_k\}$,
2. $s^{(m)}(x) = 0, \quad x \in (a, x_1) \cup (x_k, b)$,
3. $s^{(2m-\mu-1)}(x_i - 0) = s^{(2m-\mu-1)}(x_i + 0), \quad \mu\{0, 1, ..., m-1\}$
   I_i *for* $i = 1, ..., k$.

The caracterization theorem states that the solution s of the polynomial spline interpolation problem is a polynomial of $2m-1$ degree on each interior interval (x_i, x_{i+1}) and it is a polynomial of $m-1$ degree on the intervals $[a, x_1)$ and $(x_k, b]$. Furthermore, the derivative of order $2m-\mu-1$ is continuous in x_i if the value of the νth ordin derivative in x_i does not belong to Λ.

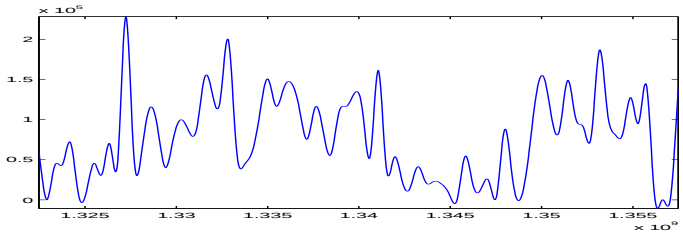


Figure 1: Spline interpolation: for internet traffic data

Definition 3 *The solution s of the polynomial spline interpolation problem is called a natural spline function of order $2m-1$.*

When $\Lambda = \{\lambda_i | \lambda_i(f) = f(x_i), i = 1, ..., n\}$, with $x_i \in [a, b], i = 1, ..., n$ and $n \geq m$, then for every $f \in H^{m,2}[a, b]$ the interpolation spline function $S_L f$ exists, is unique and the corresponding operator is called the spline operator of Lagrange type.
The function $S_L f$ may be written in the form

$$S_L f = \sum_{k=1}^{n} s_k f(c_k),$$

where $s_k k = 1, .., n$ are the fundamental interpolation spline functions. To determine these functions we use the caracterization theorem and we have

$$s_k(x) = \sum_{i=0}^{m-1} a_i^k x_i + \sum_{j=1}^{n} b_j^k (x - x_j)_+^{2m-1}, k = 1, ..., n,$$

whith $a_i^k, i = 0, ..., m-1$ and $b_j^k, j = 1, ..., n$ obtained as the solution of the following systems:

$$\begin{aligned} s_k^p(\alpha) &= 0, \ p = m, ..., 2m-1, \ \texttt{and} \ \alpha > x_n \\ s_k(x_\nu) &= \delta_{k\nu}, \ \nu = 1, ..., n \end{aligned}$$

for $k = 1, ..., n$.
We will use the third degree Lagrange type spline interpolation operator on internet traffic datas.

2 The Shepard operator

Our next approximation method is the Shepard method, introduced in 1968, which is a well suited method for interpolation of very large scattered data sets. It has the advantages of a small storage requirement and an easy generalization to additional independent variables.

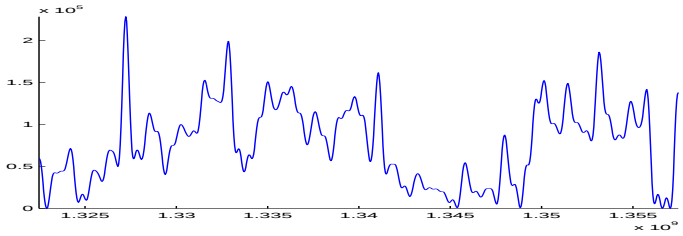


Figure 2: Shepard interpolation: for internet traffic data

Considering the interpolation points $(x_i, f(x_i))$, $i = 0, ..., N$, Shepard introduced in [5] the linear interpolation operator

$$(S_0 f)(x) = \sum_{i=0}^{N} A_{i,\mu}(x) f(x_i), \qquad \mu > 0,$$

with

$$A_{i,\mu}(x) := \frac{\|x - x_i\|^{-\mu}}{\sum_{k=0}^{N} \|x - x_k\|^{-\mu}}, \qquad i = 0, ..., N.$$

S_0 reproduces exactly only the constant functions. To avoid this, several authors, starting with Shepard himself, have suggested to apply S_0 not directly to $f(x_i)$, but to some interpolation operators $P[f, x_i](x)$ at x_i by considering the so-called combined operator:

$$(S_P f)(x) = \sum_{i=0}^{N} A_{i,\mu}(x) P[f, x_i](x), \qquad \mu > 0. \tag{1}$$

The operator S_P still interpolates f at x_i, $i = 0, ..., N$ but the algebraic degree of exactness is $\max_{i=0,...,N} \text{dex}(P[f, x_i])$.

The combined Shepard-Lagrange operator S_{L_m} is given by [3], [2]:

$$(S_{L_m} f)(x) = \sum_{i=0}^{N} A_i(x) (L_m^i f)(x),$$

where

$$(L_m^i f)(x) = \sum_{\nu=0}^{m} \frac{u_i(x)}{(x - x_{i+\nu}) u_i'(x_{i+\nu})} f(x_{i+\nu})$$

is the Lagrange polynomial corresponding to the set $\Lambda_i(f) = f(x_{i+\nu}) : \nu = 0, 1, \ldots, m\}$, $i = 0, \ldots, N$, with $x_{N+\nu} = x_{N-\nu}$.

In our next Figure we plot the graphics of $S_0 f$ considering $\mu = 2$ and the same datas.

3 Fractal functions

The tird interpolation will be the fractal interpolation on internet traffic datas.

Let (X, d) be a complete metric space, let $D(X)$ the class of all non-empty closed bounded subsets of X. Then $(D(X), h)$ is a complete metric space with the Hausdorff metric: $h : D(X) \times D(X) \to R$

$$h(A, B) := \sup\{\sup_{a\in A} \inf_{b\in B} d(a, b), \sup_{b\in B} \inf_{a\in A} d(a, b)\}$$

Let $E \subset X$, $p \geq 0$, $\epsilon > 0$, and define the Hausdorff p-dimensional measure of E:

$$\mathcal{H}^p(E) := \lim_{\epsilon\to 0} \mathcal{H}^p_\epsilon(E) = \sup_{\epsilon>0} \mathcal{H}^p_\epsilon(E),$$

where

$$\mathcal{H}^p_\epsilon(E) := \inf\{\sum_{i=1}^{\infty} |E_i|^p, E \subset \cup_{i=1}^{\infty} E_i, |E_i| < \epsilon\}$$

For each E there is a unique real number q,named the Hausdorff dimension of E, such that

$$\mathcal{H}^p(E) = \begin{cases} +\infty & if \quad 0 \leq p < q \\ 0 & if \quad q < p < \infty \end{cases}$$

B. Mandelbrot define fractal as the set of which Hausdorff dimension is noninteger.

The functions $f : I \to \mathbf{R}$, where I is a real closed interval, is named by M. F. Barnsley *fractal function* if the Hausdorff dimensions of their graphs are noninterger.

Let N a natural number, $N > 1$, and let $w_i : \ X \to X : \ i \in \{1, ..., N\}$ be continuous. Then we call $\{X, w_i : i = 1, ..., N\}$ an *iterated function sistem (IFS).*

If, for some $0 \leq k < 1$ and all $i \in \{1, ..., N\}$,

$$d(w_i(x), w_i(x')) \leq kd(x, x'), \forall\, x, x' \in X,$$

then the IFS is named *hyperbolic.*

Define $W : \mathcal{D}(X) \to \mathcal{D}(X)$ by

$$W(A) := \cup_{i=1}^{N} w_i(A),$$

where $w_i(A) = \{w_i(x) : x \in A\}$.

W is a contraction mapping if the IFS is hyperbolic:

$$h(W(A), W(B)) \leq kh(A, B) \forall\, A, B \in \mathcal{D}(X).$$

Any set $G \in D(X)$ such that $W(G) = G$ is called an *attractor* for the IFS.

Theorem 4 *(Hutchinson) Let $\{X, w_i\, i = 1, ..., N\}$ an hyperbolic IFS. There is a unique compact set $G \subset X$, such that $W(G) = G$, and*

$$G := \lim_{n\to\infty} W^n(E), \ E \in \mathcal{D}(X), W^0$$

Let $\{(x_i, y_i) \in R^2, i = 0, 1, \cdots, N\}$ be given, and $I = [x_0, x_N]$. The functions $f : I \to R$, which interpolate the data according to $f(x_i) = y_i$, $i = 0, 1, ..., N$, and whose graphs are attractors of IFS are *fractal interpolation functions*

Let $X = I \times [a, b]$ with Euclidean metric d, $I_n = [x_{n-1}, x_n]$ $u_n : I \to I_n$, $n \in \{1, 2, ..., N\}$, contractiv homeomorphism such that

$$u_n(x_0) := x_{n-1},\ u_n(x_N) := x_n,\ \forall n \in \{1, \cdots, N\}.$$

$$|u_n(c_1) - u_n(c_2)| \leq l|c_1 - c_2|,\ c_1, c_2 \in I,\ 0 \leq l < 1$$

$v_n : X \to [a, b]$ continuous, with

$$v_n(x_0, y_0) := y_{n-1},\ v_n(x_N, y_N) := y_n,\ \forall n \in \{1, \cdots, N\}.$$

$$|v_n(c, d_1) - v_n(c, d_2)| \leq q|d_1 - d_2|,\ c \in I,\ d_1, d_2 \in [a, b], 0 \leq q < 1.$$

Let $w_n : X \to X$, $n \in \{1, 2, ..., N\}$

$$w_n(x, y) = (u_n(x), v_n(x, y)).$$

$\{X, w_n : n = 1, 2, ..., N\}$ is an IFS but may not be hyperbolic.

Theorem 5 *(Barnsley) For the IFS $\{X, w_n : n = 1, 2, ..., N\}$ defined above, there is a metric d equivalent to the Euclidean metric, such that the IFS is hyperbolic with respect to d. The unique attractor G of the IFS is the graph of a continuous function $f : I \to R$ which interpolates the date set $\{(x_i, y_i) \in R^2, i = 0, 1, \cdots, N\}$*

Example: Let $\{(x_i, y_i) \in R^2, i = 0, 1, \cdots, N\}$, $N > 1$

$$w_n(x, y) = \begin{pmatrix} a_n & 0 \\ c_n & d_n \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e_n \\ f_n \end{pmatrix},$$

where $|d_n| < 1$ is given, a_n, c_n, e_n, f_n are real number such that

$$w_n(x_0, y_0) := (x_{n-1}, y_{n-1}),\ w_n(x_N, y_N) := (x_n, y_n)$$

We can solve the above equations for a_n, c_n, e_n, f_n

$$a_n = \frac{x_n - x_{n-1}}{x_N - x_0},$$

$$c_n = \frac{y_n - y_{n-1}}{x_N - x_0} - \frac{d_n(y_N - y_0)}{x_N - x_0},$$

$$e_n = \frac{x_N x_{n-1} - x_0 x_n}{x_N - x_0}$$

$$f_n = \frac{x_N y_{n-1} - x_0 y_n}{x_N - x_0} - \frac{d_n(x_N y_0 - x_0 y_N)}{x_N - x_0}.$$

w_n is a shear transformation: it maps lines parallel to the y-axis into the lines parallel to the y-axis. d_n is the vertical scaling factor.

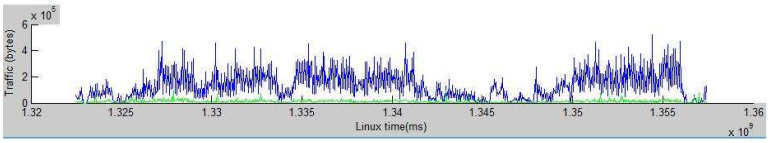


Figure 3: Fractal interpolation: for internet traffic data

References

[1] Barnsley M.F., *Fractals everywhere*, Academic Press, Orlando, Florida, 1988.

[2] Cătinaş, T., *Interpolation of scattered data*, Ed. "Casa Cărţii de Ştiinţă", Cluj-Napoca, 2007

[3] Coman Gh., Trîmbiţaş, R., *Combined Shepard univariate operators*, East Jurnal on Approximations **7** (2001), pp. 471–483

[4] Coman Gh., Birou M., Oşan C., Somogyi I., Cătinaş T., Oprişan A., Pop I., Todea I., *Interpolation operators*, Ed. "Casa Cărţii de Ştiinţă", Cluj, 2004.

[5] Shepard, D., *A two dimensional interpolation function for irregularly spaced data*, Proc. 23rd Nat. Conf. ACM, 1968, pp. 517–523.

[6] Somogyi I., András Sz., *Numerikus Analízis*, Ed. "Presa Universitară Clujeană", Cluj, 2009.

[7] Ruoyu Yan, Yingfeng Wang, *Hurst Parameter for Security Evaluation of LAN Traffic*, Information Technology Journal, 2-12, 11: 269 − 275. ($http://scialert.net/fulltext/?doi = itj.2012.269.275\&org = 11$)

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Intergenerational transmission of education in Greece: evidence from the European Social Survey 2002–2010

G. Stamatopoulou, M. Symeonaki, K. Michalopoulou

Department of Social Policy, Panteion University of Social and Political Sciences,
136 Syggrou Av., 176 71, Athens, Greece
(E-mail: gl.stamat@gmail.com, msimeon@panteion.gr, kmichal@panteion.gr)

Abstract. The present paper explores the patterns of intergenerational educational mobility in Greece and its changes for different birth cohorts born between 1930 and 1976. More specifically, we investigate the transmission of educational attainments from both father and mother through generations and over time. The main purpose is to trace the transitions of individuals between educational categories and to determine the relationship between an individual's education class and the class of his or her parents. Based on data drawn from ESS (2002–2010), our analysis provides Markov transition probability matrices and the absolute and relative mobility rates, by comparing the different rounds of the survey.
Keywords: Transmission of education, intergenerational mobility, transition probability matrices, mobility indices, cohort analysis, ESS.

1 Introduction

Intergenerational mobility can be defined as the trajectories observed from one generation to another and between different social classes. The term indicates whether and to what extent the socio-economic status of origins (measured here in terms of parental education) transmit from parents to children and has been used as a measurement of equality of opportunities.

Among other factors, education has played an important role in the study of social mobility, as it can mediate between the class of origins and the class of destination, by forming the individuals' social status. In this respect and driven by the principles of human capital theory, in the middle '60s, the Greek educational system has undergone major reforms, in order to provide more equal opportunities in education and to promote greater social mobility and fairness. As a result, an explosion of education and an improvement in individuals' educational outcomes were observed and the number of higher education graduates has rapidly increased in recent decades [4,9].

Empirical evidence illustrates that despite the expansion of education, Greece as well as other Mediterranean countries are the most immobile across Europe, as there is a linkage between paternal education and individuals outcomes [2,3,12]. However, an increase in the mobility rates is observed over time and the individuals seem to move upwards, attaining a higher educational level [4,15].

In the present paper, we investigate the intergenerational transmission of education in Greece. The main aim is to capture the extent of educational mobility through generations and to provide further evidence on the movements of individuals within the Greek educational system. Two specific research questions are examined: (i) To what degree does the educational status of both parents pass on to their children? (ii) How have the mobility rates changed over time and among different birth cohorts? Our analysis is based on data drawn from all rounds of ESS (2002–2010), except for the third one (2006) in which Greece did not participate.

The paper is organised as follows. Section 2 deals with the data and the methodology used in the analysis of intergenerational mobility. Section 3 presents the findings of the analysis, the relationship between origins and individuals' education and the respective patterns of educational mobility, while Section 4 provides the conclusions regarding the finding of Section 3.

2 Data and methods

The European Social Survey (ESS)[1] is a long-term comparative research project designed to record and document the attitudes, beliefs and behavioural patterns of the European populations. Funding via the European Commission's Framework programmes, the European Science Foundation and national academic funding agencies, the survey aims to produce comparable social indicators to be used for the European social policy. Started in 2002, the ESS is conducted every two years in more than 20 European countries. It involves national probability samples, a minimum target response rate of 70% and rigorous methodological criteria. The survey population is defined as all individuals aged 15 years and more, regardless of their nationality, citizenship or legal status. The 'homeless' and people living in collective dwellings are excluded.[8]

The ESS is one of the very few free access databases in Greece which provides the opportunity to investigate the trends of intergenerational social mobility, as it provides data on social status of both parents and individuals, even if they do not live in the same residence. In particular, we focus on raw data relating to the highest educational attainment of both parents and individuals, which were harmonised according to the latest version of International Standard Classification of Education (ISCED11).

For analytical purposes, the educational attainment has been recoded into four educational states, as indicated in Table 1.[2] Moreover, the data was weighted by applying the design weight ('dweight'), as is required by probability sampling theory.

In contrast to the economic trandition, where a regression approach is usually adopted, we base our analysis on a more sociological descriptive perspective. Thus, as in Symeonaki *et al* [14–16], we constructed Markov transition

[1] For more information on ESS visit http://www.europeansocialsurvey.org/.

[2] Since 1974 both primary and lower secondary education are compulsory, while attendance in the upper secondary schools is optional. According to OECD, the advanced vocational training or post-secondary can be allocated to tertiary education in a broader way, even if it is not tertiary level.[3]

ISCED	Educational categories	Description	States
0-1	Less than lower secondary	Primary education (un)completed	1
2	First stage of secondary completed	3-year lower secondary education	2
3	Second stage of secondary completed	3-year upper secondary education	3
4-6	Advanced education	Post-secondary or tertiary education	4

Table 1. Educational categories according to ISCED11 classification used by ESS.

probability matrices, the elements of which show the transitions that take place between educational categories. Each element $p_{ij}, \forall i, j = 1, 2, ..., n$ of a Markov matrix $\mathbf{P}$ describes the probability of an individual to move from state i (educational level of origins) to state j (individual's educational level). The elements found off the main diagonal of matrix $\mathbf{P}$ give the movements of individuals, while p_{ii} denotes the probabilities of individuals, $\forall i, j = 1, 2, ..., n$ being immobile.

Two types of indices are calculated to show the movements of individuals within the educational system: (i) relative indices such as the Prais-Shorrocks index, which indicate the rate of social fluidity and (ii) absolute mobility indices, which reflect the direction of the movements.[3] More specifically, we have computed the following indices:

The Prais-Shorrocks index [11,13] given by Equation (1):

$$M_{PS} = \left(\frac{1}{n-1}\right)(n - tr(\mathbf{P})) \tag{1}$$

where $tr(\mathbf{P})$ represents the sum of the diagonal elements of a transition matrix $\mathbf{P}$, n is the number of states and $M_{PS} \in [0, 1]$.
$M_{PS} = 1$ indicates perfect mobility and $M_{PS} = 0$ implies perfect immobility.

The Bartholomew Index[1] defined by Equation (2):

$$M_B = \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k} p_{ij}|i - j| \tag{2}$$

where p_{ij} as mentioned above denotes the probability of an individual to move from state i (social status of origins) to state j (individual's social status) and k is the number of states. The minimum value of the index is zero, which indicates perfect immobility.

The immobility ratio[7] is given by:

$$IM = \left(\frac{tr(\mathbf{P})}{n}\right) \tag{3}$$

and provides individual's rate of remaining to the social state of their origin, as well the degree of educational transmission through generations.

[3] Absolute mobility indices refer to the absolute number of individuals moving from one state to another, while the relative mobility rates are referring to the probabilities that individuals have to move upwards or downwards.[10]

Finally, Equations (4) and (5) define an upward and downward mobility index, respectively, based on the absolute number of individuals [10]:

$$u = \frac{1}{N} \sum_{j>i} n_{ij} \tag{4}$$

$$d = \frac{1}{N} \sum_{j<i} n_{ij}. \tag{5}$$

Additionally, in order to show how the mobility patterns have changed through generations, a synchronic cohort analysis was carried out. Three birth cohorts were defined (1930–1945, 1946–1960, 1961–1976) to ensure sufficiently large number of cases in each cell. Individuals aged 25 and less were not included in the analysis, as they have not in theory completed their education.[5]

3 Empirical results

In this section, we proceed with the presentation of the results of our analysis. In Table 2, the Markov transition probability matrices and the mobility indices according to father's educational profile are provided for all rounds of ESS.

ESS	**Father's state**	**Individual's state**				**Mobility indices**	
		1	2	3	4	M_{PS}	M_B
2002	1	0.492	0.189	0.239	0.080	0.734	0.758
	2	0.131	0.270	0.401	0.197		
	3	0.042	0.130	0.568	0.260		
	4	0.017	0.160	0.354	0.470		
2004	1	0.483	0.168	0.246	0.104	0.803	0.857
	2	0.126	0.126	0.437	0.311		
	3	0.030	0.131	0.475	0.364		
	4	0.018	0.188	0.288	0.506		
2008	1	0.265	0.210	0.355	0.171	0.870	0.943
	2	0.057	0.205	0.405	0.333		
	3	0.011	0.165	0.414	0.410		
	4	0.007	0.108	0.378	0.507		
2010	1	0.374	0.127	0.369	0.128	0.794	0.887
	2	0.085	0.144	0.520	0.251		
	3	0.046	0.108	0.551	0.295		
	4	0.014	0.061	0.376	0.549		

Table 2. The transition probabilities and the estimated mobility indices according to father's educational state (European Social Survey 2002–2010)

As shown, the transmission of paternal educational disadvantages to individuals attainments is obvious in the first two rounds of ESS, as children of less

educated fathers have high chances to remain to the first educational category in accordance with their origins', while those of more educated fathers are more likely to attain a higher educational level. The picture is somewhat different in the last two rounds, as more upward movements are observed. As a result, both M_{PS} and M_B are quite high for the whole surveyed population and they seem to slightly increase across the ESS rounds, indicating that father's educational profile affects the individuals' educational attainments to a lesser extent over time.

A similar pattern is detected when we examine the transition probability matrices and the extracted mobility indices according to mother's educational background (Table 3). Note that the mother's education and the individual's educational attainment are related in different rounds of ESS. However, an increasing trend of mobility rates over time is also observed.

ESS	**Mother's state**	**Individual's state**				**Mobility indices**	
		1	2	3	4	M_{PS}	M_B
2002	1	0.467	0.192	0.246	0.095	0.809	0.807
	2	0.056	0.240	0.497	0.207		
	3	0.016	0.141	0.471	0.372		
	4	0.000	0.143	0.462	0.396		
2004	1	0.462	0.163	0.255	0.120	0.838	0.945
	2	0.028	0.159	0.434	0.379		
	3	0.027	0.153	0.448	0.372		
	4	0.092	0.183	0.308	0.417		
2008	1	0.253	0.203	0.358	0.186	0.870	0.943
	2	0.010	0.161	0.389	0.440		
	3	0.011	0.207	0.428	0.355		
	4	0.000	0.141	0.329	0.529		
2010	1	0.357	0.124	0.370	0.150	0.826	0.887
	2	0.035	0.106	0.603	0.255		
	3	0.024	0.123	0.541	0.312		
	4	0.013	0.094	0.375	0.519		

Table 3. The transition probabilities and the estimated mobility indices according to mother's educational state (European Social Survey 2002–2010)

A more illustrative representation of the above results is given in Figure 1, where the progress of both mobility rates across the different rounds of the survey is provided. The increasing trend of the mobility rates based on both parents' profile over time is notable. Additionally, although the mobility rates seem to follow the same patterns for both parents, it is remarkable that the transmission of father's educational status to individuals attainments is more evident than that of mother's.

The comparison between birth cohorts gives also interesting results (Tables 4–5), as it reveals an increasing trend in mobility rates over time.

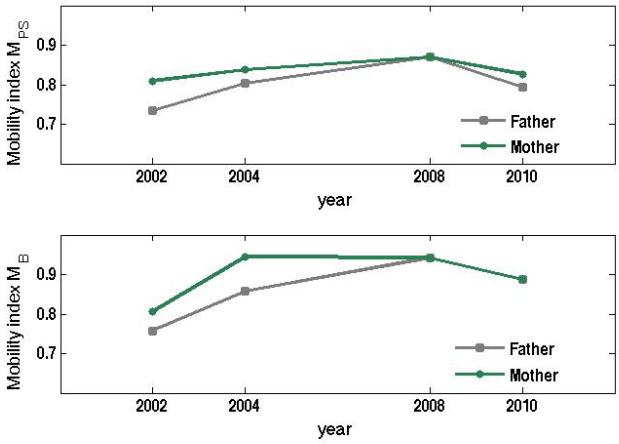


Fig. 1. The mobility rates across the European Social Survey

Birth cohorts	M_{PS}	M_B	IM	UM	DM
ESS Round 1, 2002					
1930 – 1945	0.717	0.744	0.462	0.242	0.061
1946 – 1960	0.755	0.810	0.434	0.531	0.038
1961 – 1976	0.792	0.836	0.406	0.655	0.045
ESS Round 2, 2004					
1930 – 1945	0.710	0.785	0.467	0.241	0.034
1946 – 1960	0.736	0.794	0.448	0.494	0.022
1961 – 1976	0.812	0.892	0.391	0.717	0.022
ESS Round 4, 2006					
1930 – 1945	0.528	0.487	0.604	0.336	0.022
1946 – 1960	0.900	0.935	0.325	0.626	0.036
1961 – 1976	0.825	0.960	0.381	0.781	0.026
ESS Round 5, 2008					
1930 – 1945	0.636	0.685	0.523	0.233	0.040
1946 – 1960	0.706	0.761	0.470	0.533	0.039
1961 – 1976	0.834	0.932	0.375	0.721	0.035

Table 4. Intergenerational mobility based on father's education, by birth cohorts and ESS rounds

More particularly, although both the M_{PS} and M_B differ between the rounds of ESS, the indices seem to increase through generations, indicating that both paternal and maternal educational profile affects to a lesser extent the individuals' educational attainments. The most illustrative example is provided in the fourth round of ESS (Table 5), while the Bartholomew mobility index with respect to mother's educational profile has substantially changed through generations, taking values between 0.463 (for the oldest birth cohort) and 1.004 (for the youngest birth cohort).

Birth cohorts	M_{PS}	M_B	IM	UM	DM
ESS Round 1, 2002					
1930 – 1945	0.687	0.653	0.484	0.269	0.010
1946 – 1960	0.890	0.881	0.333	0.584	0.017
1961 – 1976	0.857	0.889	0.357	0.740	0.020
ESS Round 2, 2004					
1930 – 1945	0.846	1.203	0.365	0.264	0.018
1946 – 1960	0.844	0.906	0.367	0.540	0.013
1961 – 1976	0.869	0.968	0.348	0.748	0.020
ESS Round 4, 2006					
1930 – 1945	0.533	0.463	0.600	0.346	0.009
1946 – 1960	0.805	0.838	0.396	0.656	0.004
1961 – 1976	0.885	1.004	0.336	0.852	0.012
ESS Round 5, 2008					
1930 – 1945	0.571	0.524	0.571	0.275	0.005
1946 – 1960	0.729	0.767	0.454	0.582	0.007
1961 – 1976	0.830	0.937	0.377	0.780	0.017

Table 5. Intergenerational mobility based on mother's education, by birth cohorts and ESS rounds

Consequently, the immobility ratio appears to decrease through generations, indicating that the transmission of parental background to their children's education seems to be weakened.

Regarding the directions of the movements, it is worth noting that very low rates of upward mobility for the oldest birth cohort (1930–1945) exist in all rounds but increase considerably over time. Nevertheless, the upward movements are more perceptible than the downward transitions.

4 Concluding remarks

In the present paper we attempted to provide further analysis on the relationship between parental education and individuals educational outcomes, based

on data drawn from ESS. Focusing on the method of Markov transition probability matrices and by estimating widely used mobility indices, our analysis revealed that mobility patterns have substantially changed through generations and an increase in the chances of individuals to attain a higher educational level has been noted over time. Additionally, it seems that the effect of paternal educational background on individuals' attainments is stronger, as the mobility rates appear higher with respect to mother's profile. However, in terms of social policy, there is a further need of actions to consider in order to weaken the intergenerational transmission of educational disadvantades.

References

1.D.J. Bartholomew. *Stochastic Models for social processes*, London, 1982. Wiley.
2.O. Causa and A. Johansson. Intergenerational social mobility in OECD countries. *OECD Journal: Economic Studies*, 2010:1-44, 2010.
3.S. Comi. Intergenerational mobility in Europe: Evidence from ECHP. *Universita di Milano*, 2004.
4.J. Daouli, M. Demousis, N. Giannakopoulos. Mothers, fathers and daughters: Intergenerational transmision of education in Greece. *Economics of Education Review*, 29:83-93, 2010.
5.M. van Doorn, I. Pop, M. Wolbers. Intergenerational Transmission of Education across European countries and cohorts. *European Societies*, 13:93-117, 2011
6.R. Erikson and J.H. Goldthorpe. *The Constant Flux: The Study of Class Mobility in Industrial Society*, Oxford, 1992. Clareston Press.
7.G. Heineck and R. Riphahn. Intergenerational transmission of educational attainment in Germany - the last five decades. *Journal of Economic and Statistics*, 229:36-60, 2007.
8.R. Jowel and the Central Co-ordination Team. *European Social Survey 2004/2005: Technical Report*. Centre for Comparative Social Surveys, City University. London, 2005.
9.H. Lambropoulos, G. Psacharopoulos. Educational expansion and earnings differentials in Greece. *Comparative Education Review*, 36:52-70, 1992.
10.M. Papadakis. *Techniques of Social Mobility Analysis*, Athens, 1993. In Greek, Vita.
11.S. Prais. Measuring social mobility. *Journal of the Royal Statistical Society*, Series A, 118:56-66, 1955.
12.H. Patrinos. Socioeconomic background schooling, experience, ability and monetary rewards in Greece. *Economics of Education Review*, 14:85-91, 1995.
13.A. Shorrocks. The measurement of social mobility. *Econometrica*, 46:1013-1024, 1978.
14.M. Symeonaki and G. Stamatopoulou. Exploring intergenerational educational mobility in Greece with data drawn from EU-SILC.*International Sociological Association RC04 & RC10 Conference: Social Justice and Participation. The Role of Higher Education*, Cyprus, 2011.
15.M. Symeonaki, G. Stamatopoulou. Exploring the transitions to Higher Education in Greece: issues of intergenerational educational mobility. *Policy Futures in Education* (to appear).
16.M. Symeonaki, G. Stamatopoulou, C. Michalopoulou. Intergenerational Occupational Mobility in Greece: Evidence from EU-SILC. *Proceedings of Demografics Conference*, Crete, Greece, 2012.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Delayed Heston Model: Improvement of the Volatility Surface Fitting and Pricing & Hedging of Volatility Swaps

Anatoliy Swishchuk[1] and Nelson Vadori[2]

[1] Corresponding author. University of Calgary, Department of Mathematics, 2500 University Drive NW, Calgary, AB, Canada T2N 1N4;e-mail: aswish@ucalgary.ca; Tel.: 1(403)220-3274; Fax: 1(403) 282-5150.
[2] University of Calgary, Department of Mathematics, e-mail: nvadori@ucalgary.ca.

Abstract. We present a variance drift adjusted version of the Heston model which leads to significant improvement of the market volatility surface fitting (compared to Heston). The numerical example we performed with recent market data shows a significant (44%) reduction of the average absolute calibration error [1] (calibration on Sep. 30^{th} 2011 for underlying EURUSD). Our model has two additional parameters compared to the Heston model, can be implemented very easily and was initially introduced for volatility derivatives pricing purpose. The main idea behind our model is to take into account some past history of the variance process in its (risk-neutral) diffusion. Using change of time method for continuous local martingales, we derive a closed formula for the Brockhaus&Long approximation of the volatility swap price in this model. We also consider dynamic hedging of volatility swaps using a portfolio of variance swaps.

Keywords: variance swap; volatility swap; stochastic volatility with delay; Heston model with delayed stochastic volatility, change of time; dynamic hedging.

1 Introduction

The volatility process is an important concept in financial modeling as it quantifies at each time t how likely the modeled asset log-return is to vary significantly over some short immediate time period $[t, t+\epsilon]$. This process can be stochastic or deterministic, e.g. local volatility models in which the (deterministic) volatility depends on time and spot price level. In quantitative finance, we often consider the volatility process $\sqrt{V_t}$ (where V_t is the variance process) to be stochastic as it allows to fit the observed vanilla option market prices with an acceptable bias as well as to model the risk linked with the future evolution of the volatility smile (which deterministic model cannot), namely the forward smile. Many derivatives are known to be very sensitive the forward smile, one of the most popular example being the cliquet options (options on future asset performance).

Heston model (Heston [6]; Heston and Nandi [7]) is one of the most popular stochastic volatility models in the industry as semi-closed formulas for vanilla

[1] average of the absolute differences between market and model implied BS volatilities

option prices are available, few (five) parameters need to be calibrated, and it accounts for the mean-reverting feature of the volatility.

One might be willing, in the variance diffusion, to take into account not only its current state but also its past history over some interval $[t-\tau, t]$, where $\tau > 0$ is a constant and is called the delay. Starting from the discrete-time GARCH(1,1) model (Bollerslev [1]), a first attempt was made in this direction in Kazmerchuk et al. [8], where a non-Markov delayed continuous-time GARCH model was proposed (S_t being the asset price at time t, γ, θ, α some positive constants):

$$\frac{dV_t}{dt} = \gamma\theta^2 + \frac{\alpha}{\tau}\ln^2\left(\frac{S_t}{S_{t-\tau}}\right) - (\alpha+\gamma)V_t, \tag{1}$$

this model being inherited from its discrete-time analogue:

$$\sigma_n^2 = \tilde{\gamma}\tilde{\theta}^2 + \frac{\tilde{\alpha}}{L}\ln^2\left(\frac{S_{n-1}}{S_{n-1-L}}\right) + (1-\tilde{\alpha}+\tilde{\gamma})\sigma_{n-1}^2 \tag{2}$$

The parameter θ^2 (resp. γ) can be interpreted as the value of the long-range variance (resp. variance mean-reversion speed) when the delay is equal to 0 (we will see that introducing delay modifies the value of these two model features), and α a continuous-time equivalent of the variance ARCH(1,1) autoregressive coefficient. In fact, we can interpret the right-hand side of previous diffusion equation as the sum of two terms:

- the delay-free term $\gamma(\theta^2 - V_t)$ which accounts for the mean-reverting feature of the variance process
- $\alpha\left(\frac{1}{\tau}\ln^2\left(\frac{S_t}{S_{t-\tau}}\right) - V_t\right)$ which is a pure (noisy) delay term, i.e. that vanishes when $\tau \to 0$ and takes into account the past history of the variance (via the term $\ln\left(\frac{S_t}{S_{t-\tau}}\right)$). The autoregressive coefficient α can be seen as the amplitude of this pure delay term.

J.C. Duan remarked the importance to incorporate the real world $\mathbb{P}-$drift $d_{\mathbb{P}}(t,\tau) := \int_{t-\tau}^{t}(\mu - \frac{1}{2}V_u)du$ of $\ln\left(\frac{S_t}{S_{t-\tau}}\right)$ in the model (where μ stands for the real world $\mathbb{P}-$drift of the stock price S_t), transforming the variance dynamics into:

$$\frac{dV_t}{dt} = \gamma\theta^2 + \frac{\alpha}{\tau}\left[\ln\left(\frac{S_t}{S_{t-\tau}}\right) - d_{\mathbb{P}}(t,\tau)\right]^2 - (\alpha+\gamma)V_t \tag{3}$$

The latter diffusion (3) was introduced in Swishchuk [12] and Kazmerchuk et al. [9], and the proposed model was proved to be complete and to account for the mean-reverting feature of the volatility process. This model is also non Markov as the past history $(V_u)_{u\in[t-\tau,t]}$ of the variance appears in its diffusion equation via the term $\ln\left(\frac{S_t}{S_{t-\tau}}\right)$, as it is shown in Swishchuk [12].

In the continuity of this approach, pricing of variance swaps for one-factor stochastic volatility with delay has been presented in Swishchuk [12], for multi-factor stochastic volatility in Swishchuk [13] and for one-factor stochastic volatility with delay and jumps in Swishchuk and Li [10]. Variance swap for local

Levy-based stochastic volatility with delay has been calculated in Swishchuk and Malenfant [15].

Unfortunately, the model (3) doesn't lead to (semi-)closed formulas for the vanilla options, making it difficult to use for practitioners willing to calibrate on vanilla market prices. Nevertheless, one can notice that the Heston model and the delayed continuous-time GARCH model (3) are very similar in the sense that the expected values of the variances are the same - when we make the delay tend to 0 in (3). As mentioned before, the Heston framework is very convenient, and therefore it is naturally tempting to adjust the Heston dynamics in order to incorporate the delay introduced in (3). In this way, we considered in a first approach adjusting the Heston drift by a deterministic function of time so that the expected value of the variance under the delayed Heston model is equal to the one under the delayed GARCH model (3). In addition to making our delayed Heston framework coherent with (3), this construction makes the variance process diffusion dependent not on its past history $(V_u)_{u\in[t-\tau,t]}$, but on the past history of its risk-neutral expectation $(\mathbb{E}_0^{\mathbb{Q}}(V_u))_{u\in[t-\tau,t]}$, preserving the Markov feature of the Heston model (where we denote $\mathbb{E}_t^{\mathbb{Q}}(\cdot) := \mathbb{E}^{\mathbb{Q}}(\cdot|\mathcal{F}_t)$). The purpose of sections 2 and 3 is to present the Delayed Heston model as well as some calibration results on call option prices, with a comparison to the Heston model. In sections 4 and 5, we will consider the pricing and hedging of volatility and variance swaps in this model.

Volatility and variance swaps are contracts whose payoff depend (respectively convexly and linearly) on the realized variance of the underlying asset over some specified time interval. They provide pure exposure to volatility, and therefore make it a tradable market instrument. Variance Swaps are even considered by some practitioners to be vanilla derivatives. The most commonly traded variance swaps are discretely sampled and have a payoff $P_n^V(T)$ at maturity T of the form:

$$P_n^V(T) = N\left[\frac{252}{n}\sum_{i=0}^{n}\ln^2\left(\frac{S_{i+1}}{S_i}\right) - K_{var}\right]$$

, where S_i is the asset spot price on fixing time $t_i \in [0,T]$ (usually there is one fixing time each day, but there could be more, or less), N the notional amount of the contract (in currency per unit of variance) and K_{var} the strike specified in the contract. The corresponding volatility swap payoff $P_n^v(T)$ is given by:

$$P_n^v(T) = N\left[\sqrt{\frac{252}{n}\sum_{i=0}^{n}\ln^2\left(\frac{S_{i+1}}{S_i}\right)} - K_{vol}\right]$$

One can also consider continuously sampled volatility and variance swaps (on which we will focus in this article), which payoffs are respectively defined as the limit when $n \to +\infty$ of their discretely sampled versions. Formally, if we denote $(V_t)_{t\geq 0}$ the stochastic volatility process of our asset, adapted to some brownian filtration $(\mathcal{F}_t)_{t\geq 0}$, then the continuously-sampled realized variance

V_R from initiation date of the contract $t = 0$ to maturity date $t = T$ is given by $V_R = \frac{1}{T}\int_0^T V_s ds$. The fair variance strike K_{var} is calculated such that the initial value of the contract is 0, and therefore is given by:

$$\mathbb{E}_0^{\mathbb{Q}}\left[e^{-rT}(V_R - K_{var})\right] = 0 \Rightarrow K_{var} = \mathbb{E}_0^{\mathbb{Q}}(V_R)$$

In the same way, the fair volatility strike K_{vol} is given by:

$$\mathbb{E}_0^{\mathbb{Q}}\left[e^{-rT}(\sqrt{V_R} - K_{var})\right] = 0 \Rightarrow K_{vol} = \mathbb{E}_0^{\mathbb{Q}}(\sqrt{V_R})$$

The volatility swap fair strike might be difficult to compute explicitly as we have to compute the expectation of a square-root. In Brockhaus and Long [4], the following approximation - based on Taylor expansion - was proposed to compute the expected value of the square-root of an almost surely non negative random variable Z:

$$\mathbb{E}(\sqrt{Z}) \approx \sqrt{\mathbb{E}(Z)} - \frac{Var(Z)}{8\mathbb{E}(Z)^{\frac{3}{2}}} \tag{4}$$

We will refer to this approximation in our paper as the Brockhaus&Long approximation.

Carr and Lee [5] provides an overview of the current market for volatility derivatives. They survey the early literature on the subject. They also provide relatively simple proofs of some fundamental results related to variance swaps and volatility swaps. Pricing of variance swaps for one-factor stochastic volatility with delay has been presented in Swishchuk [11], for multi-factor stochastic volatility in Swishchuk [13] and for one-factor stochastic volatility with delay and jumps in Swishchuk and Li [10]. Variance swap for local Levy-based stochastic volatility with delay has been calculated in Swishchuk and Malenfant [15]. Variance and volatility swaps in energy markets have been considered in Swishchuk [14]. Broadie and Jain [3] covers pricing and dynamic hedging of volatility derivatives in the Heston model.

The paper is organized as follows: in section 2, we present the Delayed Heston model; in section 3, we present calibration results (for underlying EURUSD on September 30th 2011) as well as a comparison with the Heston model. In section 4, we compute the price process $X_t(T) := \mathbb{E}_t^{\mathbb{Q}}(V_R)$ of the floating leg of the variance swap of maturity T, as well as the Brockhaus&Long approximation of the price process $Y_t(T) := \mathbb{E}_t^{\mathbb{Q}}(\sqrt{V_R})$ of the floating leg of the volatility swap of maturity T. This leads in particular to closed formulas for the fair volatility and variance strikes. In section 5, we consider - in this model - dynamic hedging of volatility swaps using variance swaps.

2 Presentation of the Delayed Heston model

Throughout this paper, we will assume constant risk-free rate r, dividend yield q and finite time-horizon T. We will also denote $\mathbb{E}_t^{\mathbb{Q}}(\cdot) := \mathbb{E}^{\mathbb{Q}}(\cdot|\mathcal{F}_t)$ and

$Var_t^{\mathbb{Q}}(\cdot) := Var^{\mathbb{Q}}(\cdot|\mathcal{F}_t)$

Assume the following risk-neutral $\mathbb{Q}-$ stock price dynamics ($Z_t^{\mathbb{Q}}$ and $W_t^{\mathbb{Q}}$ being two correlated standard brownian motions):

$$dS_t = (r-q)S_t dt + S_t\sqrt{V_t}dZ_t^{\mathbb{Q}}. \tag{5}$$

The well-known Heston model has the following $\mathbb{Q}-$dynamics for the variance V_t:

$$dV_t = \gamma(\theta^2 - V_t)dt + \delta\sqrt{V_t}dW_t^{\mathbb{Q}}, \tag{6}$$

where θ^2 is the long-range variance, γ the variance mean-reversion speed, δ the volatility of the variance and ρ the brownian correlation coefficient ($\left\langle W^{\mathbb{Q}}, Z^{\mathbb{Q}}\right\rangle_t = \rho t$). We also assume $S_0 = s_0$ a.e. and $V_0 = v_0$ a.e., for some positive constants v_0, s_0.

As explained in the introduction, the following delayed continuous-time GARCH dynamics have been introduced for the variance in Swishchuk [12]:

$$\frac{dV_t}{dt} = \gamma\theta^2 + \frac{\alpha}{\tau}\left[\int_{t-\tau}^{t}\sqrt{V_s}dZ_s^{\mathbb{Q}} - (\mu - r)\tau\right]^2 - (\alpha+\gamma)V_t, \tag{7}$$

where μ stands for the real world $\mathbb{P}-$drift of the stock price S_t. We can interpret the right-hand side of previous diffusion equation [2] as the sum of two terms:

- the delay-free term $\gamma(\theta^2 - V_t)$ which accounts for the mean-reverting feature of the variance process
- $\alpha\left(\frac{1}{\tau}\left[\int_{t-\tau}^{t}\sqrt{V_s}dZ_s^{\mathbb{Q}} - (\mu-r)\tau\right]^2 - V_t\right)$ which is a pure (noisy) delay term of amplitude α, i.e. that vanishes when $\tau \to 0$ and takes into account the past history of the variance via the integral $\int_{t-\tau}^{t}\sqrt{V_s}dZ_s^{\mathbb{Q}}$. As we will see below, the introduction of this pure delay term modifies the value of both the long-range variance and variance mean-reversion speed of the model.

We can see that the two models are very similar. Indeed, they both give the same expected value for V_t as the delay goes to 0 in (7), namely $\theta^2 + (V_0 - \theta^2)e^{-\gamma t}$. The idea here is to adjust the Heston dynamics (6) in order to account for the delay introduced in (7). Our approach is to adjust the drift by a deterministic function of time so that the expected value of V_t under the adjusted Heston model is the same as under (7). This approach can be seen as a correction by a pure delay term of amplitude α (in the sense of (15)) of the Heston drift by a deterministic function in order to account for the delay.

[2] note that θ^2 (resp. γ) has been defined in introduction for the delayed continuous-time GARCH model as the value of the long-range variance (resp. variance mean-reversion speed) when $\tau = 0$, therefore it has the same meaning as the Heston long-range variance (resp. variance mean-reversion speed). That is why we use the same notations in both models.

Namely, we assume the adjusted Heston dynamics:

$$dV_t = \left[\gamma(\theta^2 - V_t) + \epsilon_\tau(t)\right] dt + \delta\sqrt{V_t} dW_t^{\mathbb{Q}}, \tag{8}$$

$$\epsilon_\tau(t) := \alpha\tau(\mu - r)^2 + \frac{\alpha}{\tau}\int_{t-\tau}^{t} v_s ds - \alpha v_t, \tag{9}$$

with $v_t := \mathbb{E}_0^{\mathbb{Q}}(V_t)$. It was shown in Swishchuk [12] that v_t solves the following:

$$\frac{dv_t}{dt} = \gamma\theta^2 + \alpha\tau(\mu - r)^2 + \frac{\alpha}{\tau}\int_{t-\tau}^{t} v_s ds - (\alpha + \gamma)v_t. \tag{10}$$

And we have the following expression for v_t:

$$v_t = \theta_\tau^2 + (V_0 - \theta_\tau^2)e^{-\gamma_\tau t}, \tag{11}$$

with:

$$\theta_\tau^2 := \theta^2 + \frac{\alpha\tau(\mu - r)^2}{\gamma}, \tag{12}$$

The parameter θ_τ^2 can be interpreted as the adjusted long-range variance - that has been (positively) shifted from its original value θ^2 because of the introduction of delay. We have $\theta_\tau^2 \to \theta^2$ when $\tau \to 0$, which is coherent. We will see below that we can interpret the parameter $\gamma_\tau > 0$ as the adjusted mean-reversion speed. This parameter is given in Swishchuk [12] by a (nonzero) solution to the following equation:

$$\gamma_\tau = \alpha + \gamma + \frac{\alpha}{\gamma_\tau\tau}(1 - e^{\gamma_\tau\tau}). \tag{13}$$

By (9), (11) and (13) we get an explicit expression for the drift adjustment:

$$\epsilon_\tau(t) = \alpha\tau(\mu - r)^2 + (V_0 - \theta_\tau^2)(\gamma - \gamma_\tau)e^{-\gamma_\tau t} \tag{14}$$

The following simple property gives us some information about the correction term $\epsilon_\tau(t)$ and γ_τ, that will be useful for interpretation purpose and in the derivation of the semi-closed formulas for call options in the next section. Indeed, given (15) and (11), the parameter γ_τ can be interpreted as the adjusted variance mean-reversion speed, and we have by (13) that $\gamma_\tau \to \gamma$ when $\tau \to 0$, which is coherent.

Property 1: γ_τ is the unique solution to (13) and:

$$0 < \gamma_\tau < \gamma, \quad \lim_{\tau\to 0}\sup_{t\in\mathbb{R}^+} |\epsilon_\tau(t)| = 0 \tag{15}$$

Proof: Let's show $\gamma_\tau \geq 0$. If $\gamma_\tau < 0$ then by (13) we have $\alpha+\gamma+\frac{\alpha}{\gamma_\tau\tau}(1-e^{\gamma_\tau\tau}) < 0$, i.e. $1-e^{\gamma_\tau\tau}+\gamma_\tau\tau > -\frac{\gamma}{\alpha}\gamma_\tau\tau$. But $\tau > 0$ so $\exists x_0 > 0$ s.t. $1-e^{-x_0}-x_0 > \frac{\gamma}{\alpha}x_0$. A simple study shows that is impossible whenever $\frac{\gamma}{\alpha} \geq 0$, which is what we have by assumption. Therefore $\gamma_\tau \geq 0$, and in fact $\gamma_\tau > 0$ since it is a nonzero

solution of (13). If $\gamma \leq \gamma_\tau$ then by (13) $\gamma_\tau\tau + 1 - e^{\gamma_\tau\tau} \geq 0$. But $\gamma_\tau\tau > 0$ therefore $\exists x_0 > 0$ s.t. $x_0 + 1 - e^{x0} \geq 0$. A simple study shows that is impossible. The uniqueness comes from a similar simple study. Now, because $\gamma_\tau > 0$, we have $\sup_{t\in\mathbb{R}^+} |\epsilon_\tau(t)| \leq \alpha\tau(\mu-r)^2 + |(V_0-\theta_\tau^2)(\gamma-\gamma_\tau)|$ and $(V_0-\theta_\tau^2)(\gamma-\gamma_\tau) = \circ(1)$ by (13). So $\lim_{\tau\to 0} \alpha\tau(\mu-r)^2 + |(V_0-\theta_\tau^2)(\gamma-\gamma_\tau)| = 0$. $\triangle$

3 Calibration on call option prices and comparison to the Heston model

It is possible to get semi-closed formulas for call options in our delayed Heston model. Indeed, our model is a time-dependent Heston model with time-dependent long-range variance $\tilde{\theta}^2(t) := \theta_\tau^2 + (V_0-\theta_\tau^2)\frac{(\gamma-\gamma_\tau)}{\gamma}e^{-\gamma_\tau t}$. We perform our calibration on September 30^{th} 2011 for underlying EURUSD on the whole volatility surface (maturities from 1M to 10Y, strikes ATM, 25D Call/Put, 10D Call/Put). The implied volatility surface, the Zero Coupon curves EUR Vs. Euribor 6M and USD Vs. Libor 3M and the spot price are taken from Bloomberg (mid prices). The drift $\mu = 0.0188$ is estimated from 7.5Y of daily close prices (source: www.forexrate.co.uk).

The calibrated parameters for delayed Heston are:

$$(V_0, \gamma, \theta^2, \delta, \rho, \alpha, \tau) = (0.0343, 3.9037, 10^{-8}, 0.808, -0.5057, 71.35, 0.7821)$$

and for Heston:

$$(V_0, \gamma, \theta^2, \delta, \rho) = (0.0328, 0.5829, 0.0256, 0.3672, -0.4824)$$

The absolute calibration error (in bp of the BS volatility) for Heston model and our delayed Heston model are given below. The results show a 44% reduction of the average absolute calibration error (46bp for delayed Heston, 81bp for Heston).

	ATM	25D Call	25D Put	10D Call	10D Put
1M	152	192	41	193	67
2M	114	139	15	136	81
3M	89	109	3	110	92
4M	48	61	17	67	101
6M	5	15	34	29	85
9M	59	42	63	2	85
1Y	107	83	102	31	96
1.5Y	141	116	111	42	73
2Y	166	137	127	54	68
3Y	145	124	77	52	0
4Y	96	95	18	37	66
5Y	29	47	52	7	138
7Y	39	10	112	28	186
10Y	100	67	168	58	225

***Table 1**: Heston Absolute Calibration Error (in bp of the BS volatility)*

	ATM	25D Call	25D Put	10D Call	10D Put
1M	116	91	109	128	115
2M	44	24	59	54	88
3M	14	3	32	36	60
4M	18	28	1	5	29
6M	31	37	23	19	3
9M	45	45	56	37	57
1Y	51	47	82	50	104
1.5Y	29	30	79	49	129
2Y	24	23	83	47	139
3Y	11	9	29	30	90
4Y	41	28	14	17	38
5Y	76	55	59	5	16
7Y	71	49	58	1	14
10Y	26	8	18	47	24

Table 2: *Delayed Heston Absolute Calibration Error (in bp of the BS volatility)*

4 Pricing Variance and Volatility Swaps

In this section, we derive a closed formula for the Brockhaus&Long approximation of the volatility swap price using change of time method introduced in Swishchuk [11], as well as the price of the variance swap. Precisely, in Brockhaus and Long [4], the following approximation was presented to compute the expected value of the square-root of an almost surely non negative random variable Z: $\mathbb{E}(\sqrt{Z}) \approx \sqrt{\mathbb{E}(Z)} - \frac{Var(Z)}{8\mathbb{E}(Z)^{\frac{3}{2}}}$. We denote $V_R := \frac{1}{T}\int_0^T V_s ds$ the realized variance on $[0,T]$.

We let $X_t(T) := \mathbb{E}_t^{\mathbb{Q}}(V_R)$ (resp. $Y_t(T) := \mathbb{E}_t^{\mathbb{Q}}(\sqrt{V_R})$) the price process of the floating leg of the variance swap (resp. volatility swap) of maturity T.

Theorem 1: The price process $X_t(T)$ of the floating leg of the variance swap of maturity T in the delayed Heston model (5)-(8) is given by:

$$\begin{aligned} X_t(T) = \frac{1}{T}\int_0^t V_s ds + \frac{T-t}{T}\theta_\tau^2 + (V_t - \theta_\tau^2)\left(\frac{1-e^{-\gamma(T-t)}}{\gamma T}\right) \\ +(V_0 - \theta_\tau^2)e^{-\gamma_\tau t}\left(\frac{1-e^{-\gamma_\tau(T-t)}}{\gamma_\tau T} - \frac{1-e^{-\gamma(T-t)}}{\gamma T}\right) \end{aligned} \tag{16}$$

Proof: By definition, $X_t(T) = \mathbb{E}_t^{\mathbb{Q}}(\frac{1}{T}\int_0^T V_s ds) = \frac{1}{T}\int_0^t V_s ds + \frac{1}{T}\int_t^T \mathbb{E}_t^{\mathbb{Q}}(V_s)ds$. Let $s \geq t$. Then we have by (8) that $\mathbb{E}_t^{\mathbb{Q}}(V_s - V_t) = \mathbb{E}_t^{\mathbb{Q}}(V_s) - V_t = \int_t^s \gamma(\theta^2 - \mathbb{E}_t^{\mathbb{Q}}(V_u)) + \epsilon_\tau(u)du + \mathbb{E}_t^{\mathbb{Q}}(\int_t^s \sqrt{V_u}dW_u^{\mathbb{Q}})$. But $(\sqrt{V_t})_{t\geq 0}$ is an adapted process s.t. $\mathbb{E}^{\mathbb{Q}}(\int_0^T V_u du) < +\infty$, therefore $\int_0^t \sqrt{V_u}dW_u^{\mathbb{Q}}$ is a martingale and we have $\mathbb{E}_t^{\mathbb{Q}}(\int_t^s \sqrt{V_u}dW_u^{\mathbb{Q}}) = 0$. Therefore $\forall s \geq t \geq 0$, the function $s \to \mathbb{E}_t^{\mathbb{Q}}(V_s)$ is a solution of $y_s' = \gamma(\theta^2 - y_s) + \epsilon_\tau(s)$ with initial condition $y_t = V_t$. Simple calculations give us $\mathbb{E}_t^{\mathbb{Q}}(V_s) = \theta_\tau^2 + (V_t - \theta_\tau^2)e^{-\gamma(s-t)} + (V_0 - \theta_\tau^2)e^{-\gamma_\tau t}(e^{-\gamma_\tau(s-t)} - e^{-\gamma(s-t)})$. A calculation of $\int_t^T \mathbb{E}_t^{\mathbb{Q}}(V_s)ds$ completes the proof.

Corollary 1: The price K_{var} of the variance swap of maturity T at initiation of the contract $t = 0$ in the delayed Heston model (5)-(8) is given by:

$$K_{var} = \theta_\tau^2 + (V_0 - \theta_\tau^2)\frac{1-e^{-\gamma_\tau T}}{\gamma_\tau T} \tag{17}$$

Proof: By definition, $K_{var} = X_0(T)$.

Now, let $x_t := -(V_0 - \theta_\tau^2)e^{(\gamma-\gamma_\tau)t} + e^{\gamma t}(V_t - \theta_\tau^2)$. Then by Ito's Lemma we get:

$$dx_t = \delta e^{\gamma t}\sqrt{(x_t + (V_0 - \theta_\tau^2)e^{(\gamma-\gamma_\tau)t})e^{-\gamma t} + \theta_\tau^2}dW_t^{\mathbb{Q}}. \tag{18}$$

Which is of the form $dx_t = f(t, x_t)dW_t^{\mathbb{Q}}$ with:

$$f(t,x) := \delta e^{\gamma t}\sqrt{(x + (V_0 - \theta_\tau^2)e^{(\gamma-\gamma_\tau)t})e^{-\gamma t} + \theta_\tau^2}$$

The process $(x_t)_{t\geq 0}$ is therefore a continuous local martingale, and even a true martingale since $\mathbb{E}^{\mathbb{Q}}(\int_0^T f^2(s, x_s)ds) < \infty$. We can use the change of time method introduced in Swishchuk [11] and we get $x_t = \tilde{W}_{\phi_t}$, where $\tilde{W}_t$ is a $\mathcal{F}_{\phi_t^{-1}}-$ adapted $\mathbb{Q}-$Brownian motion, which is based on the fact that every continuous local martingale can be represented as a time-changed brownian motion. The process $(\phi_t)_{t\geq 0}$ is a.e. increasing, non negative, $\mathcal{F}_t-$ adapted and is called the change of time process. This process is also equal to the quadratic variation $\langle x\rangle_t$ of the (square-integrable) continuous martingale x_t.

Expressions of ϕ_t, ϕ_t^{-1} and $\tilde{W}_t$ are given by:

$$\phi_t = \langle x\rangle_t = \int_0^t f^2(s, x_s)\, ds \tag{19}$$

$$\tilde{W}_t = \int_0^{\phi_t^{-1}} f(s, x_s)dW_s^{\mathbb{Q}} \tag{20}$$

$$\phi_t^{-1} = \int_0^t f^{-2}\left(\phi_s^{-1}, x_{\phi_s^{-1}}\right) ds. \tag{21}$$

This immediately yields:

$$V_t = \theta_\tau^2 + (V_0 - \theta_\tau^2)e^{-\gamma_\tau t} + e^{-\gamma t}\tilde{W}_{\phi_t} \tag{22}$$

Lemma 1:

$$\mathbb{E}_t^{\mathbb{Q}}(\tilde{W}_{\phi_s}) = \tilde{W}_{\phi_{t\wedge s}} \tag{23}$$

And for $s, u \geq t$:

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}(\tilde{W}_{\phi_s}\tilde{W}_{\phi_u}) = x_t^2 + \delta^2 \Bigg[&\theta_\tau^2\left(\frac{e^{2\gamma(s\wedge u)} - e^{2\gamma t}}{2\gamma}\right) \\ &+(V_0 - \theta_\tau^2)\left(\frac{e^{(2\gamma-\gamma_\tau)(s\wedge u)} - e^{(2\gamma-\gamma_\tau)t}}{2\gamma - \gamma_\tau}\right) + x_t\left(\frac{e^{\gamma(s\wedge u)} - e^{\gamma t}}{\gamma}\right)\Bigg]\end{aligned} \tag{24}$$

Proof: (23) comes from the fact that $x_t = \tilde{W}_{\phi_t}$ is a martingale. Let $s \geq u \geq t$. Then by iterated conditioning: $\mathbb{E}_t^{\mathbb{Q}}(\tilde{W}_{\phi_s}\tilde{W}_{\phi_u}) = \mathbb{E}_t^{\mathbb{Q}}(\mathbb{E}_u^{\mathbb{Q}}(\tilde{W}_{\phi_s}\tilde{W}_{\phi_u})) = \mathbb{E}_t^{\mathbb{Q}}(\tilde{W}_{\phi_u}\mathbb{E}_u^{\mathbb{Q}}(\tilde{W}_{\phi_s})) = \mathbb{E}_t^{\mathbb{Q}}(\tilde{W}_{\phi_u}^2)$, because $x_t = \tilde{W}_{\phi_t}$ is a martingale. Now, by definition of the quadratic variation, $x_u^2 - \langle x \rangle_u$ is a martingale and therefore $\mathbb{E}_t^{\mathbb{Q}}(\tilde{W}_{\phi_u}^2) = x_t^2 - \langle x \rangle_t + \mathbb{E}_t^{\mathbb{Q}}(\langle x \rangle_u) = x_t^2 - \phi_t + \mathbb{E}_t^{\mathbb{Q}}(\phi_u) = x_t^2 - \phi_t + \phi_t + \mathbb{E}_t^{\mathbb{Q}}(\int_t^u f^2(s, x_s)\, ds)$. By definiton of $f^2(s, x_s)$ and since x_t martingale, then we have (for $s \geq t$) $\mathbb{E}_t^{\mathbb{Q}}(f^2(s, x_s)) = f^2(s, x_t)$, and so that $\mathbb{E}_t^{\mathbb{Q}}(\tilde{W}_{\phi_s}\tilde{W}_{\phi_u}) = x_t^2 + \int_t^u f^2(s, x_t)\, ds$. A simple integration completes the proof.

The following theorem gives the expression of the Brockhaus&Long approximation of the volatility swap floating leg price process $Y_t(T)$.

Theorem 2: The Brockhaus&Long approximation of the price process $Y_t(T)$ of the floating leg of the volatility swap of maturity T in the delayed Heston model (5)-(8) is given by:

$$Y_t(T) \approx \sqrt{X_t(T)} - \frac{Var_t^{\mathbb{Q}}(V_R)}{8X_t(T)^{\frac{3}{2}}} \tag{25}$$

where $X_t(T)$ is given by Theorem 1 and:

$$\begin{aligned} Var_t^{\mathbb{Q}}(V_R) = {} & \frac{x_t\delta^2}{\gamma^3 T^2}\left[e^{-\gamma t}\left(1 - e^{-2\gamma(T-t)}\right) - 2(T-t)\gamma e^{-\gamma T}\right] \\ & + \frac{\delta^2}{2\gamma^3 T^2}\left[2\theta_\tau^2\gamma(T-t) + 2(V_0 - \theta_\tau^2)\frac{\gamma}{\gamma_\tau}e^{-\gamma_\tau t} + 4\theta_\tau^2 e^{-\gamma(T-t)} - \theta_\tau^2 e^{-2\gamma(T-t)} - 3\theta_\tau^2\right] \\ & - \frac{\delta^2(V_0 - \theta_\tau^2)}{\gamma^2 T^2(\gamma_\tau^2 + 2\gamma^2 - 3\gamma\gamma_\tau)}\left[2(\gamma_\tau - 2\gamma)e^{-\gamma(T-t)-\gamma_\tau t}\right. \\ & \left. + (\gamma - \gamma_\tau)e^{-2\gamma(T-t)-\gamma_\tau t} + 2\frac{\gamma^2}{\gamma_\tau}e^{-\gamma_\tau T}\right] \end{aligned} \tag{26}$$

Proof: The (conditioned) Brockhaus&Long approximation gives us:

$$Y_t(T) = \mathbb{E}_t^{\mathbb{Q}}(\sqrt{V_R}) \approx \sqrt{\mathbb{E}_t^{\mathbb{Q}}(V_R)} - \frac{Var_t^{\mathbb{Q}}(V_R)}{8\mathbb{E}_t^{\mathbb{Q}}(V_R)^{\frac{3}{2}}} = \sqrt{X_t(T)} - \frac{Var_t^{\mathbb{Q}}(V_R)}{8X_t(T)^{\frac{3}{2}}}$$

Furthermore:

$$\begin{aligned} Var_t^{\mathbb{Q}}(V_R) &= \mathbb{E}_t^{\mathbb{Q}}((V_R - \mathbb{E}_t^{\mathbb{Q}}(V_R))^2) \\ &= \frac{1}{T^2}\mathbb{E}_t^{\mathbb{Q}}\left(\left(\int_0^T (V_s - \mathbb{E}_t^{\mathbb{Q}}(V_s))ds\right)^2\right) \end{aligned} \tag{27}$$

From (22) we have $V_t = \theta_\tau^2 + (V_0 - \theta_\tau^2)e^{-\gamma_\tau t} + e^{-\gamma t}\tilde{W}_{\phi_t}$, and since $\tilde{W}_{\phi_t}$ is a martingale, $V_s - \mathbb{E}_t^{\mathbb{Q}}(V_s) = 0$ if $s \leq t$, and $V_s - \mathbb{E}_t^{\mathbb{Q}}(V_s) = e^{-\gamma s}(\tilde{W}_{\phi_s} - x_t)$ if $s > t$.

Therefore

$$\begin{aligned} Var_t^{\mathbb{Q}}(V_R) &= \frac{1}{T^2}\mathbb{E}_t^{\mathbb{Q}}\left(\left(\int_t^T e^{-\gamma s}(\tilde{W}_{\phi_s} - x_t)ds\right)^2\right) \\ &= \frac{1}{T^2}x_t^2\left(\int_t^T e^{-\gamma s}ds\right)^2 + \frac{1}{T^2}\mathbb{E}_t^{\mathbb{Q}}\left(\left(\int_t^T e^{-\gamma s}\tilde{W}_{\phi_s}ds\right)^2\right) \\ &- \frac{2}{T^2}x_t\left(\int_t^T e^{-\gamma s}\mathbb{E}_t^{\mathbb{Q}}(\tilde{W}_{\phi_s})ds\right)\left(\int_t^T e^{-\gamma s}ds\right) \\ &= -\frac{1}{T^2}x_t^2\left(\int_t^T e^{-\gamma s}ds\right)^2 + \frac{1}{T^2}\mathbb{E}_t^{\mathbb{Q}}\left(\left(\int_t^T e^{-\gamma s}\tilde{W}_{\phi_s}ds\right)^2\right) \\ &= \frac{1}{T^2}\int_t^T\int_t^T e^{-\gamma(s+u)}\mathbb{E}_t^{\mathbb{Q}}(\tilde{W}_{\phi_s}\tilde{W}_{\phi_u})dsdu - \frac{1}{T^2}x_t^2e^{-2\gamma t}\left(\frac{1-e^{-\gamma(T-t)}}{\gamma}\right)^2 \end{aligned} \tag{28}$$

Lemma 1 and some straightforward computations complete the proof.

Corollary 2: The Brockhaus&Long approximation of the volatility swap price K_{vol} of maturity T at initiation of the contract $t = 0$ in the delayed Heston model (5)-(8) is given by:

$$K_{vol} \approx \sqrt{K_{var}} - \frac{Var^{\mathbb{Q}}(V_R)}{8K_{var}^{\frac{3}{2}}} \tag{29}$$

where K_{var} is given by Corollary 1 and:

$$\begin{aligned} Var^{\mathbb{Q}}(V_R) = \frac{\delta^2 e^{-2\gamma T}}{2T^2\gamma^3}\Big[&\theta_\tau^2\left(2\gamma T e^{2\gamma T} + 4e^{\gamma T} - 3e^{2\gamma T} - 1\right) + \frac{\gamma}{2\gamma - \gamma_\tau}(V_0 - \theta_\tau^2) \\ &\left(2e^{2\gamma T}\left(2\frac{\gamma}{\gamma_\tau} - 1\right) - 4\gamma e^{\gamma T}\left(\frac{e^{(\gamma-\gamma_\tau)T} - 1}{\gamma - \gamma_\tau}\right) + 4e^{\gamma T}\left(1 - \frac{\gamma}{\gamma_\tau}e^{(\gamma-\gamma_\tau)T}\right) - 2\right)\Big] \end{aligned} \tag{30}$$

We notice that letting $\tau \to 0$ (and therefore $\gamma_\tau \to \gamma$) we get the formula of Swishchuk [11].

Proof: We have by definition $K_{vol} = Y_0(T)$, and straightforward computations using theorem 2 finish the proof.

5 Volatility Swap Hedging

In this section, we consider dynamic hedging of volatility swaps using variances swap. In the spirit of Broadie and Jain [3], we consider a portfolio containing at time t one unit of volatility swap and β_t units of variance swaps, both of maturity T. Therefore the value Π_t of the portfolio at time t is:

$$\Pi_t = e^{-r(T-t)}\left[Y_t(T) - K_{vol} + \beta_t(X_t(T) - K_{var})\right] \tag{31}$$

The portfolio is self-financing, therefore:

$$d\Pi_t = r\Pi_t dt + e^{-r(T-t)}\left[dY_t(T) + \beta_t dX_t(T)\right] \tag{32}$$

The price processes $X_t(T)$ and $Y_t(T)$ can be expressed, denoting $I_t := \int_0^t V_s ds$ the accumulated variance at time t (known at this time):

$$X_t(T) = \mathbb{E}_t^{\mathbb{Q}}\left[\frac{1}{T}I_t + \frac{1}{T}\int_t^T V_s ds\right] = g(t, I_t, V_t) \tag{33}$$

$$Y_t(T) = \mathbb{E}_t^{\mathbb{Q}}\left[\sqrt{\frac{1}{T}I_t + \frac{1}{T}\int_t^T V_s ds}\right] = h(t, I_t, V_t) \tag{34}$$

Letting $\tilde{\theta}_t^2 := \theta_\tau^2 + (V_0 - \theta_\tau^2)\frac{(\gamma-\gamma_\tau)}{\gamma}e^{-\gamma_\tau t}$ and noticing that $dI_t = V_t dt$, by Ito's lemma we get:

$$dX_t(T) = \left[\frac{\partial g}{\partial t} + \frac{\partial g}{\partial I_t}V_t + \frac{\partial g}{\partial V_t}\gamma(\tilde{\theta}_t^2 - V_t) + \frac{1}{2}\frac{\partial^2 g}{\partial V_t^2}\delta^2 V_t\right]dt + \frac{\partial g}{\partial V_t}\delta\sqrt{V_t}dW_t^{\mathbb{Q}} \tag{35}$$

$$dY_t(T) = \left[\frac{\partial h}{\partial t} + \frac{\partial h}{\partial I_t}V_t + \frac{\partial h}{\partial V_t}\gamma(\tilde{\theta}_t^2 - V_t) + \frac{1}{2}\frac{\partial^2 h}{\partial V_t^2}\delta^2 V_t\right]dt + \frac{\partial h}{\partial V_t}\delta\sqrt{V_t}dW_t^{\mathbb{Q}} \tag{36}$$

As conditional expectations of cashflows at maturity of the contract, the price processes $X_t(T)$ and $Y_t(T)$ are by construction martingales, and therefore we should have:

$$\frac{\partial g}{\partial t} + \frac{\partial g}{\partial I_t}V_t + \frac{\partial g}{\partial V_t}\gamma(\tilde{\theta}_t^2 - V_t) + \frac{1}{2}\frac{\partial^2 g}{\partial V_t^2}\delta^2 V_t = 0 \tag{37}$$

$$\frac{\partial h}{\partial t} + \frac{\partial h}{\partial I_t}V_t + \frac{\partial h}{\partial V_t}\gamma(\tilde{\theta}_t^2 - V_t) + \frac{1}{2}\frac{\partial^2 h}{\partial V_t^2}\delta^2 V_t = 0 \tag{38}$$

The second equation, combined with some appropriate boundary conditions, was used in Broadie and Jain [3] to compute the value of the price process $Y_t(T)$, whereas we focus on its Brockhaus&Long approximation.

Therefore we get:

$$dX_t(T) = \frac{\partial g}{\partial V_t}\delta\sqrt{V_t}dW_t^{\mathbb{Q}} \tag{39}$$

$$dY_t(T) = \frac{\partial h}{\partial V_t}\delta\sqrt{V_t}dW_t^{\mathbb{Q}} \tag{40}$$

and so:

$$d\Pi_t = r\Pi_t dt + e^{-r(T-t)}\left[\frac{\partial h}{\partial V_t}\delta\sqrt{V_t}dW_t^{\mathbb{Q}} + \beta_t\frac{\partial g}{\partial V_t}\delta\sqrt{V_t}dW_t^{\mathbb{Q}}\right] \tag{41}$$

In order to dynamically hedge a volatility swap of maturity T, one should therefore hold β_t units of variance swap of maturity T, with:

$$\beta_t = -\frac{\frac{\partial h}{\partial V_t}}{\frac{\partial g}{\partial V_t}} = -\frac{\frac{\partial Y_t(T)}{\partial V_t}}{\frac{\partial X_t(T)}{\partial V_t}} \tag{42}$$

The initial hedge ratio β_0 is given by ($Var^{\mathbb{Q}}(V_R)$, K_{var} being given resp. in Corollary 2 and 1):

$$\beta_0 = -\frac{\frac{\partial Y_0(T)}{\partial V_0}}{\frac{\partial X_0(T)}{\partial V_0}} \tag{43}$$

$$\frac{\partial X_0(T)}{\partial V_0} = \frac{1-e^{-\gamma_\tau T}}{\gamma_\tau T} \tag{44}$$

$$\frac{\partial Y_0(T)}{\partial V_0} \approx \frac{\frac{\partial X_0(T)}{\partial V_0}}{2\sqrt{K_{var}}} - \frac{K_{var}\frac{\partial Var^{\mathbb{Q}}(V_R)}{\partial V_0} - \frac{3}{2}\frac{\partial X_0(T)}{\partial V_0}Var^{\mathbb{Q}}(V_R)}{8K_{var}^{\frac{5}{2}}} \tag{45}$$

$$\begin{aligned}\frac{\partial Var^{\mathbb{Q}}(V_R)}{\partial V_0} = \frac{\delta^2 e^{-2\gamma T}}{T^2\gamma^3}\frac{\gamma}{2\gamma-\gamma_\tau}\left[e^{2\gamma T}\left(2\frac{\gamma}{\gamma_\tau}-1\right)\right.\\ \left.-2\gamma e^{\gamma T}\left(\frac{e^{(\gamma-\gamma_\tau)T}-1}{\gamma-\gamma_\tau}\right)+2e^{\gamma T}\left(1-\frac{\gamma}{\gamma_\tau}e^{(\gamma-\gamma_\tau)T}\right)-1\right]\end{aligned} \tag{46}$$

The hedge ratio β_t for $t>0$ is given by ($Var_t^{\mathbb{Q}}(V_R)$, $X_t(T)$ being given resp. in Theorem 2 and 1):

$$\beta_t = -\frac{\frac{\partial Y_t(T)}{\partial V_t}}{\frac{\partial X_t(T)}{\partial V_t}} \tag{47}$$

$$\frac{\partial X_t(T)}{\partial V_t} = \frac{1-e^{-\gamma(T-t)}}{\gamma T} \tag{48}$$

$$\frac{\partial Y_t(T)}{\partial V_t} \approx \frac{\frac{\partial X_t(T)}{\partial V_t}}{2\sqrt{X_t(T)}} - \frac{X_t(T)\frac{\partial Var_t^{\mathbb{Q}}(V_R)}{\partial V_t} - \frac{3}{2}\frac{\partial X_t(T)}{\partial V_t}Var_t^{\mathbb{Q}}(V_R)}{8X_t(T)^{\frac{5}{2}}} \tag{49}$$

$$\frac{\partial Var_t^{\mathbb{Q}}(V_R)}{\partial V_t} = \frac{\delta^2}{\gamma^3 T^2}\left[1-e^{-2\gamma(T-t)}-2(T-t)\gamma e^{-\gamma(T-t)}\right] \tag{50}$$

We take the parameters that have been calibrated in section 3 and we plot the naive Volatility Swap strike $\sqrt{K_{var}}$ and the adjusted Volatility Swap strike $\sqrt{K_{var}} - \frac{Var^{\mathbb{Q}}(V_R)}{8K_{var}^{\frac{3}{2}}}$ along the maturity dimension, as well as the convexity adjustment $\frac{Var^{\mathbb{Q}}(V_R)}{8K_{var}^{\frac{3}{2}}}$:

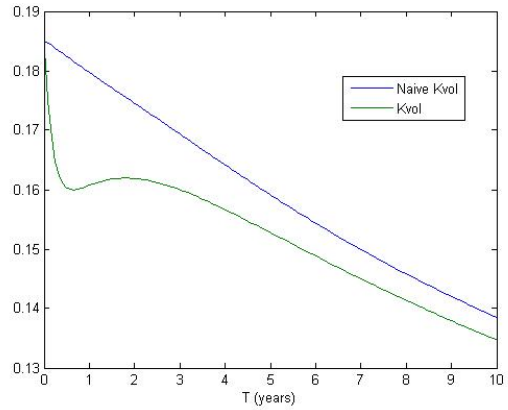


Figure 1: *Naive Volatility Swap Strike Vs. Adjusted Volatility Swap Strike*

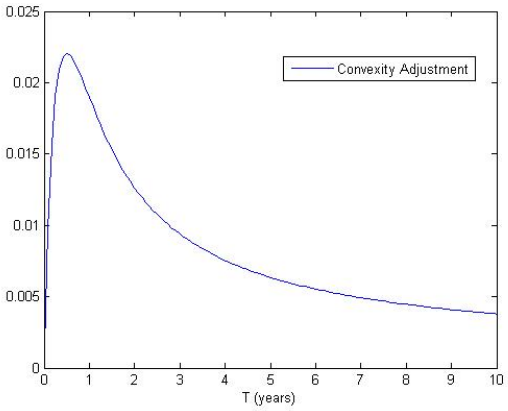


Figure 2: *Convexity Adjustment*

We also plot the initial hedge ratio β_0 along the maturity dimension.

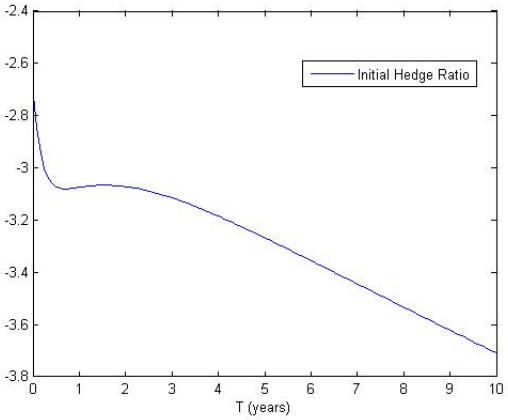


Figure 3: *Initial Hedge Ratio*

References

1.Bollerslev, T. Generalized autoregressive conditional heteroscedasticity. *J. Economics* 31, 307-27, 1986.
2.Broadie, M. and Jain, A. The effect of jumps and discrete sampling on volatility and variance swaps. *Intern. J. Theor. Appl. Finance*, vol. 11, no. 8, pp. 761-797, 2008.
3.Broadie, M and Jain, A. Pricing and Hedging Volatility Derivatives. *Journal of Derivatives*, Vol.15, No.3, 7-24, 2008.
4.Brockhaus, O. and Long, D. Volatility swaps made simple. *Risk*, January, 92-96, 2000.
5.Carr, P. and Lee, R. Volatility Derivatives. *Annu. Rev, Financ. Econ.*, 1:1-21 (doi:10.1146/annurev.financial.050808.114304), 2009.
6.Heston, S. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6, 327-343, 1993.
7.Heston, S. and Nandi, S. Derivatives on volatility: some simple solutions based on observables. *Federal Reserve Bank of Atlanta*, working paper 2000-20, November, 2000.
8.Kazmerchuk, Y., Swishchuk, A. and Wu, J. The pricing of options for security markets with delayed response. *Mathematics and Computers in Simulation* 75 6979, 2007.
9.Kazmerchuk, Y., Swishchuk, A. and Wu, J. A continuous-time GARCH model for stochastic volatility with delay. *Candian Appl. Math. Quart.*, v. 13, No. 2, 2005.
10.Swishchuk, A. and Li, X. Pricing variance swaps for stochastic volatilities with delay and jumps, *Intern. J. Stoch. Anal.*, vol. 2011 (2011), Article ID 435145, 2011.
11.Swishchuk, A. Modeling of variance and volatility swaps for financial markets with Stochastic volatilities. *Wilmott Magazine*, September issue, No. 2, pp. 64-72, 2004.
12.Swishchuk, A. Modeling and pricing of variance swaps for stochastic volatilities with delay. *Wilmott Magazine*, September, No. 19, pp. 63-73, 2005.
13.Swishchuk, A. Modeling and Pricing of Variance Swaps for Multi-Factor Stochastic Volatilities with Delay. *Canad. Appl. Math. Quart.*, Volume 14, Number 4, Winter 2006.
14.Swishchuk, A. Variance and volatility swaps in energy markets. *J. Energy Markets*, Volume 6/Number 1, Spring 2013 (3349), 2013.
15.Swishchuk, A and Malenfant, K. Variance swap for local Levy based stochastic volatility with delay. *Intern. Rev. Appl. Fin. Iss. Econ. (IRAFIE)*, Vol. 3, No. 2, 2011, pp. 432-441, 2011.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Intergenerational mobility as a distance measure between probability distribution functions

M. Symeonaki and G. Stamatopoulou

Department of Social Policy, Panteion University of Social and Political Sciences,
136 Syggrou Av., 176 71, Athens, Greece
(E-mail: msymeonaki@me.com, gl.stamat@gmail.com)

Abstract. This paper presents a methodology of estimating social intergenerational mobility as a distance or similarity measure between the parent's probability distribution function and the sibling's. Several distance and similarity measures are provided and their properties are discussed. An illustration of the methodology is presented providing the measurement of the intergenerational occupational mobility with evidence drawn from the Survey of Health, Ageing and Retirement in Europe (SHARE project), and more specifically from SHARELIFE release 1.0, concerning Greece[1].
Keywords: Intergenerational mobility, distance and similarity measures, SHARE project, occupational intergenerational mobility.

1 Introduction

The measurement of intergenerational mobility is of utmost important for the Social and Economical Sciences as it reveals information about the existence of social inequalities, the intergenerational transmission of poverty, social inclusion etc providing a measurement of the extent to which positions, social status or socio-economical thesis change from one generation to another.

There are many approaches to the measurement of intergenerational social mobility. The statistical approach (see for example Bartholomew [3], Bartholomew *et al.* [2] and Boudon [5] among others), where the concept of transition matrices is used and social mobility is studied by examining the

[1] This paper uses data from SHARE wave 4 release 1, as of November 30th 2012. The SHARE data collection has been primarily funded by the European Commission through the 5th Framework Programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life), through the 6th Framework Programme (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5- CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th Framework Programme (SHARE-PREP, No 211909, SHARE-LEAP, No 227822 and SHARE M4, No 261982). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 and OGHA 04-064) and the German Ministry of Education and Research as well as from various national sources is gratefully acknowledged (see www.share-project.org for a full list of funding institutions).
This study was supported in part by a grand from THALIS-Panteion University-Investigating Crucial Interdisciplinary Linkages in Ageing Societies

mathematical properties of these transition matrices. A number of mobility indices based on the elements of the transition probabilities are proposed in the context of the transition matrix approach [13,11,3]. Other approaches based on the estimation of transition probabilities and on empirical data can be found in Symeonaki *et al.* [14–16].

On the other hand, there is the distance approach (see for example Cowell [6] and Fields and Ok [7]). The concepts of distance and similarity are essential in both abstract and applied sciences and consequently distance measures have become an important tool in many areas including Probability and Statistics, Computer Science, Social Sciences, Pattern Recognition, Image Processing, etc. In the literature, distances, dissimilarity and similarity measures, or similarity coefficients are used to express quantitatively the similarity or dissimilarity of two data points, objects, clusters, distributions, samples, etc. Similarity measures and similarity coefficients are used to describe how similar two data points (clusters, objects, samples, distributions, etc) are, whereas distance measures or dissimilarity measures are used to examine how dissimilar two data points (clusters, objects, samples, distributions, etc) are ([9]). Many measures of similarity or dissimilarity have been proposed in the literature. Understanding the relationship between different distance measures is helpful in choosing a proper one for a particular application.

In the study of social mobility, the distance approach covers only the income intergenerational mobility, where the continuous variable of income is taken into account. Therefore, this leaves out the comparison of discrete probability distribution functions and moreover it uses income as the means of stratification. This is not always the best way, since in most cases we cannot have access to valid income data. In the present paper, we propose alternatively the measurement of intergenerational occupational mobility and provide a methodology of measuring mobility with the aid of distance or similarity measures. More specifically, we provide distance and similarity measures that can account for the dissimilarity between the parent's occupational distribution and the respective distribution of the sibling. In this case, we are actually looking for a distance or similarity measure, between two discrete probability distributions. In order to illustrate the methodology, we use data drawn from the Survey of Health, Ageing and Retirement in Europe (SHARE project) and more specifically data drawn from SHARELIFE release 1.0, concerning Greece in respect to the respondent's occupation and the occupation of main breadwinner.

The paper has been organised in the following way. Section 2 provides the preliminaries and the notation employed in the paper, whereas Section 3, presents the distance and similarity measures of occupational intergenerational mobility. Having presented the measures the similarity and distance measures between the respondent and main breadwinner in respect to their occupation are estimated for Greece, with data drawn from SHARELIFE release 1.0, in order to facilitate the theoretical aspects of Section 3. Section 5 summarises the conclusions of our analysis.

2 Preliminaries and notation

The mathematical notion of (metric) distance introduced in Frechet [8] and Hausdorff [10] is given by the following Definitions.

Definition 1. Let X be a set. A distance on X is a function $d : X \times X \to R$, where R is the set of real numbers, if and only if it satisfies the following conditions, $\forall x, y \in X$:

1. $d(x, y) \geq 0$, (non-negativity or positivity axiom),
2. $d(x, y) = 0$, if and only if $x = y$, (reflexivity axiom), and
3. $d(x, y) = d(y, x)$, (symmetry axiom).

Definition 2. Let X be a set. A metric distance on X is a function $d : X \times X \to R$, where R is the set of real numbers, if and only if it satisfies the following conditions, $\forall x, y \in X$:

1. $d(x, y) \geq 0$, (non-negativity or positivity axiom),
2. $d(x, y) = 0$, if and only if $x = y$, (identity axiom),
3. $d(x, y) = d(y, x)$, (symmetry axiom), and
4. $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

Table 1 provides a summary of the basic notation used in the remaining of the paper.

Table 1. Summary of the notation used

Notation	Description
d	denotes the number of occupational classes, i.e. the occupational space, denoted by S is equal to $S = \{1, 2, ..., d\}$,
x_i	is the number of respondents in the $i-th$ occupational class,
f_i	is the number of main breadwinners in the $i-th$ occupational class,
x	is the total number of respondents, i.e. $x = \sum_{i=1}^{d} x_i$,
f	is the total number of main breadwinners, i.e. $f = \sum_{i=1}^{d} f_i$,
p_i^x	denotes the proportion of respondents in $i-th$ occupational level, i.e. $p_i^x = \frac{x_i}{x}$,
p_i^f	denotes the proportion of fathers in $i-th$ occupational level, i.e. $p_i^f = \frac{f_i}{f}$,
$\mathbf{p}_x$	denotes the vector $\mathbf{p}_x = [p_1^x, p_2^x, ..., p_d^x]$, and
$\mathbf{p}_f$	denotes the vector $\mathbf{p}_f = [p_1^f, p_2^f, ..., p_d^f]$.

3 Intergenerational mobility based on distance and similarity measures

In the present section, following the notation introduced in Section 2 we provide a number of distance and similarity measures, which will be used in the measurement of intergenerational mobility of occupation. The most widely known distance between two points (clusters, objects, etc) is the Euclidean distance. The Euclidean distance between the vectors $\mathbf{p}_f$ and $\mathbf{p}_x$ is given by Equation (1):

$$d_{Euc} = \sqrt{\sum_{i=1}^{d} |p_i^f - p_i^x|^2}. \tag{1}$$

In Manhattan distance (Equation (2)) the distance between two points is calculated as the sum of the absolute differences of their coordinates. Its true meaning, therefore, lies in the fact that it represents the total proportional variation of parents and siblings when their distribution to occupational classes is considered.

$$d_{Man} = \sum_{i=1}^{d} |p_i^f - p_i^x|. \tag{2}$$

The Minkowski distance (Equation (3)), or L_p, is a metric on Euclidean space which can be considered as a generalization of both the Euclidean distance and the Manhattan distance. It is also true that the higher the value of p, the greater the importance given to large differences. Therefore, deciding upon an appropriate value of p, comes to the emphasis that one would like to give to the larger differences of proportions to the occupational classes between siblings and parents.

$$d_{Mk} = \sqrt[p]{\sum_{i=1}^{d} |p_i^f - p_i^x|^p}. \tag{3}$$

The Sorensen index (Equation (4)), also known as Sorensen's similarity coefficient, is a statistical index used for comparing the similarity of two samples.

$$d_{Sor} = \frac{\sum_{i=1}^{d} |p_i^f - p_i^x|}{\sum_{i=1}^{d} (p_i^f + p_i^x)} = \frac{\sum_{i=1}^{d} |p_i^f - p_i^x|}{2}. \tag{4}$$

It is obvious from Equation (2) that $d_{Sor} = \frac{d_{Man}}{2}$ and is usually used as a measure for gender segregation.

Gower's similarity index is given by Equation (5).

$$d_{Gow} = \frac{\sum_{i=1}^{d} |p_i^f - p_i^x|}{d}. \tag{5}$$

Chebyshev distance, or L^∞ metric is a metric on a vector space where the distance between two vectors is the greatest of their differences along any

coordinate dimension. Chebychev distance considers only the part for which the difference is maximum, while Manhattan distance gives equal importance to all differences. Therefore, we could assume that Chebychev distance represents the maximum proportional variation of the occupational distribution between parents and siblings.

$$d_{Cheb} = \max_i |p_i^f - p_i^m|. \quad (6)$$

4 Distance and similarity measures of intergenerational mobility with evidence from the SHARELIFE release 1.0.

The SHARE project[1,4] is a multidisciplinary, longitudinal and cross-national panel database, developed to understand the relations between health, labour force participation and institutional context of old people support in Europe. Funding mainly via the European Commission, as well as the US National Institute on Ageing and national sources, it is designed in January of 2002 and it is conducted every two years. The first wave of the survey took place in 2004–2005, in 11 European countries ranging from Nordic to Mediterranean countries, while the fourth wave took place in 2010-2011. The main purpose of survey is to provide a full picture of all aspects of ageing process and its impact in the different cultures of Europe. Moreover, using the knowledge of its predecessors, the US Health and Retirement Study (HRS) and the English Longitudinal Survey on Ageing (ELSA), the survey aims to collect comparable data useful for the policies planned and applied in the European Union. In order to succeed its purposes, the survey divided into 21 modules. Except for the coverscreen (CV) and the demographic (DN) modules, it covers a large variety of subjects, such as physical and mental health, behavioural risks, employment and pensions, social support etc. The target population of the survey is all the non-institutionalised population aged more than 50 years old, as well as their spouse, regardless of their age.

SHARELIFE [12] is the third wave of data collection for SHARE, which focuses on people's life histories. SHARELIFE assembled more comprehensive information on significant areas of our respondents lives, including partners and children, housing and work history, detailed questions on health and health care, etc. SHARELIFE thus complements the SHARE panel data by providing life history information to enhance the understanding of how early life experiences and events throughout life influence the circumstances of older people. With this variety SHARELIFE constitutes an unique cross-national, interdisciplinary database for research in the fields of sociology, economics, gerontology, and demography. SHARELIFE links individual micro data over the respondents entire life with institutional macro data on the welfare state. It thereby allows assessing the full effect of welfare state interventions on the life of the individual. Changes in institutional settings that influence individual decisions are of specific interest to evaluate policies throughout Europe. SHARELIFE follows a Life History Calendar (LHC) approach, which has been

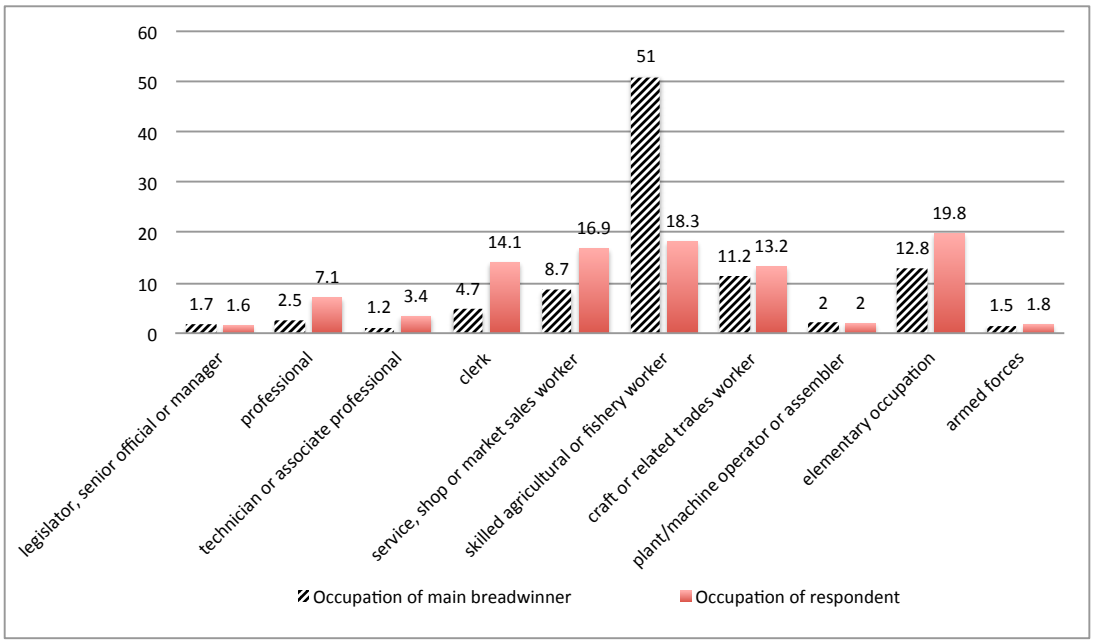


Fig. 1. Distribution of job description between main breadwinner and respondent. [2] *Source: SHARELIFE, release 1.0*, $N = 2,938$.

designed to help respondents in remembering past events more accurately. Using the life history calendar technique has been shown to improve the accuracy of the retrospective information given by respondents [12]. In the present section, we provide the experimental results obtained by using the similarity and distance measures of Section 3, with data drawn from SHARELIFE release 1.0 (SHARE, wave 3) for Greece, in respect to the respondent's occupation and the occupation of main breadwinner.

In Figure 1 the probability distributions of the main breadwinner and the respondent are exhibited. The categorisation used is the one provided in Table 2. The shortest distance, i.e. the Euclidean distance, for example will be equal to $d_{Euc} = 0.361$. Other distance measures can be found in Table 3 and Table 4. In the case of Chebyshev distance, only the maximum difference is estimated, and therefore $d_{Cheb} = 0.327$, being the difference in the proportions in the $6-$th occupational class, which is actually a considerable distance, caused by a notable drop in the percentage of individuals in the Agricultural or Fishery occupations.

5 Conclusions

The justification for using distance and similarity measures or similarity coefficients as a measure of intergenerational (or intragenerational) social mobility is established upon the understanding that high mobility indicates different distribution probabilities between siblings and parents (or different distributions

Table 2. Respondent's and main breadwinner's job description

Field of subject, highest qualification	**Main breadwinner**	**Respondent**
Legislator, senior official or manager	1.7	1.6
Professional	2.5	7.1
Technician or associate professional	1.2	3.4
Clerk	4.7	14.1
Service, shop or market sales worker	8.7	16.9
Skilled agricultural or fishery worker	51.0	18.3
Craft or related trades worker	11.2	13.2
Plant/machine operator or assembler	2.0	2.0
Elementary occupation	12.8	19.8
Armed forces	1.5	1.8
Total	100.0	100.0

[3] *Source: SHARELIFE, release 1.0*, $N = 2,938$.

Table 3. Distance and similarity measures between main breadwinners and respondents

1.	Euclidean distance	$d_{Euc} = 0.665$
2.	Manhattan distance	$d_{Man} = 0.361$
3.	Sorensen index	$d_{Mk} = 0.333$
4.	Gower simmilarity index	$d_{Gow} = 0.067$
5.	Chebyshev distance	$d_{Cheb} = 0.327$

[4] *Source: SHARELIFE, release 1.0*, $N = 2,938$.

during one's lifetime) across the occupational, educational or social classes; the more different the distributions, the more mobile individuals are in respect to the social ladder, the bigger the distance measures (or the smaller the similarity measures). Therefore, it is important to consider distance measures, to understand the meaning of their values and to be able to choose between them.

References

1.C. Barang, V. Eudier, N. Sirven. SHARE, the Survey of Health, Ageing, and Retirement in Europe goes longitudinal. *Issues in health economics*, n 137, Institute for research and information in health economics, IRDES, 2008.
2.D. J. Bartholomew, A. F. Forbes, S. I. McClean. *Statistical techniques for manpower planning*. 1991. John Wiley & Sons.

Table 4. The Minkowski distance for different values of p between main breadwinners and respondents

p	d_{Mk}
$p = 1$	$d_{Mk} = 0.361$
$p = 2$	$d_{Mk} = 0.815$
$p = 3$	$d_{Mk} = 0.19185$
$p = 4$	$d_{Mk} = 0.1064$
$p = 5$	$d_{Mk} = 0.06125$
$p = 6$	$d_{Mk} = 0.03498$
$p = 10$	$d_{Mk} = 0.00374$
$p = 20$	$d_{Mk} = 0.0000$

[5] *Source: SHARELIFE, release 1.0*, $N = 2,938$.

3.D. J. Bartholomew. *Stochastic Models for social processes*. London, 1982. Wiley.
4.A. Börsch-Supan and H. Jürges (eds.). *The Survey of Health, Ageing and Retirement in Europe - Methodology*. Mannheim, 2005. Mannheim Research Institute for the Economics of Aging (MEA).
5.R. Boudon. *Mathematical Structures of Social Mobility*. Amsterdam, 1973. Elsevier.
6.F. A. Cowell. Measures of distributional change: an axiomatic approach. *Review of Economic Studies*, 52:135-151, 1985.
7.G. S. Fields, E. Ok. Measurement movement of incomes. *Economica*, 66:455-471, 1999.
8.M. Frechet. Sur quelques points du calcul fonctionnel (These). *Rend. Circ. Mat. Palermo*, 22:1-74, 1906.
9.Gan, G., Ma, C. and Wu, J. (2007), Data Clustering: Theory, Algorithms, and Applications, ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, PA.
10.F. Hausdorff. *Grundzuege der Mengenlehre*. Leipzig, 1914. Viet.
11.G. Heineck and R. Riphahn. Intergenerational transmission of educational attainment in Germany - the last five decades. *Journal of Economic and Statistics*, 229:36-60, 2007.
12.SHARELIFE, Survey of Health, Ageing and Retirement in Europe, Release Guide 1, Mannheim Research Institute for the Economics of Aging (MEA).
13.A. F. Shorrocks. The measurement of social mobility. *Econometrica*, 46:1013-1024, 1978.
14.M. Symeonaki, O. Filopoulou and G. Stamatopoulou, Measuring Intergenerational Educational Mobility in Greece, ASMDA, Rome, Italy, 7-10 June, 2011.
15.M. Symeonaki, G. Stamatopoulou and K. Michalopoulou, Intergenerational Occupational Mobility in Greece: Evidence from EU-SILC, Demographic Analysis and Research International Conference, Chania, Greece, 5-8 June, 2012.
16.M. Symeonaki and G. Stamatopoulou, Exploring the transition to Higher Education in Greece: issues of intergenerational educational mobility In I. Lamprianou (ed.) "International perspectives of transition to Higher Education: issues of social justice and participation", Policy Futures in Education, 2013 in press.
17.A. F. Shorrocks. Ranking income distributions. *Economica* 50:3-17, 1983.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Asymptotic behavior of kernel density and mode estimates for censored and associated data

Tatachak Abdelkader[1], Ferrani Yacine[2], and Ould Said Elias[3]

[1] Lab. MSTD, Faculty of Mathematics, USTHB, BP 32, 16111, Algeria (E-mail: atatachak@usthb.dz)
[2] Faculty of Sciences, Department of Mathematics, UMBB, Av. de l'indépendance, 35000, Algeria (E-mail: yferrani@umbb.dz)
[3] Univ. Lille Nord de France, F-59000 Lille, France & ULCO, LMPA, F-62228 Calais, France (E-mail: ouldsaid@lmpa.univ-littoral.fr)

Abstract. The aim of this paper is to investigate the asymptotic behavior of a kernel density estimator for randomly right-censored data under association hypothesis. As a result, we state the optimal uniform almost sure convergence rate of the estimator and then, as an application, we derive the almost sure convergence rate of a new kernel mode estimator for the true mode of the underlying density function.
Keywords: Associated data, Censored data, Kaplan-Meier estimator, Kernel mode estimation, Strong uniform consistency.

1 Introduction

Let $\{Y_n, n \geq 1\}$ be a sequence of random variables (r.v.'s) defined on a common probability space $(\Omega, \mathscr{A}, \mathbb{P})$. Recall that the r.v.'s $\{Y_i; 1 \leq i \leq n\}$ are said to be positively associated (PA), if for every pair of functions $\psi_1(\cdot)$ and $\psi_2(\cdot)$ from $\mathbb{R}^n$ to $\mathbb{R}$, which are nondecreasing componentwise, it holds that:

$$\mathrm{Cov}[\psi_1(Y_i, 1 \leq i \leq n), \psi_2(Y_j, 1 \leq j \leq n)] \geq 0,$$

whenever this covariance exists. Infinitely many r.v.'s are said to be PA, if any finite subset of them is a set of PA.
In classical statistical inference, the observed r.v.'s of interest are generally assumed to be independent and identically distributed (i.i.d.). However in some real lifetime systems, it is quite common for the components to be dependent. For example, in reliability and survival analysis, the r.v.'s of lifetimes for components are not independent but associated.
The concept of association was introduced and defined by [1] mainly for the sake of applications. For instance, positive association occurs often in certain reliability theory problems, as well as in some important models employed in statistical mechanics, in certain testing hypotheses problems and in clinical trial studies. For more details and examples (see [6],[9], [4]) and the references therein.
Here, we deal with density estimation for a randomly right censored model given by nonnegative collections of stationary PA r.v.'s $\{Y_i, 1 \leq i \leq n\}$ (survival times) with a common distribution function (d.f.) F and a probability density function (p.d.f.) $f = dF$, and $\{C_i, 1 \leq i \leq n\}$ (censoring times) with

d.f. G. Moreover, we assume that the censoring times are i.i.d. and are independent of the survival times.
In a right censored model, one observes only the n pairs (Z_i, δ_i) with $Z_i = \min(Y_i, C_i)$ and $\delta_i = \mathbf{1}_{\{Y_i \leq C_i\}}$, where $\mathbf{1}_{\{A\}}$ denotes the indicator function of the event A. The independence of the Y_i's from the C_i's ensures the identifiability of the model and then, the distribution H of Z_1 satisfies $\bar{H} := 1 - H = (1-G)(1-F)$. The problem at hand is that of drawing nonparametric inference about f and its mode θ based on the actually observed n pairs (Z_i, δ_i). In this kind of model, it is well known that the classical empirical distribution does not estimate consistently the d.f.'s F and G. Therefore, [7] proposed consistent estimators F_n and G_n for F and G, respectively defined by

$$F_n(t) = 1-\prod_{i=1}^{n}\left[1-\frac{\delta_{(i)}}{n-i+1}\right]^{\mathbf{1}_{\{Z_{(i)}\leq t\}}}, G_n(t) = 1-\prod_{i=1}^{n}\left[1-\frac{1-\delta_{(i)}}{n-i+1}\right]^{\mathbf{1}_{\{Z_{(i)}\leq t\}}}.$$

Here, $Z_{(1)}, Z_{(2)}, \ldots, Z_{(n)}$ denote the order statistics of $Z_1, Z_2, \ldots, Z_n$ and $\delta_{(i)}$ is the concomitant of $Z_{(i)}$. These estimators were studied in depth by many authors in both i.i.d. and strong mixing cases. To quote only a few of them, we refer the reader to [12],[5],[13],[8],[2],[3] and the references therein.
Now, assume that f possesses a unique peak at the mode θ, that is

$$\theta := \arg\sup_{t\in\mathbb{R}} f(t).$$

Based on smoothing techniques, a kernel quasi-estimator for f is given by

$$\tilde{f}_n(t) = \frac{1}{nh_n}\sum_{i=1}^{n}\frac{\delta_i}{1-G(Z_i)}K\left(\frac{t-Z_i}{h_n}\right), \tag{1}$$

where K is a p.d.f. (so-called kernel function) and $h_n =: h > 0$ is a sequence of bandwidths tending to zero along with n. In practice, the estimator in (1) can not be computed since G is unknown, but it plays a great role in establishing our results. Therefore, the feasible estimator for f is to replace (plug-in) $G(\cdot)$ by its Kaplan-Meier estimate $G_n(\cdot)$, and then we get

$$\hat{f}_n(t) = \frac{1}{nh}\sum_{i=1}^{n}\frac{\delta_i}{1-G_n(Z_i)}K\left(\frac{t-Z_i}{h}\right). \tag{2}$$

Hence, a natural kernel estimator of the mode θ is

$$\hat{\theta}_n := \arg\max_{t\in\mathbb{R}} \hat{f}_n(t).$$

Note that in absence of censoring ($\delta_i = 1$), the estimator $G_n(\cdot)$ vanishes everywhere and then (2) becomes the popular kernel density estimator, say $f_n(\cdot)$ (see [10]).
In this study, we establish a strong uniform consistency result with a rate over a compact set for the kernel density estimator of f defined in (2), when the survival times are randomly right-censored by the censoring times and form a stationary positively associated sequence. As an application, we derive the almost sure rate of convergence for the kernel mode estimator $\hat{\theta}_n$ of the true mode θ, defined in (3) below.

2 Assumptions and main results

In the sequel, for any d.f. L, let τ_L be defined by

$$\tau_L = \inf\{t : L(t) = 1\} \leq \infty.$$

Then for the marginal d.f. H of the Z_i's, it holds that $\tau_H = \tau_F \wedge \tau_G$.
And, let $\mathscr{C} := [0, \tau] \subset [0, \tau_H[$ be a compact set such that $\theta \in \mathscr{C}$. We then, without loss of generality, reduce our definition of the mode to the real $\theta := \arg\max_{t\in\mathscr{C}} f(t)$, which implies the necessity to change the above definition of the kernel mode estimator into

$$\hat{\theta}_n := \arg\max_{t\in\mathscr{C}} \hat{f}_n(t). \tag{3}$$

All along this paper, we use c to denote a positive constant which may take a different value for each appearance.
We will make use of the assumptions gathered hereafter for easy reference.

M1. $\{Y_i; i \geq 1\}$ is a stationary sequence of PA r.v.'s with marginal density $f = dF$ and having finite second moment,

M2. The censoring time variables $\{C_i; i \geq 1\}$ are i.i.d. r.v.'s with d.f. G, and are independent of $\{Y_i; i \geq 1\}$,

M3. The covariance term satisfies: $\rho(r) := \sup_{j:|\ell-j|\geq r} Cov(Y_j, Y_\ell)$ for all $\ell \geq 1$ and $r > 0$, where $\rho(r) \leq \gamma_0 e^{-\gamma r}$ for some positive constants γ_0 and γ.

K. K is a Lipschitz p.d.f. with compact support satisfying $\int uK(u)du = 0$.

D1. The p.d.f. $f(\cdot)$ is twice continuously differentiable on $\mathscr{C}$ with second derivative $f^{(2)}(\cdot)$ such that $f^{(2)}(\theta) < 0$,

D2. The joint p.d.f. $f_{1,j}(\cdot,\cdot)$ of (Y_1, Y_{1+j}) satisfies: $\sup_{j>1}\sup_{u,v\in\mathscr{C}} |f_{1,j}(u,v) - f(u)f(v)| \leq c$,

D3. The mode θ satisfies the following property : for any $\varepsilon > 0$ and t, there exists a $\delta > 0$ such that $|\theta - t| \geq \varepsilon \Rightarrow |f(\theta) - f(t)| \geq \delta$.

H. The bandwidth h satisfies: $h \to 0$, $nh \to \infty$ and $\frac{\log^5 n}{nh} \to 0$, as $n \to +\infty$.

Now, we are in a position to list our main results.

Proposition 1 *Suppose that $\tilde{K}_1(\cdot), \tilde{K}_2(\cdot), \ldots, \tilde{K}_n(\cdot)$ are positively associated r.v.'s with mean zero, defined on $(\Omega, \mathscr{A}, \mathbb{P})$ and let $\psi : \mathbb{N}^2 \longrightarrow \mathbb{N}$ be a function defined by $\psi(u, v) = uv$. Furthermore, assume that there exist constants M, L_1, $L_2 < \infty$, μ, $\lambda \geq 0$ and a non-increasing sequence of real coefficients $(\phi(n))_{n\geq 1}$. Then for all u-tuples $(s_1, \ldots, s_u)$ and all v-tuples $(w_1, \ldots, w_v)$ with $1 \leq s_1 \leq \cdots \leq s_u \leq w_1 \leq \cdots \leq w_v \leq n$, we have*

(a). $$Cov\left(\prod_{i=s_1}^{s_u} \tilde{K}_i(t), \prod_{j=w_1}^{w_v} \tilde{K}_j(t)\right) \leq c^{u+v} h\psi(u, v)\phi(w_1 - s_u),$$

(b). $\sum_{s=0}^{\infty}(s+1)^{k_0}\phi(s) \leq L_1 L_2^{k_0}(k_0!)^{\mu},\ \forall k_0 \geq 0,$

(c). $\mathbb{E}\left[\left|\tilde{K}_i\right|^{k_0}\right] \leq (k_0!)^{\lambda} M^{k_0}.$

Theorem 1 *Under Assumptions* **M1**-**M3**, **K**, **D1**, **D2** *and* **H** *we have*

$$\sup_{t\in\mathscr{C}}\left|\tilde{f}_n(t) - \mathbb{E}\left[\tilde{f}_n(t)\right]\right| = O\left(\sqrt{\frac{\log n}{nh}}\right) \text{ a.s., as } n \longrightarrow \infty.$$

Theorem 2 *Under the assumptions of Theorem 1 we have*

$$\sup_{t\in\mathscr{C}}|\hat{f}_n(t) - f(t)| = O\left\{\max\left(\sqrt{\frac{\log n}{nh}}, h^2\right)\right\} \text{ a.s., as } n \longrightarrow \infty.$$

Remark 1 *Note that if we choose* $h = O\left(\left(\frac{\log n}{n}\right)^{1/5}\right)$, *which is the optimal choice with respect to the almost sure uniform convergence criterion in kernel density estimation (see [11]), then we get*

$$\sup_{t\in\mathscr{C}}|\hat{f}_n(t) - f(t)| = O\left(\left(\frac{\log n}{n}\right)^{2/5}\right) \text{ a.s., as } n \longrightarrow \infty,$$

which is optimal in the minimax sense.

Corollary 1 *Under the assumptions of Theorem 1 and* **D3** *we have*

$$|\hat{\theta}_n - \theta| = O\left\{\max\left(\left(\frac{\log n}{nh}\right)^{1/4}, h\right)\right\} \text{ a.s., as } n \longrightarrow \infty.$$

In addition, if we choose $h = O\left(\left(\frac{\log n}{n}\right)^{1/5}\right)$, *then*

$$|\hat{\theta}_n - \theta| = O\left(\left(\frac{\log n}{n}\right)^{1/5}\right) \text{ a.s., as } n \longrightarrow \infty.$$

References

1.J. Esary, F. Proschan and D. Walkup. Association of random variables with applications, *Annals of Mathematical Statistics*, **38**, 1466–1476, 1967.

2.Z. Cai. Asymptotic properties of Kaplan-Meier estimator for censored dependent data, *Statistics and Probability Lettters*, **37**, 381–389, 1998.

3.Z. Cai and G. G. Roussas. Kaplan-Meier estimator under association, *Journal of Multivariate Analysis*, **67**, 318–348, 1998.

4.J. Dedecker, P. Doukhan, G. Lang, J. R. Léon, S. Louhichi and C. Prieur. *Weak Dependence: With Examples and Applications,* Lecture Notes in Statistics, 2007.

5.P. Deheuvels and J. Einmahl. (2000). Functional limit laws for the increments of Kaplan-Meier product limit processes and applications. *Annals of Probability*, **28**, 1301–1335, 2000.
6.P. Doukhan and S. Louhichi. A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Application*, **84**, 313–342, 1999.
7.E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations, *Journal of the Americal Statistical Association*, **53**, 457–481, 1958.
8.J. Lecoutre and E. Ould Saïd. Kaplan-Meier estimate under strong mixing. Journal of Statistical Planning and Inference, **44**, 359–369, 1995.
9.S. Louhichi. Weak convergence for empirical processes of associated sequences, *Les Annals de l'Institut Henri Poincarré: Probabilités et Statistiques*, **36**, 547–567, 2000.
10.E. Parzen. On estimation of a probability density function and mode, *The Annals of Mathematical Statistics*, **33**, 1065–1076, 1962.
11.W. Stute. A law of the logarithm for kernel density estimators, *Annals of Probability*, **10**, 414–422, 1982.
12.W. Stute and J. L. Wang. The strong law under random censorship, *Annals of Statistics*, **21**, 1591–1607, 1993.
13.Z. Ying and L. J. Wei. The Kaplan-Meier Estimate for Dependent Failure Time Observations. *Journal of Multivariate Analysis*, **50**, 50,17–29, 1994.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Minimum Pseudodistance Estimators and Applications to Portfolio Optimization

Aida Toma[1,2] and Samuela Leoni-Aubin[3]

[1] Department of Applied Mathematics, Bucharest Academy of Economic Studies, Piaţa Romană 6, Bucharest, Romania (E-mail: aida_toma@yahoo.com)
[2] "Gh. Mihoc - C. Iacob" Institute of Mathematical Statistics and Applied Mathematics, Calea 13 Septembrie 13, Bucharest, Romania
[3] INSA Lyon, ICJ, 20, Rue Albert Einstein, 69621 Villeurbanne Cedex, France (E-mail: samuela.leoni@insa-lyon.fr)

Abstract. We present a class of portfolios that have better stability properties than the traditional mean-variance portfolios. These portfolios are constructed using robust minimum pseudodistance estimators as inputs in the optimization procedure. Our numerical results on empirical financial data confirm the advantages of the new approach.
Keywords: Portfolio optimization, Robustness, Pseudodistances.

1 Introduction

In recent years, researchers and practitioners accorded much attention to portfolio optimization and risk management, these being important components in the activity of the investment societies, banks and insurance companies.

The theory of Markowitz[12] introduced the idea of investment diversification, showing that the investor should take into account for his choice, not only the characteristics of each financial asset, but also the correlation of each asset with the other assets. The practitioners often use the Markowitz's model for portfolio optimization, although in recent literature has been proved the fact that, in some conditions, the traditional approach may lead to irrelevant sub-optimal portfolios. Among the reasons of this drawback, an important role is played by the unbounded, arbitrary influence of the extreme returns on the procedure. The mean-variance optimization determines the optimal portfolio which maximizes an objective function whose input parameters are the mean and the covariance matrix of the asset returns. In the classical approach, these parameters are estimated using the maximum likelihood technique. In the presence of atypical observations in the sample (these observations can be arbitrarily large returns or arbitrarily small returns), the biases of these estimations can be arbitrarily large and may induce significant changes in the composition of the portfolio. In order to construct efficient portfolios which will be stable with respect with the presence of atypical observation in the sample, some authors suggested to use robust estimators for the vector of mean returns and for the covariance matrix of the returns (see for example DeMiguel and Nogales[6], Fabozzi *et al.*[7], Perret-Gentil and Victoria-Feser[13]). Although the contributions of these authors have the merit to consider the role of robust estimation for improving the mean-variance portfolios, it is a well known fact

that the traditional robust estimators suffer important looses in efficiency in comparison with the maximum likelihood estimators.

In this context, we suggest replacing the classical maximum likelihood estimators with minimum pseudodistance estimators. These estimators are robust with respect to local deviations from normality, assuring the fact that the resulted mean-variance optimal portfolios reflect the statistical properties for the majority of the data and their characteristics are not affected by the possible abnormal strong influence of some few observations with extreme values. As the divergences, the pseudodistances are generalizations of distances between probability measures, usually not satisfying the triangle inequality. The use of divergence measures in statistical inference is an important issue. We refer to the book Basu[2] for a variety of statistical methods based on divergences, as well as to some recent papers such as Broniatowski and Keziou[4], Bouzebda and Keziou[3], Mattheou and Karagrigoriou[11], Toma and Broniatowski[15], Toma and Leoni-Aubin[16] and Vonta *et al.*[18] which develop divergence based estimation and test procedures. The class of pseudodistances that we consider is indexed by a tuning parameter $\alpha \geq 0$. The choice of this constant determines the robustness and efficiency properties of the corresponding minimum pseudodistance estimator. The pseudodistance indexed by $\alpha = 0$ coincides with the modified Kullback-Leibler divergence and the estimator which minimizes this pseudodistance is the classical maximum likelihood estimator itself. It is known the fact that, when the model is correctly specified, the maximum likelihood estimator has the highest asymptotic efficiency, but in the case of misspecification, this estimator gives completely erroneous results. Instead, the minimum pseudodistance estimators associated with strictly positive α gain robustness with the price of a relatively small loss of efficiency.

Our empirical investigations on financial data show that the investment strategies based on using minimum pseudodistance estimators are less exposed to the extreme risks. The choice of the parameter α, associated to the estimation method, allows for a flexible treatment for different periods of time: when the volatility is low, the data can be approximated by the normal model, so the choice of α close to zero leads to efficient estimations. When the volatility is high, the choice of α between 0.2 and 0.5 significantly reduces the bias of estimations and leads to relevant conclusions.

Thus, the proposed robust methods significantly reduce the effect of atypical observations from samples, have high efficiency and therefore they can represent viable alternatives to the existing methods.

2 Minimum pseudodistance estimators

Broniatowski *et al.*[5] introduced a class of parametric estimators called minimum pseudodistance estimators. The family of pseudodistances is indexed by a positive tuning parameter α and is defined as follows. For P and Q two probability measures with densities p, respectively q, with respect to the Lebesgue measure, the pseudodistance of order α is defined by

$$R_\alpha(P,Q) := \frac{1}{\alpha+1}\ln\int p^\alpha \mathrm{d}P + \frac{1}{\alpha(\alpha+1)}\ln\int q^\alpha \mathrm{d}Q - \frac{1}{\alpha}\ln\int p^\alpha \mathrm{d}Q \qquad (1)$$

for $\alpha > 0$. The following limit relation holds

$$R_\alpha(P,Q) \to R_0(P,Q) := \int \ln \frac{q}{p} \mathrm{d}Q \quad \text{for} \quad \alpha \downarrow 0.$$

Basu *et al.*[1] and Jones *et al.*[9] considered similar classes of pseudodistances in order to define robust and efficient estimates. We also note that the pseudodistances (1) are related to the Renyi entropy functionals as studies by Källberg *et al.*[10].

Let $(P_\theta)_{\theta\in\Theta}$ be a parametric model, with the parameter space $\Theta \subset \mathbb{R}^d$, and assume that each probability measure P_θ has a density p_θ with respect to the Lebesgue measure. Let $X^1, \ldots, X^T$ be a sample on P_θ. A minimum pseudodistance estimator $\widehat{\theta}_n$ of the parameter θ is defined by

$$\widehat{\theta}_n := \arg\inf_\theta R_\alpha(P_\theta, P_n),$$

where P_n is the empirical measure pertaining to the sample, or equivalently as

$$\widehat{\theta}_n = \begin{cases} \arg\sup_\theta \frac{1}{TC_\alpha(\theta)} \sum_{i=1}^T p_\theta^\alpha(X^i) & \text{if } \alpha > 0 \\ \arg\sup_\theta \frac{1}{T} \sum_{i=1}^T \ln p_\theta(X^i) & \text{if } \alpha = 0 \end{cases}$$

where

$$C_\alpha(\theta) = \left(\int p_\theta^{\alpha+1} d\lambda \right)^{\frac{\alpha}{\alpha+1}}.$$

Note that, the particular case $\alpha = 0$ corresponds to the definition of the maximum likelihood estimator of the parameter θ.

When p_θ is the N-variate normal density with $\theta = (\mu, \Sigma)$

$$p_\theta(x) = \left(\frac{1}{2\pi}\right)^{N/2} \sqrt{\det \Sigma^{-1}} \exp\left(-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right)$$

the minimum pseudodistance estimators of $\theta = (\mu, \Sigma)$, corresponding to positive α, may be written as

$$\widehat{\theta}_n = \arg\sup_\theta (\sqrt{\det \Sigma^{-1}})^{\frac{\alpha}{\alpha+1}} \sum_{i=1}^T \exp\left(-\frac{\alpha}{2}\|X^i - \mu\|^2_{\Sigma^{-1}}\right),$$

where $\|X^i - \mu\|^2_{\Sigma^{-1}} = (x-\mu)^t \Sigma^{-1}(x-\mu)$. By differentiation with respect to μ and Σ, the minimum pseudodistance estimators of these parameters are solutions of the system

$$\mu = \sum_{i=1}^T \frac{\exp(-\frac{\alpha}{2}\|X^i - \mu\|^2_{\Sigma^{-1}})}{\sum_{i=1}^T \exp(-\frac{\alpha}{2}\|X^i - \mu\|^2_{\Sigma^{-1}})} X^i$$

$$\Sigma = \sum_{i=1}^T \frac{(\alpha+1)\exp\left(-\frac{\alpha}{2}\|X^i - \mu\|^2_{\Sigma^{-1}}\right)}{\sum_{i=1}^T \exp(-\frac{\alpha}{2}\|X^i - \mu\|^2_{\Sigma^{-1}})} (X^i - \mu)(X^i - \mu)^t.$$

The definitions and properties of the minimum pseudodistance estimators in the unidimensional normal case, as well as some empirical results in this case are given in Toma[14].

For studying the robustness of an estimator the influence function is often used. Recall that, a map T defined on a set of probability measures and parameter space valued is a statistical functional corresponding to an estimator $\widehat{\theta}_n$ of the parameter θ, if $\widehat{\theta}_n = T(P_n)$. As it is known, the influence function of T at P_θ is defined by

$$\mathrm{IF}(x; T, P_\theta) := \left.\frac{\partial T(\widetilde{P}_{\varepsilon x})}{\partial \varepsilon}\right|_{\varepsilon=0}$$

where $\widetilde{P}_{\varepsilon x} := (1-\varepsilon)P_\theta + \varepsilon\delta_x$, $\varepsilon > 0$, δ_x being the Dirac measure putting all mass at x (see Hampel *et al.*[8]). Whenever the influence function is bounded with respect to x, the corresponding estimator is called robust.

When P_θ is the N-variate normal model and $\theta = (\mu, \Sigma)$ is the parameter to be estimated, the influence functions of the minimum pseudodistance estimators of $\theta = (\mu, \Sigma)$ are

$$\mathrm{IF}(x; \mu, P_{\mu,\Sigma}) = (\sqrt{\alpha+1})^{N+2}(x-\mu)\exp\left(-\frac{\alpha}{2}\|x-\mu\|^2_{\Sigma^{-1}}\right) \tag{2}$$

$$\mathrm{IF}(x; \Sigma, P_{\mu,\Sigma}) = (\sqrt{\alpha+1})^{N+4}\left[(x-\mu)(x-\mu)^t - \frac{1}{\alpha+1}\Sigma\right]\exp\left(-\frac{\alpha}{2}\|x-\mu\|^2_{\Sigma^{-1}}\right) \tag{3}$$

In the particular case $\alpha = 0$, the influence functions of the maximum likelihood estimators of location and covariance are obtained. As it is well known, these influence functions are unbounded and therefore the maximum likelihood estimators are not robust. When $\alpha > 0$, the estimators gain robustness, in this case the influence functions (2) and (3) being bounded with respect to x.

We mention that the pseudodistances (1) have also been considered by Toma and Leoni-Aubin[17] for defining a new class of optimal robust M-estimators using the Hampel's infinitesimal approach.

3 Portfolio optimization

To introduce the general problem, let us suppose to have a portfolio of N financial assets. Let $X := (X_1, \ldots, X_N)^t$ be the random vector of the asset returns, X_i representing the random variable associated to the return of the i^{th} asset, $i = 1, \ldots, N$. Let $\mu = (\mu_1, \ldots, \mu_N)^t := \mathrm{E}(X)$ be the vector of the expected returns and $\Sigma := \mathrm{cov}(X)$ be the covariance matrix of the returns of the assets. Denote by p_i the weight of the i^{th} asset in the portfolio and let $p := (p_1, \ldots, p_N)^t$. Then, the total return of the portfolio is given by the random variable $\sum_{i=1}^N p_i X_i$. The expected return of the portfolio is

$$R(p) := \mathrm{E}(\sum_{i=1}^N p_i X_i) = p^t\mu$$

and the variance of the portfolio return is

$$S(p) := \mathrm{Var}(\sum_{i=1}^{N} p_i X_i) = p^t \Sigma p.$$

The Markowitz's approach for optimal portfolio selection consists of determining the portfolio p^* solution to the following optimization problem. For a given positive value of the parameter λ, representing the investor's risk aversion, the portfolio p^* is solution of

$$\arg\max_p \{R(p) - \frac{\lambda}{2} S(p)\}$$

with the constraint $p^t e_N = 1$, where $e_N := (1, \ldots, 1)$ is the N-dimensional vector of ones. The solution to this optimization problem is given by

$$p^* = \frac{1}{\lambda} \Sigma^{-1} (\mu - \eta e_N)$$

where

$$\eta = \frac{e_N^t \sum^{-1} \mu - \lambda}{e_N^t \sum^{-1} e_N}.$$

Different positive values of λ correspond to different investment strategies and determine the so-called mean-variance efficient frontier. If we gradually increase λ from zero and for each instance solve the optimization problem, we obtain each portfolio along the efficient frontier. This is the case when short selling is allowed. When short selling is not allowed, all the portfolio weights p_i, $i = 1, \ldots, N$, have positive values.

Traditionally, the unknown parameters μ and Σ are estimated using their sample counterparts, namely the maximum likelihood estimators under multivariate normal distribution. It is known that the portfolio optimization based on sample mean and covariance performs poorly in practice. Since the maximum likelihood estimators of μ and Σ, which are inputs in the optimization procedure, are very sensitive to outlying observations, the weights of the resulted portfolio, which are outputs of this procedure, may be substantially affected by such observations.

Our approach for robust portfolio optimization is based on using robust minimum pseudodistance estimators of location and covariance. The robustness of the mean vector and covariance matrix minimum pseudodistance estimators is transferred to the weights of the resulted portfolio, since the influence functions of weights estimators are proportional with the influence functions of the mean and covariance matrix estimators.

4 Empirical results

We consider 172 monthly log-returns of 8 MSCI Indexes (France, Germany, Italy, Japan, Pacific Ex JP, Spain, United Kingdom and USA) from January

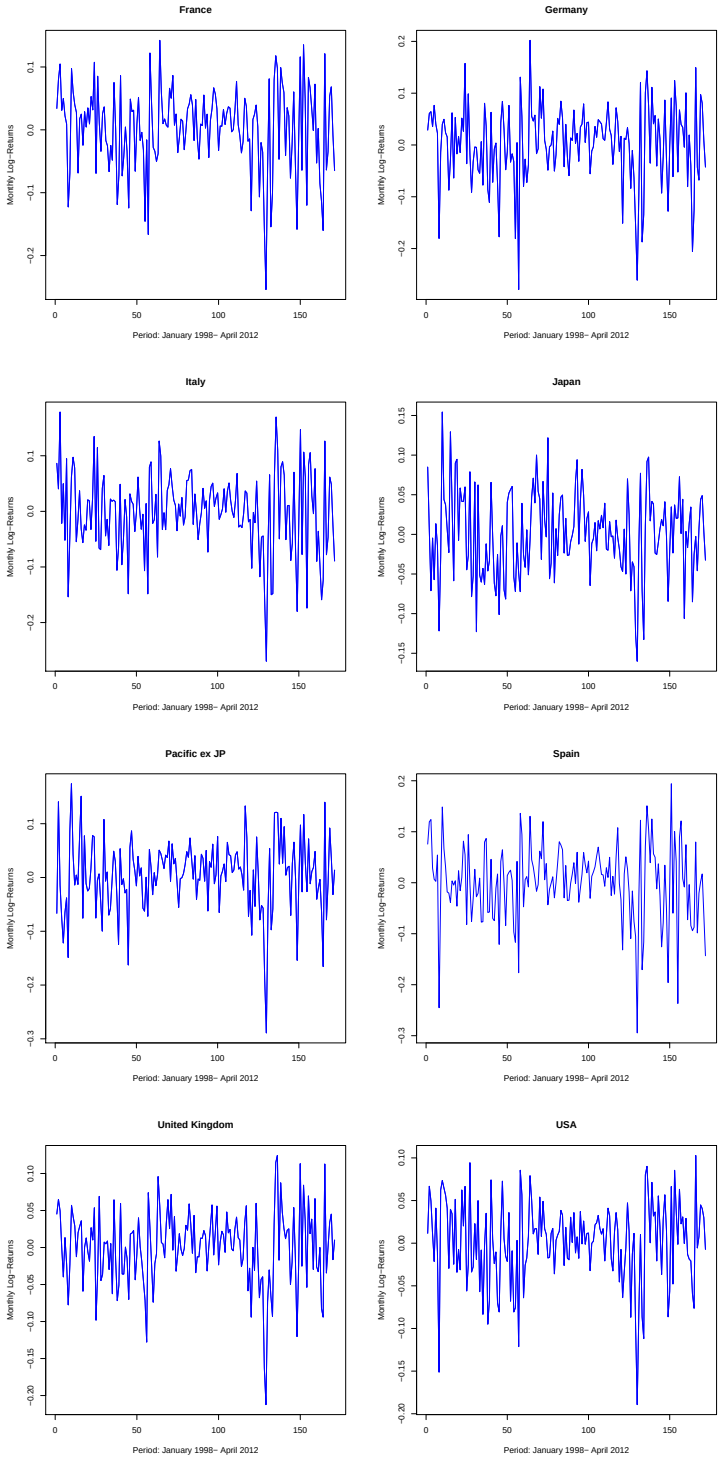


Fig. 1. Monthly log-returns of the indexes

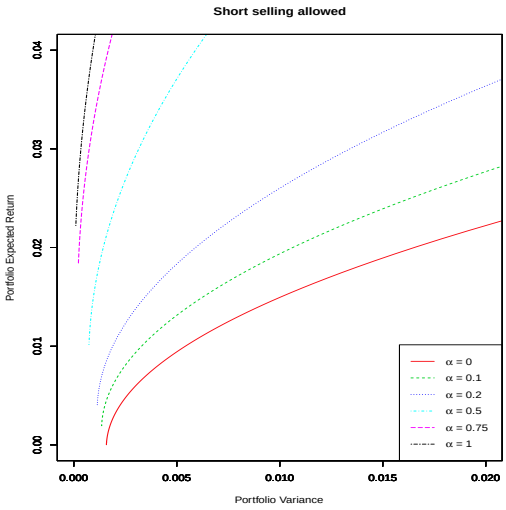


Fig. 2. Mean-variance efficient frontiers

1998 to April 2012 with the aim to construct robust and efficient portfolios. The data are provided by MSCI (www.msci.com) and are presented in Figure 1. We notice that some log-returns are extreme with very low or high values. Such extreme log-returns heavily bias the sample mean and covariance estimators and this bias can be seen when comparing these estimators with the robust minimum pseudodistance estimators.

Minimum pseudodistance estimates of the mean vector and of the covariance matrix computed for different values of the tuning parameter α are used to determine efficient frontiers. In Figure 2 we plot efficient frontiers for the case "short selling allowed". The frontiers based on the minimum pseudodistance estimations dominate those based on the classical maximum likelihood estimations, yielding portfolios with larger expected returns and smaller risks. Given a value of the portfolio variance, the higher is the value of α, the higher the value of the portfolio return is. In order to have a trade-off between robustness and efficiency of the estimation procedures, values of α between 0.2 and 0.5 would be preferred, since for larger values of α, the efficiency of the estimates is too low.

Thus, the proposed robust estimates reduce the volatility effects which typically affects the results of the traditional approaches and therefore they represent an alternative for robust portfolio optimization.

Acknowledgments

This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, project number PN-II-RU-TE-2012-3-0007.

References

1.A. Basu, I.R. Harris, N.L. Hjort and M.C. Jones. Robust and efficient estimation by minimizing a density power divergence. *Biometrika*, 85, 549–559, 1998.
2.A. Basu, H. Shioya and C. Park. *Statistical inference: the minimum distance approach*, Chapman & Hall, 2011.
3.S. Bouzebda and A. Keziou. New estimates and tests of independence in semiparametric copula models. *Kybernetica*, 46, 178–201, 2010.
4.M. Broniatowski and A. Keziou. Parametric estimation and tests through divergences and the duality technique. *Journal of Multivariate Analysis*, 100, 16–36, 2009.
5.M. Broniatowski, A. Toma and I. Vajda. Decomposable pseudodistances and applications in statistical estimation. *Journal of Statistical Planning and Inference*, 142, 2574–2585, 2012.
6.V. DeMiguel and F.J. Nogales. Portfolio selection with robust estimation. *Operations Research*, 57, 560–577, 2009.
7.F.J. Fabozzi, D. Huang, and G. Zhou. Robust portfolios: contributions from operations research and finance. *Ann. Oper. Res.*, 176, 191–220, 2010.
8.F.R. Hampel, E. Ronchetti, P.J. Rousseeuw and W. Stahel. *Robust statistics: the approach based on influence functions*, Wiley, New York, 1986.
9.M.C. Jones, N.L. Hjort, I.R. Harris and A. Basu. A comparison of related density based minimum divergence estimators. *Biometrika*, 88, 865–873, 2001.
10.D. Källberg, N. Leonenko, O. Seleznjev. Statistical inference for Renyi entropy functionals. *Conceptual Modelling and its Theoretical Foundations*, 1, 36–51, 2012.
11.K. Mattheou and A. Karagrigoriou. A new family of divergence measures for tests of fit. *Australian and New Zeeland Journal of Statistics*, 52, 187–200.
12.H.M. Markowitz. Mean-variance analysis in portfolio choice and capital markets. *Journal of Finance*, 7, 77–91, 1952.
13.C. Perret-Gentil and M.P. Victoria-Feser. Robust mean-variance portfolio selection. FAME Research Paper, no. 140, 2005.
14.A. Toma. Robust estimations for financial returns: an approach based on pseudodistance minimization. *Economic Computation and Economic Cybernetics Studies and Research*, 46, 117–131, 2012.
15.A. Toma and M. Broniatowski. Dual divergence estimators and tests: Robustness results. *Journal of Multivariate Analysis*, 102, 20–36, 2011.
16.A. Toma and S. Leoni-Aubin. Robust tests based on dual divergence estimators and saddlepoint approximations. *Journal of Multivariate Analysis*, 101, 1143–1155, 2010.
17.A. Toma and S. Leoni-Aubin. Optimal robust M-estimators using Renyi pseudodistances. *Journal of Multivariate Analysis*, 115, 359–373, 2013.
18.F. Vonta, K. Mattheou and A. Karagrigoriou. On properties of the (phi,a)-power divergence family with applications in goodness of fit tests. *Methodology and Computing in Applied Probability*, 14, 335–356, 2012.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

A Dairy Market Information and Intelligence System

Arturo Valdivia[1], José A. Espinosa[2], Arturo G. Valdivia[3], Georgel Moctezuma[2] , Alejandra Vélez[2], José L. Jolalpa[2], Arturo González[2], América A. Luna[2], José J. Espinoza[2], Sergio F. Góngora[2], Carlos Arriaga[2], Ernesto Martínez[4] and José Dávalos[5]

[1] University de Barcelona, 585 Gran Via de les Corts Catalanas, 08007 Barcelona, Spain. (E-mail: arturo.valdivia@ub.edu)
[2] National Institute for Agricultural, Livestock and Forestry Research of Mexico, 5 Avenida Progreso, 04010 Mexico City, Mexico. (E-mail: espinosa.jose@inifap.gob.mx)
[3] Autonomous University of Aguascalientes, Aguascalientes, Mexico (E-mail: avaldiv@correo.uaa.mx)
[4] Autonomous University of Mexico State, 100 Instituto Literario, 50000 Toluca, Mexico. (E-mail: fernestom@yahoo.com.mx)
[5] National Autonomous University of Mexico, 3000 Avenida Insurgentes Sur, 04510 Mexico City, Mexico (E-mail: jldf@unam.mx)

Abstract. In this paper we present the implementation and initial results of the project *InfoLeche-México*, an information and intelligence system for the bovine milk market in Mexico. The variables of interest were selected following the a Delphi methodology, including farmers and experts from several Mexican research institutions and educative institutions (national an state universities). As a result of this panel of experts, nearly 1450 variables were defined. The up-to-date value of these variables is uploaded to a content management system which acts both as data warehouse and knowledge community environment. Once the data is the system, a series of target variables are computed and interactively presented to the user. All the processed data is public but previously anonymized. In addition, every month the research team of the project elaborates a newsletter summing up a series recommendations.

Keywords: Computing-aided decision support, Data warehouse, Delphi methodology, Stochastic modelling, Risk indicators, Time series analysis, Knowledge systems, open-source content management system

1 Introduction

In the past decades the subject of market intelligence has become more relevant, due the increasing globalization of the economy. Researchers have emphasised the value of knowledge for increasing competitive advantage (Amit and Schoemaker [1], Asoh Derek et al. [2], Grant [6]), and empirical evidence from consolidated companies (*e.g.*, Virgin or Dell Computers) shows that developing a market orientation is likely to enhance business performance (Pulendran et al. [13] and Cravens et al. [4]). Consequently, strategies based on products are thus turning into business strategies based on knowledge (Nonaka [11]).

In Mexico the literature on this topic, encompassing empirical studies from the agricultural sector, is thin. On the other hand, the *Sistema Nacional de Información e Integración de Mercados*, a national service in charge of offering information about the price behavior of the main agricultural, livestock and fishing commodities does not fully cover the dairy market commodities, and does not uses the information that comes directly from production units. In this scenario, the project *InfoLeche-México* addresses the problem of developing a specialized dairy market information and intelligence system for the bovine milk market in Mexico. This project aims to set a knowledge community that puts together both researchers and the main actors of an economical activity. We believe that this interaction between scholars and practitioners will create an environment for listening, understanding and responding to the market and the competition. The main interaction between the members of this knowledge community is meant to take place in a web application designed by *InfoLeche-México*. This web application acts also as data warehouse since the up-to-date value of a series of variables of interest is continuously uploaded and preprocessed. All the processed data is public but previously anonymized.

In this paper we present the methodology, implementation and initial results of the project *InfoLeche-México*. The paper is organized as follows. In Section 2 we schematize the workflow of the project, and we present the role of each type of participant. In Section 3 we describe the variables of interest for our study, along with the the methodology followed in order obtain them. In Section 4 we discuss the implementation of the web application that will serves as host to our knowledge community. Finally in Section 5 we present the current state of project and some of the initial results and some topics of current interest.

2 Information flow

The general scheme for the information flow of the market information and intelligence system *InfoLeche-México* is depicted in Figure 1. The first part of the information flow consists in the continuous update of the current value a series of variables of interest –see Section 3. This data recollection is carried out by a group of *collaborators* which regularly visit the participant production units in order to fill in a predefined spread sheet file. This follow-up file is uploaded to the web application –see Section 4. A team of *validators* review the data uploaded by the collaborators, in order to validate it and subsequently be added to database. The moment the new data is uploaded, the web application automatically computes a series of target variables that helps the validators with their task. Once the validation process is complete, a module of the web application takes of updating the database with the new raw and preprocessed data.

For the final part of the information process, the raw and preprocessed data is analyzed by the research team in order derive a series of indicators and recommendations which are afterwards presented in the web platform as

newsletters and other digital materials. These materials as well as the anonymized raw and preprocessed data is accessible to all the guest visitors of the web page. In addition, the platform contains a collection of media and research articles related to the problems and topics addressed by *InfoLeche-México*. We remark here the challenge of addressing an heterogeneous knowledge community encompassing scholars and practitioners.

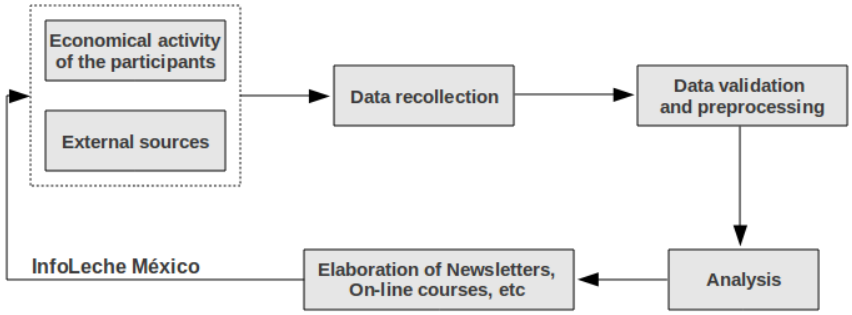


Fig 1. The image depicts the information flow in the project.

Even though the initial data analysis strongly relies in time series techniques, our intention is to use the database obtained in order to study different continuous-time models for the variables of interest. These models may improve our understanding of the some of the variables, and may be used to forecast specific risk indicators. This study corresponds to a second stage of the project.

3 Variables of interest

According to Figure 1, the information flow starts by selecting the market variables that will be relevant to our study, *i.e.*, variables with influence in the competitiveness and rentability of the production systems, as well as variables describing the interaction between farmers and the rest of the production chain. To this matter, the three-step process (see Figure 2) has been followed. The first step in the process consists of an intensive literature review covering economical, technological, social, environmental, competitive, productive and marketing aspects. This review lead to the identification of a preliminary list of 54 of variables. The preliminary list of variables were concentrated in the form of a table as shown in Figure 3. These variables are divided in three big categories:

(a) Input and output of commodities, services, infrastructure, equipment and installations used in farms. In this category 26 variables were defined, including, for instance, *corn silage*, *concentrated feeds*, *family labor*, *artificial insemination*, and *milking machines*.

(b) National context variables. In this category 16 variables were defined, including, for instance, *milk production volume by Mexican states*, *average milk price*, *grains price*, *national milk consumption*, *government subsidies*, and *interest rates*.

(c) International context variables. In this category 11 variables were defined, including, for instance, *importation of replacement heifers*, *infectious diseases of international concern*, *productions costs and government subsidies in other countries*.

An impact matrix analysis and cross impact analysis (see Bas [3]) was used in order to establish, respectively, a priorization and relationship among the variables.

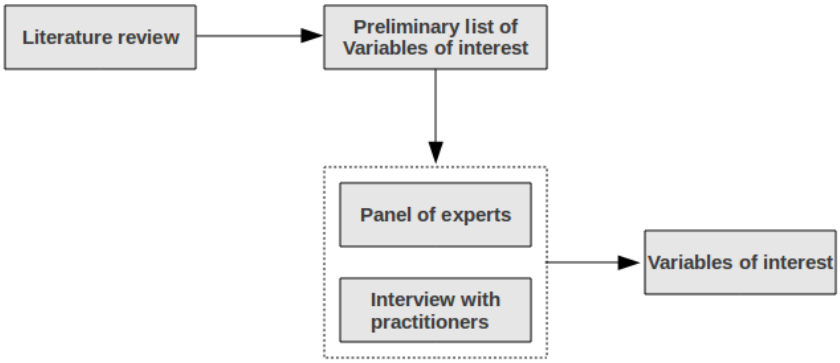


Fig 2. The image depicts the process followed in order to define the variables of interest for the project.

As a second step of the process, and following the methodology of Rey et. al. [14], a Delphi panel of experts was proposed in order to validate the preliminary list of variables. The panel was formed with nearly 60 experts, 90% of them coming from academic and research institutions, and the remaining 10% coming from government departments and farmers organizations. Each member was invited to the panel by taking into consideration its experience in the field (more than 20 years), reputation among pairs, and manifest willingness to participate. The panel members discussed the present and future (up to the year 2030) importance of the variables, taking into account both under two scenarios: a trend and optimistic scenarios. In addition, a series of interviews with practitioners were performed in order to warranty a clear conceptualization of the identified variables.

As a final step of the process, following the methodology of Gómez de Castro et al. [7], the research team of the project have elaborated an initial list of 48 variables. The statistical measures used were the first and third quartile, the interquartile difference an the median. The value scale was 1 to 10, wherein the 1 meant almost no importance and 10 was extremely important. From these 48 variables, a complete list consisting of around 1450 variables of interest were defined in order to account for important interactions between the original 48. The identification of the important interactions was obtained, similarly, during the literature review, the Delphi panel and the interview with practitioners. The detailed process can be found in Espinosa-García
et al. [5].

Variables	Definición de la variable	Referencias bibliográficas
Montos de financiamiento (créditos) para la producción de leche.	Cantidad monetaria otorgada por el sector público (FIRA y Financiera Rural) y el sector privado (Banca Comercial) para financiar mediante créditos (avío y refaccionario) la actividad lechera en México.	González *et al.*, 2011. Ricossa, 2004. Cervantes *et al.*, 2001.
Apoyo (subsidios) para fomentar la producción de leche nacional y por estado.	Cantidad monetaria otorgada por el sector público (gobiernos federal y estatal) para fomentar la producción de leche.	Cervantes *et al.*, 2001. Téllez, 1994.
Consumo nacional de leche.	Cantidad de leche (fluida y en polvo) que consume el país: producción interna de leche más importaciones menos exportaciones.	García, 1992. Nacional Financiera, 1995. Dornbusch y Fischer, 1992. Ricossa, 2004.

Fig 3. The image shows a portion of the preliminary list variables of interest identified after the literature review. In the first column a short name for the variable is established, whereas the second column contains its description. The third column lists the literature references that lead to identify the variable.

4 The web application

InfoLeche-México is built on an open source application licensed under the Gnu Public License (see http://www.gnu.org/). The main reason behind this choice is that we aim for (1) a robust and secure application that does not require to be programmed from scratch, (2) whose features are easy to extend and improve, independently from initial development team. It is apparent that these characteristics leads to a fast and efficient development of a functional and longevous web application (cf. James and Noble [8], and Miles and Miles [10]).

After considering various possibilities, we have chosen to use the framework of *Drupal*, which is an open source content management platform powering millions of websites and applications. It's built, used, and supported by an active and diverse community of people around the world (see http://drupal.org/). Many prominent organizations around the world –from entertainment companies to government departments– have built their web sites on *Drupal*; some of them may be found listed in http://www.drupalsites.net. For more information about *Drupal,* we refer the interested reader to its official page.

According to the workflow depicted in Figure 1, the roles guest, collaborator, validator, researcher and invited researcher were we defined. Each of them find different and specific tools in the platform once they log in. While the tools for the first three roles relay on preexisting modules of *Drupal*; the last two roles required the development of a particular module that takes care of the concentration of the data from all the participants into a single *comma separated value* file that can be downloaded in order to carry out a further analysis in a specialized program as R (see the official web page of the R project, http://www.r-project.org/). Once this new module is finished, we shall submit this new module to the Drupal contributed modules repertory. In this sense *InfoLeche-México* is coherent with the open source philosophy.

5 Current state of the project and first results

Currently, the prototype for the web application is finished. The collaborators have already started the continuous uploading of the data coming from all the participating production units, thus enriching the database. The participating researchers are already using the prototype in order to validate and analyze the data uploaded by the collaborators. As a result, the first newsletter presented by *InfoLeche-México* elaborated in October of last year. Since then, newsletters have been prepared to appear on monthly bases. Some of the topics discussed so far in the newsletter are: the increment in milk powder imports; the behavior of grain prices in Mexico; and a comparative analysis between the states of Estados de México and Guanajuato, taking the variables described in Section [2] as reference for the comparison.

As mentioned in the introduction, *InfoLeche-México* is meant to serve both as a data warehouse and knowledge community. Consequently, apart from the continuous evaluation of the web application and the products described in Section [3], one of the ongoing tasks consists in enlarging the database of media and scientific articles related to the problems addressed by *InfoLeche-México*.

Conclusions

The project *InfoLeche-México* serves as an example of implementation of a knowledge community that puts together both researchers and the main actors of

an economic activity. This community gathers and analyzes up-to-date data coming from the production process and the market. As a result of the analysis, a series recommendations are presented to the participating businesses and the general public in order to aid their decision making. It is worth noticing, on the one hand, that this project is the first public service in Mexico of this type, and that it benefits particularly to smallest links (*e.g.*, family farms) of the production chain of milk in Mexico. On the second hand, let us remark that this community also facilitates an environment for outlining future research projects, thus ensuring the longevity of the project.

We believe that the discussion and implementation of the project *InfoLeche-México*, presented in this paper, may provide a concrete example for similar projects.

InfoLeche-México is a part of the macro-project "Improving productivity, competitiveness and sustainability of the production chain of bovine milk in Mexico"; an overall description we refer to Núñez-Hernández et al. [12]. The macro-project is directed by the National Institute for Agricultural, Livestock and Forestry Research of Mexico (INIFAP); Secretariat of Agriculture, Livestock, Rural Development, Fisheries and Food (SAGARPA); and the Confederation of Produce Foundations (COFUPRO). And it is supported by the National Council of Science and Technology (financial support No. SAGARPA –CONACyT 144591).

References

1. R. Amit and P.J.H. Shoemaker. Strategic assets and organizational rents. *Strategic Management Journal*, 14, 1, 33–46, 1993.

2. A. Asoh Derek, B. Salvatore and D. Peter. Alignment: the missing link in knowledge management research. In *Fourth European Conference on Knowledge Management*, Fergal McGrath University of Limerick, 39–47, 2003.

3. E. Bas. *Prospectiva, herramientas para la gestión estratégica del cambio*. Ariel, España, 1999.

4. D. Cravens, G. Greenley, N. F. Piercy and S.F. Slater. Mapping the path to market leadership. *Marketing Management*, 7, 3, 1998.

5. J. A. Espinosa-García, G. Moctezuma-López, A. G. Valdivia-Flores, J. L. Jolalpa-Barrera, S. F. Góngora-González, E. Martínez, A. Vélez-Izquierdo, A. T. González-Orozco, J. J. Espinoza-Arellano, A. A. Luna-Estrada, J. L. Dávalos-Flores, C. Arriaga-Jordán. *Variables del mercado de insumos y productos que inciden en la competitividad de la cadena de leche de bovino en México*. In Avances de Investigación para la producción de de la Productividad, Competitividad y Sustentabilidad de la Cadena Productiva de

Leche de Bovino en Mexico" Etapa 1. Libro Científico Núm. 4., INIFAP, Centro de Investigación Regional Norte Centro. Coahuila, 2012.

6. R.M. Grant. The resource-based theory of competitive advantage: implications for strategy formulation. *California Management Review*, 33, 3, 114–135, 1991.

7. A. Gómez de Castro, S. Valle, A. Maestrey, V. Trujillo, O. Alfaro, O. Mengo and M. Medina. *La dimensión de "futuro" en la construcción de la sostenibilidad institucional*. International Service for National Agricultural Research project "Nuevo Paradigma", San José, Costa Rica, 2001.

8. T. James and M. Noble. *Drupal 7 Business Solutions*. Packt Publishing, Birmingham, 2012.

9. M.A. Justo and A.P. Parra. *Inteligencia de Mercado de Productos Diferenciados Comercialización de hortalizas en fresco*. Working document No. 30 (available at www.inta.gov.ar/ies) 2004.

10. E. Miles and L. Miles. *Drupal's Building Blocks: Quickly Building Web Sites with CCK, Views and Panels*. Peason Education, Indiana, 2001.

11. I. Nonaka. A dynamic theory of organizational knowledge creation. *Organization Science*, 5, 1, 14–37, 1994.

12. G. Núñez-Hernández. *Avances de investigación para la producción de leche en México. Macroproyecto "Mejoramiento de la productividad, competitividad y sustentabilidad de la cadena productiva de leche de bovino en México" Etapa 1*. Libro Científico Núm. 4., INIFAP, Centro de Investigación Regional Norte Centro. Coahuila, 2012.

13. S. Pulendran, R. Speed and R. Widing. The antecedents and consequences of market orientation in Australia. *Australian Journal of Management*, 25, 2, 2000.

14. J.C. Rey, V. Trujillo, A. Sánchez, E. Mazzani, F. Carreño, R. Salazar and J. González. *Evaluación del Contexto Institucional para Definir el Rol de los INIAs en la Investigación Y Gestión de los Recursos Naturales*. International Service for National Agricultural Research, The Hague, Netherlands, 1999.

15. A. R. Saldaña, J. A. Espinosa-García, G. Moctezuma-López, A. Ayala, C. A. Tapia, R. M. Ríos, S. M. Valle, and J.M. Gomes de C. *Proyecto Quo Vadis: El Futuro de la Investigación Agropecuaria y Forestal y la Inovación Institucional de México*. INIFAP, México D.F., 2006.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Estimating the model with fixed and random effects by a robust method*

Jan Ámos Víšek[1]

Institute of Economic Studies, Faculty of Social Sciences, Charles University, Opletalova 26, Prague, the Czech Republic (E-mail: visek@fsv.cuni.cz)

Abstract. Regression model with fixed and random effects estimated by modified versions of the *Ordinary Least Squares* (OLS) is a standard tool of panel data analysis. However, it is vulnerable to the bad effects of influential observations (contamination or atypical observations). Robustification of the OLS by means of the *Least Weighted Squares* (LWS) and modifications of LWS in an analogous way as the OLS were modified for estimating the fixed- and random-effects-model rid of this disadvantage. The numerical study reveals the reliability of the respective algorithm. The results of this study were collected in a file which is possible to find on web, address is given below. Patterns of these results were included into the paper (extent of which is limited). The possibility to reach nearly the full efficiency of estimation in the case when there are no influential points is demonstrated at the mentioned file on the web.

Keywords: Linear regression model, the least weighted squares, fixed and random effects, numerical simulations.

1 An introduction

Atypical observations in a data set can cause misleading conclusions of the regression analysis. That was the reason for building up the robust methods for identifying the true underlying model in the various frameworks. One such framework is the regression model with the fixed and random effects and we can meet with proposals of robust estimators of this model based on the idea of *M-estimators*, of the *Least Median of Squares* (LMS) or of the *Least Trimmed Squares* (LTS), see e. g. Bramati and Croux [2], Dehon [6], Kott [12], Rocke [14], Veradi and Croux [17] or Veradi and Wagner [18]. The former require studentization of residuals - to reach the *scale-* and *regression-equvivariance* of the estimators in question, see Bickel [1]. Bickel also showed that the studentization has to be done by a *scale-equivariant* and *regression-invariant* estimator of standard deviation of disturbances but there are only a few robust estimators of standard deviation of disturbances which are *scale-equivariant* and *regression-invariant*, see Croux and Rousseeuw [4], Jurečková and Sen [11] or Víšek [26]. Moreover, all these estimators employs the preliminary robust *scale-* and *regression-equvivariant* estimator of regression model. The latter proposals based on LMS and LTS can be very sensitive to "inliers", see Víšek [20] or [27]. The discussion on the sensitivity of highly robust estimators to inliers has been started by Hettmansperger and Sheather [9]. Although their shocking results appeared later to be due to the bad algorithm they used (see again Víšek [20] and also Boček and Lachout [3]), they gave an inspiration for studies of sensitivity of robust methods to the changes of data inside the main cloud of them (Víšek [20],[22],[24]). The *Least Weighted Squares* (LWS) which employs the

* Research was supported by grant of GA ČR number 13-01930S panel P402.

idea of smooth decrease of the influence of atypical observations by means of prescribing the weights to the order statistics of the squared residuals rather than to the squared residuals directly, rid of both these problems. The advantage of the method is that it can be adjusted to the level of contamination by an adaptive selection of the weights (which is possible to perform in a reasonable time due to the fast algorithm we have for computing the estimator).

2 The model and estimators

For any n, $T \in N$ (the set of all positive integers) and $\beta^0 \in R^p$ (p-dimensional Euclidean space) we will study the linear regression model

$$Y_{it} = X'_{it}\beta^0 + u_i + e_{it}, \quad i = 1, 2, ..., n, \quad t = 1, 2, ..., T \tag{1}$$

with Y_{it}'s being the response variables, X_{it}'s staying for p-dimensional random explanatory variables, u_i's for the effects and e_{it}'s for the disturbances.

Definition 1. If $cov(X_{itj}, u_i) = 0$ for all $i = 1, 2, ..., n$, $t = 1, 2, ..., T$ and $j = 1, 2, ..., p$, the model (1) is called the *random effects model*, otherwise we speak about the *fixed effects model*, see e. g. Judge [10].

For any $\beta \in R^p$, $i \in N$ and $t \in \{1, 2, ..., T\}$ we will consider the residual of the (i, t)-th observation given as

$$r_{it}(\beta) = Y_{it} - X'_{it} \cdot \beta \tag{2}$$

and we denote by $r^2_{(\ell)}(\beta)$ the ℓ-th order statistic among the squared residuals $r^2_{it}(\beta), i = 1, 2, ..., n, t = 1, 2, ..., T$, i. e. we have

$$r^2_{(1)}(\beta) \le r^2_{(2)}(\beta) \le ... \le r^2_{(n \cdot T)}(\beta).$$

Finally, let's recall the classical *Ordinary Least Squares* and their robust version, the *Least Weighted Squares*.

Definition 2. Let $w_\ell \in [0, 1]$, $\ell = 1, 2, ..., n \cdot T$ be weights. The estimators

$$\hat\beta^{(OLS,n,T)} = \underset{\beta \in R^p}{arg\,min} \sum_{i=1}^{n} \sum_{t=1}^{T} r^2_{it}(\beta) \text{ and } \hat\beta^{(LWS,n,T,w)} = \underset{\beta \in R^p}{arg\,min} \sum_{\ell=1}^{n \cdot T} w_\ell\, r^2_{(\ell)}(\beta)$$

are called the *Ordinary Least Squares* (*OLS*) and the *Least Weighted Squares* (*LWS*) estimator, respectively (Víšek [21]).

Remark *First of all, notice that $\hat\beta^{(LWS,n,T,w)}$ is L-estimator, as it assigns weights to the order statistics of squared residuals, and in the general case it is not S-estimator. It follows from the fact that $w_\ell \cdot r^2_{(\ell)}(\beta)$ need not be monotone in ℓ (even for the monotone weights) while the elements of the sum defining S-estimators is assumed to be monotone, see e. g. Davies [5], see also Šindelář [16]. Special cases of $\hat\beta^{(LWS,n,T,w)}$ are the Least Median of Squares (Rousseeuw [15]) and the Least Trimmed Squares (Hampel et al. [8]).*

Denoting the empirical distribution function of the absolute values of residuals $r_{it}(\beta)$'s by $F^{(n)}_\beta(r)$, a straightforward derivation shows that $\hat\beta^{(LWS,n,T,w)}$ is one of solutions of the *normal equations*

$$\sum_{i=1}^{n} \sum_{t=1}^{T} w\left(F^{(n)}_\beta(|r_{it}(\beta)|)\right) X_{it}(Y_{it} - X'_{it} \cdot \beta) = 0. \tag{3}$$

Then the generalization of the classical Kolmogorov-Smirnov result on uniform convergence of empirical d. f. for the *regression model framework* (see Víšek [28][1])

$$\sup_{\beta\in R^p}\sup_{-\infty<r<\infty}\sqrt{n}\left|F_\beta^{(n)}(r)-(n\cdot T)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}F_\beta^{(it)}(r)\right|=O_p(1)$$

allows to prove the consistency of $\hat\beta^{(LWS,n,T,w)}$ under the following conditions:

Conditions $C\,1$ *The sequence* $\left\{\left\{(X_{it}',e_{it})'\right\}_{t=1}^{T}\right\}_{i=1}^{\infty}$ *is sequence* T*-tuples of mutually independent* $(p+1)$*-dimensional random variables (r.v.'s) distributed according to distribution functions (d.f.)* $F_{X,e_{it}}(x,r)=F_X(x)\cdot F_{e_{it}}(r)$ *where* $F_{e_{it}}(r)=F_e(r\sigma_{it}^{-1})$ *with* $I\!Ee_{it}=0$, $\mathrm{var}\,(e_{it})=\sigma_{it}^2$ *and* $\lim\limits_{n\to\infty}\frac{1}{n\cdot T}\sum_{i=1}^{n}\sum_{t=1}^{T}\sigma_{it}^2=1$. *Moreover,* $F_e(r)$ *is absolutely continuous with bounded density* $f_e(r)$. *Further, there is* $q>1$ *so that* $I\!E\left\|X_1\right\|^{2q}<\infty$. *Finally,* $\{u_t\}_{t=1}^{\infty}$ *is a sequence of independent and identically distributed r. v.'s, independent from the sequence* $\left\{\left\{(X_{it}',e_{it})'\right\}_{t=1}^{T}\right\}_{i=1}^{\infty}$, *with d. f.* $F_u(u)$ *with finite* σ_u^2.

Conditions $C\,2$ *The weights are generated as* $w_\ell=w\left(\frac{\ell-1}{n\cdot T}\right)$ *where* $w(u)$ *is a continuous, nonincreasing function,* $w:[0,1]\to[0,1]$ *with* $w(0)=1$.

Conditions $C\,3$ *Put* $F_\beta^{(it)}(r)=P\left(|r_{it}(\beta)|\le r\right)$ *(remember (2)). For any* $n\in N$ *there is the only solution of*

$$I\!E\left\{\sum_{i=1}^{n}\sum_{t=1}^{T}\left[w\left(F_\beta^{(it)}(|r_{it}(\beta)|)\right)X_{it}\left(Y_{it}-X_{it}'\beta\right)\right]\right\}=0 \qquad (4)$$

namely β^0.

Theorem 1. *Let* **Conditions** $C\,1$, $C\,2$ *and* $C\,3$ *be fulfilled. Then any sequence* $\left\{\hat\beta^{(LWS,n,T,w)}\right\}_{n=1}^{\infty}$ *of the solutions of normal equations (3) is weakly consistent.*

The proof is a direct reformulation of the result Víšek [27] from the *cross-sectional-data framework* to the *panel-data framework* (possibly with fixed or random effects). The result can be - under a bit stronger conditions - strengthen to $\sqrt{n}$-consistency of $\hat\beta^{(LWS,n,T,w)}$, Víšek [25].

3 Fixed and random effects estimation

Let's recall classical solutions (assuming normality of disturbances). For model with random effects we have (see (1))

$$Y_{it}=X_{it}'\beta^0+v_{it},\qquad i=1,2,...,n,\quad t=1,2,...,T \qquad (5)$$

where $v_{it}=u_i+e_{it}$, $I\!Ev_{it}=0$, $\mathrm{cov}(X_{itj},v_{it})=0$ *and* $I\!E\left[v_{it},v_{is}\right]=\mathrm{var}(u_i)=\sigma_u^2$. It implies that $\hat\beta^{(OLS,n,T)}$ is not efficient due to the correlation between disturbances. The estmation can be improved either by utilizing the *Generalized Least Squares* or - equivalently - by considering slightly modified data

$$\tilde Y_{it}=Y_{it}-\lambda\overline{Y}_i\ \text{ and }\ \tilde X_{it}=X_{it}-\lambda\overline{X}_i,\ \text{ with }\ \lambda=1-\sigma_e^2\cdot\left(\sigma_e^2+T\cdot\sigma_u^2\right)^{-1}, \qquad (6)$$

[1] The main technical tool for the proof of this generalization was the Skorohod embedding into Wiener process, see Portnoy [13].

(where $\overline{Y}_i = T^{-1}\sum_{t=1}^{T} Y_{it}$ and $\overline{X}_i = T^{-1}\sum_{t=1}^{T} X_{it}$) and applying $\hat{\beta}^{(OLS,n,T)}$ on $\tilde{Y}_{it}$'s and $\tilde{X}_{it}$'s, see e. g. Wooldridge [29]. The variances σ_e^2 and σ_u^2 can be estimated employing $r_{it}(\hat{\beta}^{(OLS,n,T)})$ by the classical estimators for the variance of disturbances, say $\hat{\sigma}_v^2$ and $\hat{\sigma}_u^2$, taking into account that $\sigma_e^2 = \sigma_v^2 - \sigma_u^2$ and then we can use $\hat{\lambda}$ instead of λ. Applying then *OLS* on the transformed data we obtain *RE-estimate* (which is below in tables denoted as $\hat{\beta}^{RE}$).

Robustification of this classical estimation consists of substituting $\hat{\beta}^{(OLS,n,T)}, \hat{\sigma}_v^2$ and $\hat{\sigma}_u^2$ by $\hat{\beta}^{(LWS,n,T,w)}, \hat{\sigma}_{LWS,v}^2$ and $\hat{\sigma}_{LWS,u}^2$, respectively. Such estimator is denoted in the tables below as the *Random Weighted Effects* estimator $\hat{\beta}^{RWE}$. However, we are not able to derive analytically an improvement (if any) of $\hat{\beta}^{RWE}$ compared to $\hat{\beta}^{LWS}$, hence the numerical study - we shall discuss the results in section Conclusions.

Assuming framework with fixed effects, we have in model (6) $I\!E v_{it} = 0$ with $\mathrm{cov}(X_{itj}, v_{it}) \neq 0$. It implies the inconsistency and biasedness of $\hat{\beta}^{(OLS,n,T)}$. Among all possible remedies it seems the most attractive *fixed-effect-estimation* (or alternatively called the *within-the-group transformation*) seems to be more attractive. It considers data

$$\tilde{Y}_{it} = Y_{it} - \overline{Y}_i \ \text{ and } \ \tilde{X}_{it} = X_{it} - \overline{X}_i, \ \text{ with } \overline{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it} \ \text{ and } \ \overline{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it} \quad (7)$$

($\tilde{X}_{it1} \equiv 0$). Hence, let's put $\tilde{V}_{itj} = \tilde{X}_{itj+1}$ and $\gamma_j^0 = \beta_{j+1}^0$ for $i = 1, 2, ..., n, t = 1, 2, ..., T$ and $j = 1, 2, ...p-1,$, we have

$$\tilde{Y}_{it} = \tilde{V}_{it}'\gamma^0 + e_{it}, \quad \text{with} \quad I\!E[\tilde{V}_{itj} \cdot e_{it}] = 0 \quad (8)$$

and hence $\hat{\gamma}^{(OLS,n,T)}$ is unbiased, consistent, etc., under normality of e_{it}'s even reaching Rao-Cramer lower bound for covariance matrix (if we consider ordering by means of positive definiteness). The intercept can be additionally estimated by $\hat{\beta}_1^{(OLS,n,T)} = \frac{1}{n}\sum_{i=1}^{n}\left[\overline{Y}_i - \overline{X}_i'\hat{\gamma}^{(OLS,n,T)}\right]$.

So, having estimated robustly mean values of response and explanatory variables (by *LWS* estimators for location) and employing transformation (7) with these estimates instead of with $\overline{Y}_i$'s and $\overline{X}_i$'s, we can finally employ $\hat{\gamma}^{(LWS,n,T,w)}$ (i. e. analogy of $\hat{\beta}^{(LWS,n,T,w)}$ for the model (8)) we compute the *Fixed Weighted Effects Estimator* (denoted below as $\hat{\beta}^{FWE}$).

4 Numerical study

Having generated data

$$\left\{ \left\{ \left\{ X_{it}^{(k)}, u_i^{(k)} \right\}_{t=1}^{20} \right\}_{i=1}^{50}, \ \left\{ \varepsilon_{it}^{(k)} \right\}_{i=1}^{50} \right\}_{k=1}^{500} \text{ and } \left\{ \left\{ \left\{ \sigma_{it}^{(k)} \right\}_{t=1}^{20} \right\}_{i=1}^{50} \right\}_{k=1}^{500}$$

where all variables in the first group were distributed according to the standard normal distribution while in the second group according to the uniform distribution on $[0.5, 1.5]$, we put $\tilde{X}_{itj}^{(k)} = X_{itj}^{(k)} + u_i^{(k)}, e_{it}^{(k)} = \varepsilon_{it}^{(k)} \cdot \sigma_{it}^{(k)}$ for $i = 1, 2, ..., 50 \ \ t = 1, 2, ..., 20$ and $j = 1, 2, ..., 5$. Then employing (1) we computed two groups (for the second one we employed $\tilde{X}_{it}^{(k)}$'s instead of $X_{it}^{(k)}$'s)

$$\left\{ \left\{ \left\{ Y_{it}^{(k)}, X_{it}^{(k)} \right\}_{t=1}^{20} \right\}_{i=1}^{50} \right\}_{k=1}^{500} \quad \text{and} \quad \left\{ \left\{ \left\{ \tilde{Y}_{it}^{(k)}, \tilde{X}_{it}^{(k)} \right\}_{t=1}^{20} \right\}_{i=1}^{50} \right\}_{k=1}^{500}.$$

The first group represents data for the model with random effects, the second one the model with fixed effects. Then the estimates of regression coefficients were computed - for both groups all estimators, i. e. $\hat{\beta}^{(OLS,n,T)}$, $\hat{\beta}^{FE}$, $\hat{\beta}^{RE}$, $\hat{\beta}^{(LWS,n,T,w)}$, $\hat{\beta}^{FWE}$ and $\hat{\beta}^{RWE}$ were computed, to offer the reader a possibility to create an idea how e. g. estimator proposed for the model with the fixed effects works when estimating coefficients of the model with random effects etc.. So, we obtained, say $\left\{ \hat{\beta}^{(index,k)} = (\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_5)' \right\}_{k=1}^{500}$

(where index "attains values" *OLS, RE, FE, LWS, RWE* and *FWE* indicating the method employed for the computation).

The weight function $w(r) : [0,1] \to [0,1]$ was selected in an optimal way - see a brief discussion in the file on web (see the address at the bottom of this page). We used the weights:

$$w_i = 1 \quad \text{for} \quad i = 1, 2, ..., h, \quad h \in \{1, 2, ..., n\},$$
$$w_i = 0 \quad \text{for} \quad i = g, g+1, ..., n, \quad g \in \{1, 2, ..., n\}, \; g > h$$

and for $h \le i \le g$ the weights w_i's decreased linearly from 1 to 0.

The empirical means and empirical variances of estimates of coefficients (over these 500 repetitions indicated above)

$$\hat{\beta}_j^{(index)} = \frac{1}{500} \sum_{k=1}^{500} \hat{\beta}_j^{(index,k)} \text{ and } \widehat{\text{var}}\left(\hat{\beta}_j^{(index)}\right) = \frac{1}{500} \sum_{k=1}^{500} \left[\hat{\beta}_j^{(index,k)}\right]^2 - \left[\hat{\beta}_j^{(index)}\right]^2$$

are reported below in tables[2].

TABLE 1

True coeffs β^0	**3**	**1**	**2**	**-4**	**5**
These coefficients were used in the whole numerical study.					
The disturbances are heteroscedastic, both, the disturbances and the effects, are independent from explanatory variables. Contamination is created by outliers, values of which are equal to -2.5 multiple of original values of response variable.					
Contamination level is equal to 0.5%.					
$\hat{\beta}^{OLS}$(var($\hat{\beta}^{OLS}$))	$2.95_{(0.016)}$	$0.93_{(0.040)}$	$1.87_{(0.042)}$	$-3.73_{(0.060)}$	$4.63_{(0.076)}$
$\hat{\beta}^{FE}$(var($\hat{\beta}^{FE}$))	$0.00_{(0.000)}$	$0.93_{(0.039)}$	$1.87_{(0.041)}$	$-3.73_{(0.059)}$	$4.63_{(0.074)}$
$\hat{\beta}^{RE}$(var($\hat{\beta}^{RE}$))	$2.95_{(0.016)}$	$0.93_{(0.039)}$	$1.87_{(0.041)}$	$-3.73_{(0.059)}$	$4.63_{(0.074)}$
$\hat{\beta}^{LWS}$(var($\hat{\beta}^{LWS}$))	$3.00_{(0.015)}$	$0.99_{(0.007)}$	$2.00_{(0.006)}$	$-4.00_{(0.006)}$	$4.99_{(0.006)}$
$\hat{\beta}^{FWE}$(var($\hat{\beta}^{FWE}$))	$0.00_{(0.000)}$	$1.00_{(0.003)}$	$2.00_{(0.003)}$	$-4.00_{(0.003)}$	$4.99_{(0.003)}$
$\hat{\beta}^{RWE}$(var($\hat{\beta}^{RWE}$))	$3.01_{(0.158)}$	$1.00_{(0.003)}$	$2.00_{(0.003)}$	$-4.00_{(0.003)}$	$4.99_{(0.003)}$
Contamination level is equal to 3%.					
$\hat{\beta}^{OLS}$(var($\hat{\beta}^{OLS}$))	$2.66_{(0.038)}$	$0.62_{(0.141)}$	$1.24_{(0.159)}$	$-2.49_{(0.201)}$	$3.11_{(0.196)}$
$\hat{\beta}^{FE}$(var($\hat{\beta}^{FE}$))	$0.00_{(0.000)}$	$0.62_{(0.147)}$	$1.24_{(0.160)}$	$-2.49_{(0.206)}$	$3.10_{(0.196)}$

[2] We have not enough space for presenting more results of numerical study but they can be found on http://samba.fsv.cuni.cz/~visek/asdma2013/.

$\hat{\beta}^{RE}(\text{var}(\hat{\beta}^{RE}))$	$2.66_{(0.038)}$	$0.62_{(0.141)}$	$1.24_{(0.159)}$	$-2.49_{(0.202)}$	$3.10_{(0.196)}$
$\hat{\beta}^{LWS}(\text{var}(\hat{\beta}^{LWS}))$	$3.00_{(0.014)}$	$1.00_{(0.006)}$	$2.00_{(0.006)}$	$-4.00_{(0.006)}$	$5.00_{(0.005)}$
$\hat{\beta}^{FWE}(\text{var}(\hat{\beta}^{FWE}))$	$0.00_{(0.000)}$	$1.00_{(0.004)}$	$1.99_{(0.004)}$	$-3.98_{(0.004)}$	$4.97_{(0.004)}$
$\hat{\beta}^{RWE}(\text{var}(\hat{\beta}^{RWE}))$	$3.51_{(0.686)}$	$1.00_{(0.004)}$	$1.99_{(0.004)}$	$-3.98_{(0.004)}$	$4.97_{(0.004)}$
Contamination level is equal to 15%.					
$\hat{\beta}^{OLS}(\text{var}(\hat{\beta}^{OLS}))$	$1.39_{(0.096)}$	$-0.23_{(0.249)}$	$-0.48_{(0.258)}$	$0.88_{(0.367)}$	$-1.12_{(0.387)}$
$\hat{\beta}^{FE}(\text{var}(\hat{\beta}^{FE}))$	$0.00_{(0.000)}$	$-0.24_{(0.255)}$	$-0.48_{(0.271)}$	$0.88_{(0.371)}$	$-1.11_{(0.402)}$
$\hat{\beta}^{RE}(\text{var}(\hat{\beta}^{RE}))$	$1.39_{(0.096)}$	$-0.23_{(0.249)}$	$-0.48_{(0.259)}$	$0.88_{(0.365)}$	$-1.12_{(0.387)}$
$\hat{\beta}^{LWS}(\text{var}(\hat{\beta}^{LWS}))$	$2.99_{(0.011)}$	$0.99_{(0.006)}$	$1.99_{(0.005)}$	$-3.98_{(0.006)}$	$4.98_{(0.005)}$
$\hat{\beta}^{FWE}(\text{var}(\hat{\beta}^{FWE}))$	$0.00_{(0.000)}$	$0.94_{(0.011)}$	$1.88_{(0.012)}$	$-3.76_{(0.013)}$	$4.70_{(0.017)}$
$\hat{\beta}^{RWE}(\text{var}(\hat{\beta}^{RWE}))$	$9.05_{(30.625)}$	$0.95_{(0.012)}$	$1.89_{(0.013)}$	$-3.79_{(0.014)}$	$4.73_{(0.017)}$

TABLE 2

True coeffs $\boldsymbol{\beta^0}$	**3**	**1**	**2**	**-4**	**5**
The disturbances are heteroscedastic, independent from explanatory variables but the effects are correlated with them. Contamination by leverage points, values of which are equal to the 10th of original values and response with minus sign.					
Contamination level is equal to 0.5 %					
$\hat{\beta}^{OLS}(\mathbf{var}(\hat{\beta}^{OLS}))$	$2.36_{(1.761)}$	$-0.00_{(1.922)}$	$0.66_{(2.852)}$	$-2.91_{(3.543)}$	$2.35_{(6.582)}$
$\hat{\beta}^{FE}(\mathbf{var}(\hat{\beta}^{FE}))$	$0.00_{(0.000)}$	$-0.42_{(1.555)}$	$0.24_{(2.073)}$	$-3.39_{(4.343)}$	$1.97_{(5.629)}$
$\hat{\beta}^{RE}(\mathbf{var}(\hat{\beta}^{RE}))$	$0.88_{(15.404)}$	$-0.25_{(1.514)}$	$0.41_{(2.185)}$	$-3.30_{(3.625)}$	$2.22_{(5.867)}$
$\hat{\beta}^{LWS}(\mathbf{var}(\hat{\beta}^{LWS}))$	$3.00_{(0.002)}$	$1.20_{(0.002)}$	$2.20_{(0.002)}$	$-3.81_{(0.002)}$	$5.20_{(0.002)}$
$\hat{\beta}^{FWE}(\mathbf{var}(\hat{\beta}^{FWE}))$	$0.00_{(0.000)}$	$1.00_{(0.003)}$	$1.99_{(0.003)}$	$-4.01_{(0.003)}$	$4.99_{(0.003)}$
$\hat{\beta}^{RWE}(\mathbf{var}(\hat{\beta}^{RWE}))$	$3.00_{(0.018)}$	$1.07_{(0.006)}$	$2.06_{(0.006)}$	$-3.94_{(0.006)}$	$5.07_{(0.005)}$
Contamination level is equal to 3%					
$\hat{\beta}^{OLS}(\mathbf{var}(\hat{\beta}^{OLS}))$	$-0.87_{(1.575)}$	$-0.81_{(0.459)}$	$-1.40_{(0.628)}$	$1.86_{(1.601)}$	$-3.01_{(1.339)}$
$\hat{\beta}^{FE}(\mathbf{var}(\hat{\beta}^{FE}))$	$0.00_{(0.000)}$	$-0.65_{(0.470)}$	$-1.21_{(0.555)}$	$2.04_{(1.894)}$	$-2.80_{(1.072)}$
$\hat{\beta}^{RE}(\mathbf{var}(\hat{\beta}^{RE}))$	$-2.79_{(1.786)}$	$-0.99_{(0.304)}$	$-1.56_{(0.476)}$	$1.66_{(1.564)}$	$-3.16_{(1.259)}$
$\hat{\beta}^{LWS}(\mathbf{var}(\hat{\beta}^{LWS}))$	$3.00_{(0.004)}$	$1.19_{(0.006)}$	$2.18_{(0.010)}$	$-3.82_{(0.012)}$	$5.19_{(0.007)}$
$\hat{\beta}^{FWE}(\mathbf{var}(\hat{\beta}^{FWE}))$	$0.00_{(0.000)}$	$0.52_{(0.232)}$	$1.51_{(0.237)}$	$-4.40_{(0.151)}$	$4.48_{(0.268)}$
$\hat{\beta}^{RWE}(\mathbf{var}(\hat{\beta}^{RWE}))$	$3.01_{(2.816)}$	$0.56_{(0.215)}$	$1.54_{(0.214)}$	$-4.35_{(0.141)}$	$4.50_{(0.272)}$
Contamination level is equal to 15 %					
$\hat{\beta}^{OLS}(\mathbf{var}(\hat{\beta}^{OLS}))$	$-2.62_{(0.019)}$	$-0.95_{(0.004)}$	$-1.88_{(0.005)}$	$3.65_{(0.012)}$	$-4.64_{(0.009)}$
$\hat{\beta}^{FE}(\mathbf{var}(\hat{\beta}^{FE}))$	$0.00_{(0.000)}$	$-0.48_{(0.082)}$	$-1.38_{(0.071)}$	$3.96_{(0.084)}$	$-4.04_{(0.066)}$
$\hat{\beta}^{RE}(\mathbf{var}(\hat{\beta}^{RE}))$	$-3.41_{(0.009)}$	$-1.06_{(0.002)}$	$-1.98_{(0.004)}$	$3.54_{(0.012)}$	$-4.75_{(0.008)}$
$\hat{\beta}^{LWS}(\mathbf{var}(\hat{\beta}^{LWS}))$	$3.00_{(0.004)}$	$1.19_{(0.008)}$	$2.19_{(0.018)}$	$-3.80_{(0.006)}$	$5.19_{(0.009)}$
$\hat{\beta}^{FWE}(\mathbf{var}(\hat{\beta}^{FWE}))$	$0.00_{(0.000)}$	$-0.18_{(0.593)}$	$0.67_{(0.596)}$	$-4.27_{(0.394)}$	$3.19_{(1.053)}$
$\hat{\beta}^{RWE}(\mathbf{var}(\hat{\beta}^{RWE}))$	$-0.10_{(14.701)}$	$-0.24_{(0.658)}$	$0.32_{(1.388)}$	$-3.24_{(6.859)}$	$2.12_{(7.626)}$

5 Conclusions

It is evident that even very low level of contamination (in the form of outliers as well as of leverage points) causes problems to $\hat{\beta}^{(OLS,n)}$ in efficiency and increasing level of contamination brings problems also in bias. Robustfication by $\hat{\beta}^{(LWS,n,T,w)}$ helps a lot and $\hat{\beta}^{FWE}$ and $\hat{\beta}^{RWE}$ attain an improvement even in efficiency of robust estimation. Notice however that improvement is not extremely significant. Moreover it is true only up to 10% for contamination in the case of outliers and up to 7% of contamination in the case of leverage points. For higher levels of contamination $\hat{\beta}^{FWE}$ and $\hat{\beta}^{RWE}$ even worsen the estimation and it is better to employ only $\hat{\beta}^{(LWS,n,T,w)}$. Further, the deviating from normal distribution of disturbances would probably worsen the situation - similarly as in the case of well-known results by Ronald Aylmer Fisher [7]. But it has to be confirmed by much more extended simulations.

At the end of numerical study we have briefly addressed the problem of efficiency of robust estimation and optimality of the selection of weight function.. Firstly, we assumed high contamination and hence we tried to depress the influential points very resolutely. But the data were generated without any contamination. Our selection of weights resulted in the large variances of $\hat{\beta}^{(LWS,n,T,w)}$, $\hat{\beta}^{FWE}$ and $\hat{\beta}^{RWE}$ but the values of estimators were the same as values $\hat{\beta}^{(OLS,n,T)}$, $\hat{\beta}^{FE}$ and $\hat{\beta}^{RE}$. The fact that the values of non-robust estimators were not different from the values of robust estimators indicated the absence or at least low level of contamination. That was why we modified weights so that they depressed less a possible contamination but the situation repeated. So, we modified the weights once again, adjusting them for a very low contamination. For such an adjustment of weights the efficiency of both estimators is the same. Due to the speed of the algorithm we can afford it for the real data without any problems as well as in the simulation study. The results can be seen in file which is on the address http://samba.fsv.cuni.cz/~visek/asdma2013/.

References

1.Bickel, P. J. (1975): One-step Huber estimates in the linear model. *JASA 70, 428–433.*

2.Bramati, M. C., Croux, C. (2007): Robust estimators for the fixed effects panel data model. *Econometric Journal 10, 521 - 540.*

3.Boček, P., P. Lachout (1993): Linear programming approach to LMS-estimation. *Memorial volume of Comput. Statist. & Data Analysis 19(1995), 129 - 134.*

4.Croux, C., Rousseeuw, P. J. (1992): A class of high-breakdown scale estimators based on subranges. *Comm. in Statist.- Theory and Methods 21, 1935 - 1951.*

5.Davies, L. (1990): The asymptotics of S-estimators in the linear regression model. *Annals of Statistics 18, 1651 - 1675.*

6.Dehon, C., Gassner, M., Veradi, V. (2009): Beware of 'good' outliers and overoptimistic conclussios. *Oxford Bull. of Economics and Statist. 71, 437 - 452.*

7.Fisher, R. A. (1920): A mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the mean squares error. *Mon. Not. Roy. Astr. Soc. vol. 80, pp. 758–770.*

8.Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986): *Robust Statistics – The Approach Based on Influence Functions.* New York: J.Wiley.

9.Hettmansperger, T. P., S. J. Sheather (1992): A Cautionary Note on the Method of Least Median Squares. *The American Statistician 46, 79–83.*
10.Judge, G. G., Griffiths, W. E.,Hill,R. C., Lutkepohl, H., Lee, T. C. (1985): *The Theory and Practice of Econometrics.* New York: J.Wiley.
11.Jurečková, J., Sen, P. K. (1984): On adaptive scale-equivariant M-estimators in linear models. *Statistics and Decisions 2 (1984), Suppl. Issue No. 1.*
12.Kott,P. (1989): Robust small domain estimation using random effects modelling. *Survey Methodology 15, 1 - 12.*
13.Portnoy, S. (1983): Tightness of the sequence of empiric c.d.f. processes defined from regression fractiles. In *Robust and Nonlinear Time-Series Analysis (J. Franke, W. Hărdle, D. Martin, eds.), 231 - 246. Springer-Verlag, NY 1983.*
14.Rocke, D. M. (1991): Robustness and balance in the mixed models. *Biometrics 47, 303 - 309.*
15.Rousseeuw, P.J. (1984): Least median of square regression. *JASA 79, 871-880.*
16.Šindelář, J. (1994): On L-estimators viewed as M-estimators. *Kybernetika,30 (1994), 5, 551–562.*
17.Veradi,V., Croux, C. (2009): Robust regression in Stata. *The Stata Journal 9 (3), 439 - 453.*
18.Veradi,V., Wagner, J. (2010): Robust estimation of of linear fixed effects panel data model with an application to the exporter productivity premium. *Discussion paper No 4928, IZA, Bonn.*
19.Víšek, J. Á. (1996): A cautionary note on the method of Least Median of Squares reconsidered. *Trans. Twelfth Prague Conf. on Inform. Theory, Statist. Dec. Functions and Rand. Proc., Academy of Sciences of the Czech Republic, 254 - 259.*
20.Víšek, J. Á. (1996): Sensitivity analysis of M-estimates. *Annals of the Institute of Statistical Mathematics 48, 469-495.*
21.Víšek, J. Á. (2000): Regression with high breakdown point. *Proceeding of the conference ROBUST 2000, 324 - 356, eds. J. Antoch & G. Dohnal, published by The Union of the Czech Mathematicians and Physicists 2001.*
22.Víšek, J. Á. (2002): Sensitivity analysis of M-estimates of nonlinear regression model: Influence of data subsets. *Annals of the Institute of Statistical Mathematics 54, 261 - 290.*
23.Víšek, J. Á. (2006): Instrumental Weighted Variables - algorithm. *Proc. COMPSTAT 2006, eds. A. Rizzi & M. Vichi, Physica-Verlag (Springer Company) Heidelberg 2006, 777-786.*
24.Víšek, J. Á. (2006): The least trimmed squares. Sensitivity study. *Proc. Prague Stochastics 2006, eds. M. Hušková & M. Janžura, matfyzpress, 728-738.*
25.Víšek, J. Á. (2010): Weak $\sqrt{n}$-consistency of the least weighted squares under heteroscedasticity. *Acta Universit. Carolinae., Mathematica, 2/51, 71 - 82.*
26.Víšek, J. Á. (2010): Robust error-term-scale estimate. *IMS Collections. Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: Festschrift for Jana Jurečková, Vol. 7(2010), 254 - 267.*
27.Víšek, J. Á. (2011): Consistency of the least weighted squares under heteroscedasticity. *Kybernetika 47 , 179-206.*
28.Víšek, J. Á. (2011): Empirical distribution function under heteroscedasticity. *Statistics 45, 497-508.*
29.Wooldridge, J. M. (2006): *Introductory Econometrics. A Modern Approach.* MIT Press, Cambridge, Massachusetts, second edition 2009.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

A Variable Window Scan Statistic for MA(1) Process

Xiao Wang and Joseph Glaz

Department of Statistics, University of Connecticut, Storrs, CT 06269 USA

Abstract. Approximations for the distribution of a scan statistic for independent and identically distributed (iid) observations from a normal distribution have been investigated in Glaz *et al.*[5]. Based on these approximations, a variable window scan statistic has been introduced in Wang and Glaz[6]. In this article we derive a product-type approximation and Bonferroni-type inequalities for the distribution of a scan statistic for observations from a moving average process of order one, where the error component has a normal distribution with mean of zero and variance of one. We also investigate the performance of an approximation for a sequence of moving sums of 1-dependent observations that has been derived in Haiman[7,8]. A nice feature of this approximation is that an error bound can be evaluated as well. Numerical results to evaluate the accuracy of the approximations and inequalities are obtained via an algorithm established by Genz and Bretz[9]. We also present an algorithm to implement a variable window scan statistic.
Keywords: Moving Sum, Approximation and Inequality, Moving Average, R Algorithm.

1 Introduction

The distribution of moving sums of a sequence of independent and identically distributed (iid) normal variables has been investigated in Bauer and Hackl[10,11]. In Glaz and Naus[12], accurate approximations and inequalities were derived for the distribution of the maximum of moving sums of iid discrete random variables and a connection to a scan statistic, as a testing procedure for the detection of structural changes in an observed process, has been discussed. In Glaz *et al.*[5] accurate approximations and inequalities for the distribution of moving sums of iid normal random variables have been derived and a connection to scan statistics has been discussed. The performance of a variable window scan statistic for normal data has been investigated in Wang and Glaz[6]. In this article, we extend the approximation and inequalities investigated in Glaz *et al.*[5] to a sequence of observations generated from a MA(1) process. A variable window scan statistic, for detecting a local change in the mean of the Gaussian white noise component of a MA(1) process, is also introduced and an algorithm for its implementation is presented.

The article is organized as follows. In Section 2, a fixed window scan statistic is introduced based on moving sums of observations from a MA(1) process. Two approximations and Bonferroni-type inequalities are presented. In Section 3, a variable window scan statistic is derived and a simulation algorithm is proposed for comparing its performance with fixed window scan statistics. In Section 4, based on an algorithm in Genz and Bretz[9], for selected values of

the parameters in the MA(1) process, numeric results are presented to evaluate the approximations and inequalities discussed in Section 2. In Section 5, we present some concluding remarks.

2 Fixed window scan statistic

2.1 Moving sums of MA(1) observations

Let $X_1, ..., X_M, ...$ be a sequence of observations from a MA(1) process, $X_t = \omega_t + \theta\omega_{t-1}$, where ω_t is the Gaussian white noise component with mean $\mu = 0$ and some known variance σ^2. Without loss of generality, we will assume $\sigma^2 = 1$. Let $Y_{i-m+1,i} = \sum_{n=i-m+1}^{i} X_n$, $m \leq i \leq M$, denote a moving sum of length m. For a scanning window size $2 \leq m < M$, where M is the region to be scanned, define the *scan statistic*

$$S_{m,M} = \max_{m \leq i \leq M} \{Y_{i-m+1,i}\}. \tag{1}$$

which consists of a sequence of $M - m + 1$ moving sums of length m. It's noted that $\{X_i; 1 \leq i \leq M\}$ follow a multivariate normal distribution owing to the Gaussian distribution of the white noise component, thus $\{Y_{i-m+1,i}; m \leq i \leq M\}$ have a joint multivariate normal distribution with a mean vector of zeros and covariance matrix $\Sigma = \{\sigma_{i,j}\}$, where $\sigma_{i,j} = cov(Y_{i,i+m-1}, Y_{j,j+m-1})$. After a routine derivation, the following covariance matrix is obtained,

$$\sigma_{i,j} = \begin{cases} (j+m-i)(1+\theta^2) + 2\theta(j+m-i) & when\ i-j < m \\ \theta & when\ i-j = m \\ m(1+\theta^2) + 2\theta(m-1) & when\ i = j \\ 0. & otherwise \end{cases}$$

To study the distribution of the scan statistic defined above, for $2 \leq m \leq M$ and $-\infty < t < \infty$, let

$$G_{m,t}(M) = P\left(max\{Y_{i-m+1,m}; m \leq i \leq M\} < t\right). \tag{2}$$

Thus the distribution of the scan statistic $S_{m,M}$ is given by

$$P\left(S_{m,M} < t\right) = G_{m,t}(M). \tag{3}$$

And the probability of $S_{m,M}$ exceeding threshold t is $1 - G_{m,t}(M)$. When the values of m, M and t are clearly understood, we abbreviate $G_{m,t}(M)$ and $S_{m,M}$ to $G(M)$ and S_m, respectively.

2.2 Approximations

A product-type approximation for $G(M)$ can be obtained by following the same approach in Glaz *et al.*[5]. Let $M = Km + \upsilon$, where $K \geq 3, m \geq 2$ and

$0 \leq \upsilon \leq m-1$ and all of them are integers. Then, for $2 \leq L \leq K-1$

$$G(M) = P\left\{\max_{m \leq k \leq M} Y_{k-m+1,k} < t\right\} = P\left(\bigcap_{j=1}^{K} E_j\right)$$
$$= P\left(\bigcap_{i=1}^{L-1} E_i\right) \prod_{j=L}^{K} P\left(E_j | \bigcap_{h=1}^{j-1} E_h\right), \quad (4)$$

where for $1 \leq j \leq K-1$

$$E_j = \left(\max_{jm \leq k \leq (j+1)m} Y_{k-m+1,k} < t\right),$$

which denotes the event of not exceeding threshold t within a block of $m+1$ consecutive moving sums of length m, and

$$E_K = \left(\max_{Km \leq k \leq Km+\upsilon} Y_{k-m+1,k} < t\right).$$

By conditioning on most recent $L \geq 2$ events E_j in (4), we propose the following approximation for $G(M)$:

$$G(M) \approx P\left(\bigcap_{i=1}^{L-1} E_i\right) \left[\prod_{j=L}^{K-1} P\left(E_j | \bigcap_{h=j-L+1}^{j-1} E_h\right)\right] P\left(E_K | \bigcap_{p=K-L+1}^{K-1} E_p\right)$$
$$= P\left(\bigcap_{i=1}^{L} E_i\right) \left\{\prod_{j=L+1}^{K-1} \left[\frac{P\left(\bigcap_{h=j-L+1}^{j} E_h\right)}{P\left(\bigcap_{h=j-L+1}^{j-1} E_h\right)}\right]\right\} \frac{P\left(\bigcap_{p=K-L-1}^{K} E_p\right)}{P\left(\bigcap_{p=K-L1}^{K-1} E_p\right)}$$
$$= G((L+1)m) \left[\frac{G((L+1)m)}{G(Lm)}\right]^{K-L-1} \frac{G(Lm+\upsilon)}{G(Lm)}. \quad (5)$$

Equation (5) follows from the fact that for $2 \leq L \leq K-1$, the events $\{E_j; 1 \leq j \leq K-1\}$ and $\{E_{j-L+1} \cap \cdots \cap E_j; 2 \leq j \leq K-L+1\}$ are stationary and that for $1 \leq L \leq K-2$, $P\left(\bigcap_{i=1}^{L} E_i\right) = G((L+1)m)$ and $P(E_{K-L} \cap \cdots \cap E_K) = G((L+1)m+\upsilon)$. For $L=2$, the above approximation reduces to:

$$G(M) \approx G(3m) \left[\frac{G(3m)}{G(2m)}\right]^{K-3} \frac{G(2m+\upsilon)}{G(2m)}.$$

If $M = Km$, approximation (5) is simplified to

$$G(M) \approx G((L+1)m) \left[\frac{G((L+1)m)}{G(Lm)}\right]^{K-L-1}. \quad (6)$$

Another approximation for $G(M)$ can be seen in Haiman[7,8] for iid discrete random variables. These approximations are valid as well for continuous observations from a MA(1) process. An error bound is provided here as well. For any t and $M \geq 3m$, such that $1 - G(2m) \leq .025$ and $3.3M[1 - G(2M)]^2 \leq 1$, the following approximation for $G(M)$ is obtained :

$$G(M) \approx \frac{2G(2m) - G(3m)}{\left[1 + G(2m) - G(3m) + 2\left(G(2m) - G(3m)\right)^2\right]^{M/m-1}}, \tag{7}$$

with an error bound of approximately

$$3.3[1 - G(2m)]^2(M/m - 1). \tag{8}$$

2.3 Inequalities

Because of the dependency existed among X_i's, the inequalities derived for iid continuous observations in Glaz *et al.*[5] are not valid, so we now proceed to derive second order Bonferroni-type inequalities for $G(M) = 1 - P(S_m \geq t)$ defined above. For $1 \leq i \leq M - m + 1$, define

$$A_i = (Y_{i,i+m-1} \geq t),$$

and

$$B_i = (Y_{i,i+m-1} \leq t) = (A_i^c).$$

Then,

$$\begin{aligned} P(S_m \geq t) &= P(\bigcup_{i=1}^{M-m+1} A_i) \\ &= P(\bigcup_{i=1}^{M-m+1} B_i^c), \end{aligned}$$

we will derive Bonferroni-type inequalities in terms of the events B_i, where $1 \leq i \leq M - m + 1$. It follows from Hunter[1] that

$$\begin{aligned} P(\bigcup_{i=1}^{M-m+1} B_i^c) &\leq \sum_{i=1}^{M-m+1} P(B_i^c) - \sum_{i=1}^{M-m} P(B_i^c \bigcap B_{i+1}^c) \\ &= (M - m + 1)P(B_1^c) - (M - m)P(B_1^c \bigcap B_2^c). \end{aligned}$$

Substitute

$$P(B_1^c) = 1 - P(B_1)$$

and

$$P(B_1^c \bigcap B_2^c) = 1 - P(B_1 \bigcup B_2) = 1 - 2P(B_1) + P(B_1 \bigcap B_2)$$

in the above inequality to get

$$P(\bigcup_{i=1}^{M-m+1} B_i^c) \leq 1 - (M - m)P(B_1 \bigcap B_2) + (M - m - 1)P(B_1).$$

We get the following lower bound:

$$G(M) \geq (M-m)P(B_1 \bigcap B_2) - (M-m-1)P(B_1). \tag{9}$$

To derive an upper bound for $G(M)$, we employ the inequality from Kwerel[2] to get

$$G(M) \leq 1 - \frac{2s_1}{b} + \frac{2s_2}{b(b-1)}, \tag{10}$$

where b is integer part of $2 + 2s_2/s_1$,

$$s_1 = (M-m+1)(1-P(B_1))$$

and

$$\begin{aligned} s_2 &= \sum_{j=2}^{M-m+1}\sum_{i=1}^{j-1} P(B_i^c \bigcap B_j^c) = \sum_{j=2}^{M-m+1}\sum_{i=1}^{j-1}[1-2P(B_1)+P(B_i \bigcap B_j)] \\ &= 0.5(M-m+1)(M-m)[1-2P(B_1)] + \sum_{j=2}^{M-m+1}\sum_{i=1}^{j-1} P(B_i \bigcap B_j) \\ &= 0.5(M-m+1)(M-m)[1-2P(B_1)] + \sum_{j=2}^{m}(M-m+2-j)P(B_1 \bigcap B_j) \\ &\quad + 0.5(M-2m+1)(M-2m+2)[P(B_1)]^2. \end{aligned}$$

The last equality follows from the fact that B_i and B_j are independent if $j-i \geq m$, thus $P(B_i \bigcap B_j) = [P(B_1)]^2$.

3 Variable window scan statistic

The scan statistic and its approximations in Section 2 were derived with a scanning window of a fixed length m. It can be used to detect a local change, which happens within a sequence of m observations, in the mean of Gaussian white noise component from a MA(1) process. We test the null hypothesis, H_0, that assumes X_i's, $1 \leq i \leq M$, from a MA(1) process with the Gaussian white noise mean $\mu_0 = 0$ and variance 1, against the alternative hypothesis, H_1, that assumes the Gaussian white noise mean $\mu_1 > 0$ within a subsequence of m consecutive observations whose location is unknown. It has been proved that the generalized likelihood ratio test will reject the null hypothesis when S_m exceeds a value t, which is determined from $P(S_m \geq t|H_0) = \alpha$, where α is a specified significance level of the testing procedure.

But most of the time both the size and the location of a local change is unknown, and we propose here a variable window scan statistic based on a sequence of n fixed window size scan statistics: $S_{m_1}, ..., S_{m_n}$, where $2 \leq m_1 < m_2 < .... < m_n \leq M/4$, the lengths of the n scanning windows, are predetermined by the experimenter. For $1 \leq j \leq n$, let t_j be the observed value

of S_{m_j} and $p_j = P(S_{m_j} \geq t_j \mid H_0)$ be the associated p-value. To test H_0 vs H_1 we propose the following test statistic:

$$P^{(1)}_{\min} = \min\{p_j; 1 \leq j \leq n\}, \tag{11}$$

the *minimum P-value statistic*.

Since the exact distribution for the $P^{(1)}_{min}$ is unknown, we will need to, for a given significant level α, evaluate the critical value p_α,

$$P_{H_0}\left(P^{(1)}_{min} \leq p_\alpha\right) = \alpha,$$

via simulation. To simulate $P^{(1)}_{min}$, firstly we generate M observations from a MA(1) process under the null hypothesis, then we scan the whole region with multiple moving windows of sizes m_1, m_2, ... and m_n. The observed values of the fixed window scan statistics, $S_{m_1}, ..., S_{m_n}$, are recorded as $t_1, t_2, ..., t_n$, respectively. At the next stage, R algorithms for multivariate normal probabilities from Genz and Bretz[9] are employed to evaluate the observed p values $p_j = P(S_{m_j} \geq t_j \mid H_0)$, $1 \leq j \leq n$. Then the minimum of these p values is recorded and the whole process will be repeated N times. p_α will be recorded as the $\alpha * 100$ percentile of the simulated distribution of $P^{(1)}_{\min}$.

To evaluate the performance of the variable window scan statistic, $P^{(1)}_{min}$, against that of fixed window scan statistics, in detecting a local change of the mean, with an unknown size, in Gaussian white noise of a MA(1) process, we design the following simulation study to compare the power of these scan statistics.

1. Generate M observations from a MA(1) process with $\mu_0 = 0$ and $\sigma^2 = 1$, and in a specified window of length m replace them with observations with mean $\mu_1 > 0$.

2. Scan the whole region with pre-selected window sizes, and let $t_1, ..., t_n$ be the observed values of the fixed window scan statistics $S_{m_1}, ..., S_{m_n}$, respectively.

3. For a fixed window of length m_j and a specified significance level α, evaluate $p_j = 1 - P(S_{m_j} < t_j)$, and reject H_0 if $p_j < \alpha$.

4. For the variable window scan statistic, $P^{(1)}_{min} = min\{p_j; 1 \leq j \leq k\}$, reject H_0 if $P^{(1)}_{min} < p_\alpha$.

5. Repeat steps 1-4 N times and record the frequency we have rejected H_0 with the fixed and variable window scan statistics, respectively.

4 Numerical Results

In Tables 1 and 2, for selected values of parameters M, m, θ and t, approximations and bounds are evaluated for tail probabilities of the statistic defined

in Section 2, for observations from a MA(1) process with a Gaussian white noise mean 0 and variance 1. These numerical results are obtained from the R algorithm for the multivariate normal distribution in Genz and Bretz[9].

In Tables 1 and 2, $APPRX1$, LB and UB are evaluated via the approximation and bounds for $G(M)$ in (6), (9) and (10), respectively. $APPRX1$ represents the approximation in (6) when $L = 2$, and $L = 4$ in Table 2. $APPRX2$ and $ErrorBound$ are evaluated via the approximation and error bound in (7) and (8) respectively. 10^6 simulation runs were used to evaluate the sequences in Haiman's approximation. In Tables 1 and 2, the error bounds have been evaluated only for a restricted range of probabilities, as specified in Section 2, based on Haiman[8].

t	18	19	20	21	22	23
LB	.0353	.0173	.0079	.0035	.0015	.0006
APPRX 1	.0659	.0294	.0192	.0060	.0034	.0007
APPRX 2	.0592	.0282	.0128	.0053	.0021	.0013
Error Bound	.0091	.0061	.0043	.0026	.0020	.0011
UB	.0767	.0353	.0156	.0066	.0027	.0010
BruteForce	.0622	.0288	.0135	.0055	.0021	.0007

Table 1. Approximation and bounds for $P(S_{m,M} \geqslant t)$, M=1500, m=20, θ=0.1, L=2

t	18	19	20	21	22	23
LB	.0353	.0173	.0079	.0035	.0015	.0006
APPRX 1	.0623	.0303	.0164	.0079	.0025	.0011
APPRX 2	.0609	.0256	.0149	.0054	.0022	.0008
Error Bound	.0092	.0061	.0043	.0025	.0014	.0012
UB	.0767	.0353	.0156	.0066	.0027	.0010
BruteForce	.0596	.0295	.0141	.0065	.0023	.0010

Table 2. Approximation and bounds for $P(S_{m,M} \geqslant t)$, M=1500, m=20, θ=0.1, L=4

Based on the numerical results presented in Tables 1 and 2, one can conclude that the approximations and bounds are quite accurate. Their accuracy was confirmed by $BruteForce$ with 10,000 simulated sequences of moving sums of MA(1) observations. When comparing Table 2, with $L = 4$ in (6) to Table 1 with $L = 2$, we do see more accurate approximations when using a longer subsequence of observations.

5 Concluding Remarks

In this article approximations and inequalities have been derived for the distribution of moving sums of observations from a MA(1) process. These approximations and bounds yield approximations and bounds for scan statistic

probabilities that can be employed in detecting a local change in the mean of Gaussian white noise of a MA(1) process. Based on the numerical results in Section 4, we can conclude that the approximations and bounds for scan statistic probabilities associated with the distributions of moving sums are accurate and stable. Hence, the algorithm for a variable window scan statistic presented in Section 3 can be effectively implemented.

References

1.D. Hunter. An Upper Bound for the Probability of a Union. *Journal of Applied Probability*, 13, 597-603, 1976.
2.M.S. Kwerel. Most Stringent bounds on the probability of the union and intersection of m events. *Journal of Applied Probability*, 12, 612-619, 1975.
3.J. Glaz, J. Naus and S. Wallenstein. *Scan Statistics*, Springer, New York, 2001.
4.A. Westlund. Sequential moving sums of squares of OLS residuals in parameter stability testing. *Quality and Quantity*, 18, 261-273, 1984.
5.J. Glaz, J. Naus and X. Wang. Approximations and Inequalities for Moving Sums. *Methodology and Computing in Applied Probability*, 14, 3, 597–616, 2012.
6.X. Wang and J. Glaz. Variable window scan statistics for normal data. *Communications in Statistics - Theory and Methods*, 2013.
7.G. Haiman. First passage time for some stationary process. *Stochastic Processes and their Applications*, 80, 231-248, 1999.
8.G. Haiman. Estimating the distribution of one-dimensional discrete scan statistics viewed as extremes of 1-dependent stationary sequences. *Journal of Statistical Planning and Inference*, 137, 821-828, 2007.
9.A. Genz and F. Bretz.*Computation of Multivariate Normal and T Probabilities*, Springer, 2009
10.P. Bauer and P. Hackl. The use of MOSUMS for quality control. *Technometrics*, 20, 431-436, 1978.
11.P. Bauer and P. Hackl. An extension of the MOSUM technique to quality control. *Technometrics*, 22, 1-7, 1980.
12.J. Glaz and J. Naus. Tight bounds and approximations for scan statistics probabilities for discrete data. *Annals of Applied Probability*, 1, 306-318,1991.
13.G. Haiman and C. Preda. Estimation for the Distribution of Two-dimensional Discrete Scan Statistics. *Methodology and Computing in Applied Probability*, 8, 373-382, 2006.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Criteria for Transient Behavior of Symmetric Branching Random Walks on Z and $\mathbf{Z}^2$

Elena B. Yarovaya

Department of Probability Theory
Faculty of Mechanics and Mathematics
Lomonosov Moscow State University
Moscow, Russia
(e-mail: yarovaya@mech.math.msu.su)

Abstract. We consider a continuous-time symmetric branching random walk on $\mathbf{Z}^d$, where the particles are born and die at the origin. The transition rates of the random walk are assumed to be given by $a(x,y) \approx |y-x|^{-(d+\alpha)}$, where $\alpha \in (0,2)$ and $x, y \in \mathbf{Z}^d$. As a result, the condition of the finite variance of jumps is not valid, and random walk may be transient even on low-dimensional lattices ($d = 1, 2$). Criteria for transient behavior of branching random walk on $\mathbf{Z}$ and $\mathbf{Z}^2$ are obtained.
Keywords: Branching random walks, Heavy tails, Non homogeneous environments.

1 Introduction

We consider a branching random walk (BRW) on $\mathbf{Z}^d$ with one branching source at the origin, where particles are born and die. For such processes, the key point is the study of the Green function whose properties are closely related to the spectral properties of the operator determining the law of the underlying random walk. For large times, in terms of the Green function one can determine the mean number of returns of each particle to the source which may be treated as the mean number of random walk loops both with self-intersections and without them. This problem was solved for BRWs under assumption that the underlying random walk has the finite variance of jumps (references will be given below). For BRWs with the infinite variance of the random walk jumps, the answer to this question is of independent interest see, e.g., Yarovaya[12].

The symmetric BRWs with one source of branching were considered, for example, in Albeverio *et al.*[1], Bogachev and Yarovaya[3], Yarovaya[10] where the evolution of the mean numbers of particles at an arbitrary lattice point was defined by the structure of the spectrum of the linear operator $H = A + \beta\Delta_0$. Here, the random walk generator $A = (a(z))$, $z \in \mathbf{Z}^d$ is a bounded self-adjoint operator in $l^2(\mathbf{Z}^d)$, and the operator Δ_0 is given by the equality $\Delta_0 = \delta_0\delta_0^T$ where $\delta_0 = \delta_0(\cdot)$ denotes the column vector on $\mathbf{Z}^d$ assuming the unit value at the origin and zero otherwise. The parameter β

characterizes the source intensity. In these works, one of the basic assumptions was the condition of finiteness of the variance of random walk jumps:

$$\sum_{z\in\mathbf{Z}^d} |z|^2 a(z) < \infty. \tag{1}$$

If this condition is satisfied then the BRW is recurrent on $\mathbf{Z}$ and $\mathbf{Z}^2$ but loses this property on $\mathbf{Z}^d$ for $d \geq 3$, see, e.g., Spitzer[9], Yarovaya[10].

In Albeverio *et al.*[1], Bogachev and Yarovaya[3], Yarovaya[10] it was established also that for $d \geq 3$ there exists a threshold value $\beta_c > 0$ of the source intensity β such that when $\beta > \beta_c$ the numbers of particles grow exponentially both at an arbitrary point and over the entire lattice. At the same time, for $\beta \leq \beta_c$ there cannot be such exponential growth. From this point of view, the number β_c may be regarded as the critical value. The case $d = 1, 2$ differs from the case of higher dimensions in that here the threshold value vanishes: $\beta_c = 0$, and hence the exponential growth in the particle numbers is observed for all $\beta > 0$.

In Yarovaya[10] it was shown that $\beta_c = G_0^{-1}$ where $G_\lambda = G_\lambda(0,0)$ is the Green function of the operator A. Moreover, for $d \geq 3$ the inequality $G_0 < \infty$, or equivalently $\beta_c > 0$, follows from (1).

The random walks (without branching) with infinite variance of jumps have been investigated recently by numerous researchers. Let us mention the book of A. Borovkov and K. Borovkov[4] with a comprehensive bibliography. To the best of our knowledge, one of the first papers considering BRW with heavy tails without assumption (1) was Yarovaya[12]. Let us note that (1) is not satisfied if $a(z) \approx z^{-(d+\alpha)}$ with $\alpha \in (0, 2)$. In this case the variance of jumps may become infinite, and the corresponding random walk may become transient even on $\mathbf{Z}$ and $\mathbf{Z}^2$.

The difference between the BRW with finite variance of jumps and the BRW with heavy tails is especially demonstrative for the processes of pure birth at the source. The lack of death in such BRWs causes the inequality $\beta > 0$. Therefore, in the BRWs on $\mathbf{Z}$ and $\mathbf{Z}^2$ with finite variance of jumps only an exponential growth in the numbers of particles is possible for all values of β (there are no critical and subcritical cases), whereas for the BRWs with heavy tails the exponential growth takes place only when $\beta > \beta_c = G_0^{-1} > 0$. Therefore, determination of new conditions for BRWs with heavy tails, for which the expected number of visits of the initial state by a particle is finite, allows one to reconsider the previous results related to finite variance of jumps for the new situation where BRWs can be transient even on $\mathbf{Z}$ and $\mathbf{Z}^2$.

Section 2 gives the general facts about the operator A generating the symmetrical random walk on $\mathbf{Z}^d$. No assumptions about the variance of jumps are used here. In Sec. 3 a condition on the random walk is imposed under which (1) is violated. The main technical result of the paper is the estimate of the growth rate of the Fourier transform for transient intensities of the random walk with heavy tails formulated in Theorem 2. The proof of Theorem 2 for the case $d = 2$ complements the paper of Yarovaya[12]. Theorem 3

is the principal result for BRWs stating that the underlying random walk can be transient for low-dimensional lattices, in particular, for $\mathbf{Z}$ if $\alpha \in (0,1)$ and for $\mathbf{Z}^2$ if $\alpha \in (0,2)$. In Sec. 4 the obtained results are applied to the study of BRW with heavy tails, and the necessary and sufficient conditions are given for exponential growth in the number of particles both at an arbitrary point and the entire lattice.

2 The generator A of a symmetric random walk

Consider a real function $a(\cdot) \in l^1(\mathbf{Z}^d)$. Then

$$\sum_z |a(z-z')| = \sum_{z'} |a(z-z')| = c,$$

where $c = \|a\|_{l^1} := \sum_{z\in\mathbf{Z}^d} |a(z)|$. By Schur's Lemma, see, e.g., Shubin[8] the expression

$$(Au)(z) := \sum_{z'\in\mathbf{Z}^d} a(z-z')u(z') \tag{2}$$

determines, for every $p \in [1,\infty]$, a linear bounded operator

$$A : l^p(\mathbf{Z}^d) \to l^p(\mathbf{Z}^d),$$

for which

$$\|A\|_{l^p} \le c.$$

Suppose that

A1: $a(z) \ge 0$ if $z \ne 0$ and $a(0) = -\sum_{z\ne 0} a(z) \le 0$.

Then the operator A can be rewritten in the form

$$(Au)(z) - a(0)u(z) = \sum_{z'\in\mathbf{Z}^d, z'\ne z} a(z-z')u(z'),$$

and, using $\sum_{z\ne z'} |a(z-z')| = \sum_{z'\ne z} |a(z-z')| = |a(0)|$, we get by Schur's Lemma, see Shubin[8], the estimation

$$\|A - a(0)I\|_p \le |a(0)|, \quad \forall\ p \in [1,\infty]. \tag{3}$$

Hence, for the spectrum $\sigma_p(A)$ of the operator $A : l^p(\mathbf{Z}^d) \to l^p(\mathbf{Z}^d)$ we have

$$\sigma_p(A) \subseteq \{z \in \mathbf{C} : |z - a(0)| \le |a(0)|\}, \quad p \in [1,\infty]. \tag{4}$$

Moreover, if we assume that

A2: $a(z) = a(-z)$ for every $z \in \mathbf{Z}^d$,

then the kernel of the operator A becomes symmetric. Then the operator A is self-adjoint in the Hilbert space $l^2(\mathbf{Z}^d)$, so its spectrum lies on the real line. Together with (4) it means that

$$\sigma_2(A) \subseteq [2a(0), 0]. \tag{5}$$

Now we are ready to formulate

Theorem 1. *Let $a(\cdot) \in l^1(\mathbf{Z}^d)$ be a real function such that A1 is valid. Then expression (2) determines, for every $p \in [1, \infty]$, a linear bounded operator $A : l^p(\mathbf{Z}^d) \to l^p(\mathbf{Z}^d)$ satisfying (3) and (4). Moreover, if A2 is valid, then A is a self-adjoined operator in the Hilbert space $l^2(\mathbf{Z}^d)$ and its spectrum $\sigma_2(A)$ satisfies (5).*

It is convenient to formulate some additional properties of the operator A in terms of the Fourier transform of a function $a(\cdot)$. Let the function $\phi(\theta)$ be as follows:

$$\phi(\theta) = \sum_{z \in \mathbf{Z}^d} a(z) e^{i\langle z, \theta \rangle}, \quad \theta \in [-\pi, \pi]^d, \tag{6}$$

where $\langle \cdot, \cdot \rangle$ is a standard scalar product on $\mathbf{Z}^d$.

By A2 the function $\phi(\cdot)$ can be represented in the form

$$\phi(\theta) = \sum_{z \in \mathbf{Z}^d} a(z) \cos\langle z, \theta \rangle, \quad \theta \in [-\pi, \pi]^d,$$

and thus it is a real-valued function. By A1 we have

$$\phi(\theta) = \sum_{z \in \mathbf{Z}^d, z \neq 0} a(z) \left(\cos\langle z, \theta \rangle - 1\right) \leq 0, \quad \theta \in [-\pi, \pi]^d. \tag{7}$$

Since clearly $\phi(0) = 0$ then $z = 0$ is a local maximum of the function $\phi(\cdot)$.

Definition 1. If for every $z \in \mathbf{Z}^d$ there exist a set of vectors $z_1, z_2, \ldots, z_k \in \mathbf{Z}^d$, such that $z = \sum_{i=1}^k z_i$ and $a(z_i) \neq 0$, $i = 1, 2, \ldots, k$, then the function $a(\cdot)$ is called *irreducible*.

Now, we impose one more restriction on the function $a(\cdot)$.

A3: the function $a(\cdot)$ is irreducible.

We consider a random walk X_t, $t \geq 0$, where X_t is the location of a particle on $\mathbf{Z}^d$ at time t. The generator A of a random walk X_t is defined by (2) and satisfies A1–A3. The probability $p(h, x, y)$ that the particle staying at the point $x \neq 0$ in a small time h jumps to an arbitrary point y obeys the equalities

$$p(h, x, y) = a(x, y)h + o(h) = a(y - x)h + o(h) \quad \text{for} \quad y \neq x,$$
$$p(h, x, x) = 1 + a(x, x)h + o(h) = 1 + a(0)h + o(h).$$

As shown, for example, in Yarovaya[10], the random walk transition probability $p(t,x,y)$ satisfies the system of Kolmogorov's backward equations

$$\partial_t p(t,x,y) = Ap(t,x,y), \quad p(t,x,y) = \delta_y(x). \tag{8}$$

Put

$$G_\lambda(x,y) := \int_0^\infty e^{-\lambda t} p(t,x,y)dt, \quad \lambda > 0. \tag{9}$$

By A1 (see, e.g., Yarovaya[10]) the series $\sum_{x\in\mathbf{Z}^d} p(t,x,y)$ is absolutely convergent for every t and y. Hence we can apply to equation (8) the discrete Fourier transform

$$\tilde{p}(t,\theta,y) = \mathbf{E}_x e^{i(X_t,\theta)} = \sum_{x\in\mathbf{Z}^d} p(t,x,y)e^{i(x,\theta)},$$

where $\mathbf{E}_x$ stands for the expectation under the condition that $X_0 = \delta_x(X_0)$. In this case equation (8) takes the form

$$\partial_t \tilde{p}(t,\theta,y) = \phi(\theta)\tilde{p}(t,\theta,y), \quad \tilde{p}(0,\theta,y) = e^{i(\theta,y)}.$$

Hence, $\tilde{p}(t,\theta,y) = e^{\phi(\theta)t}e^{i(\theta,y)}$, and application of the Fourier inversion to the last equality for $t \geq 0$ yields

$$p(t,x,y) = \frac{1}{(2\pi)^d} \int\limits_{[-\pi,\pi]^d} e^{\phi(\theta)t + i(\theta,y-x)}\, d\theta, \quad x,y \in \mathbf{Z}^d.$$

From this we derive another representation for the Green function (9):

$$G_\lambda(x,y) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \frac{e^{i(\theta,\, y-x)}}{\lambda - \phi(\theta)}\, d\theta, \quad x,y \in \mathbf{Z}^d, \quad \lambda > 0. \tag{10}$$

Definition 2. A random walk X_t, $t \geq 0$, will be called transient if the Green function $G_0 < \infty$ and recurrent if $G_0 = \infty$.

3 Symmetric random walks with infinite variance of jumps

Given a continuous positive function $H : [-\pi,\pi]^d \to \mathbf{R}$ satisfying

$$H(x) = H(-x), \quad \forall x \in [-\pi,\pi]^d, \tag{11}$$

denote by H_0 and H^0 the lower and upper bounds for $H(x)$:

$$0 < H_0 \leq H(x) \leq H^0, \quad \forall x \in [-\pi,\pi]^d. \tag{12}$$

Let us suppose now that the irreducible function $a(\cdot)$ satisfies an extra condition (in Yarovaya[12] this assumption is denoted as A5):

A4′: $a(z) = \frac{H\left(\frac{z}{|z|}\right)}{|z|^{d+\alpha}}$, where $\alpha \in (0,2)$, $z \in \mathbf{Z}^d$ and $z \neq 0$.

As well-known (see, e.g., Gradshteyn and Ryzhik[6, Problem 4.642]),

$$\sum_{z\in\mathbf{Z}^d, z\neq 0} \frac{1}{|z|^\nu} < \infty \Longleftrightarrow \nu > d.$$

Hence by (12)

$$\sum_{z\in\mathbf{Z}^d, z\neq 0} a(z) < \infty \Longleftrightarrow \alpha > 0.$$

Now, if we define the value $a(0)$ as

$$a(0) = -\sum_{z\in\mathbf{Z}^d, z\neq 0} a(z) \leq 0,$$

then the condition A1 will be valid for $a(\cdot)$.

The validity of the condition A2 for $a(\cdot)$ follows from (11). At the same time the assumption (1) does not hold for the function $a(\cdot)$ when $0 < \alpha < 2$. So, we have proved the following

Lemma 1. *If an irreducible function $a(z)$ satisfies A4′, then A1 and A2 hold, and*

$$\sum_{z\in\mathbf{Z}^d, z\neq 0} |z|^2 a(z) = \infty.$$

Let us denote by $\phi(\theta)$ the Fourier transform of the function $a(\cdot)$. In virtue of A1, A2 and A4′ the function $a(\cdot)$ is absolutely summable on $\mathbf{Z}^d$. Since the functions $e^{i(z,\theta)}$ are uniformly bounded for all real vectors z and quantities θ the function $\phi(\theta)$ is defined and continuous on the d-dimensional cube $[-\pi,\pi]^d$. We have thus proved the following

Lemma 2. *If A4′ holds for a scalar function $a(z)$, then its Fourier transform $\phi(\theta)$ is correctly defined, continuous on the d-dimensional cube $[-\pi,\pi]^d$ and takes real values. Moreover, $\phi(\theta)$ is of the form (7).*

In particular, when $a(z)$ satisfies A4′ we get

$$\phi(\theta) = \sum_{z\in\mathbf{Z}^d, z\neq 0} H\left(\frac{z}{|z|}\right) |z|^{-(d+\alpha)} \left(\cos\langle z,\theta\rangle - 1\right), \quad \theta \in [-\pi,\pi]^d. \tag{13}$$

Theorem 2. *If $\alpha \in (0,2)$ and A4′ holds for an irreducible function $a(\cdot)$, then*

$$\phi(\theta) \leq -C|\theta|^\alpha, \quad \theta \in [-\pi,\pi]^d,$$

with a constant $C > 0$.

The proof of Theorem 2 is based on the representation (13) and the following lemma.

Lemma 3. *If $d = 1, 2$ and $\alpha \in (0, 2)$ then for the functions*

$$S_{1,\alpha}(\theta) = \sum_{n\geq 1} \frac{1}{n^{1+\alpha}} \left(1 - \cos n\theta\right), \quad S_{2,\alpha}(\theta) = \sum_{n\geq 1} \frac{1}{n^{2+\alpha}} \sin n\theta,$$

we have

$$S_{1,\alpha}(\theta) \sim \frac{1}{\alpha} \Gamma(1-\alpha) \cos\left(\frac{1}{2}\pi\alpha\right) \theta^{\alpha}, \tag{14}$$

$$S_{2,\alpha}(\theta) \sim \theta \sum_{n\geq 1} \frac{1}{n^{1+\alpha}} - \frac{1}{\alpha(\alpha+1)} \Gamma(1-\alpha) \cos\left(\frac{1}{2}\pi\alpha\right) \theta^{\alpha+1}, \tag{15}$$

as $\theta \to 0$.

Let us remark that the value $\Gamma(1-\alpha)\cos\left(\frac{1}{2}\pi\alpha\right)$ in (14) and (15) is formally not defined when $\alpha = 1$. For $\alpha = 1$ it should be understood as the limit: $\lim_{\alpha\to 1-} \Gamma(1-\alpha)\cos\left(\frac{1}{2}\pi\alpha\right) = \frac{\pi}{2}$.

Proof. Put

$$S_{\alpha}(\theta) = \sum_{n\geq 1} \frac{1}{n^{\alpha}} \sin n\theta. \tag{16}$$

As is known, see, e.g. Zygmund[13, Vol. I, Ch. II, Sec.13], for $0 < \alpha < 2$ the function $S_{\alpha}(\theta)$ is defined on the interval $0 < \theta < \pi$ and satisfies

$$S_{\alpha}(\theta) \sim \Gamma(1-\alpha) \cos\left(\frac{1}{2}\pi\alpha\right) \theta^{\alpha-1}, \quad \theta \to 0+. \tag{17}$$

Since $S_{1,\alpha}(0) = 0$ and $S'_{1,\alpha}(\theta) = S_{\alpha}(\theta)$ then by integrating (17) over $[0, \theta]$ we get (14). Taking into account that $S_{2,\alpha}(0) = 0$, $S'_{2,\alpha}(0) = \sum_{n\geq 1} \frac{1}{n^{1+\alpha}}$, $S''_{2,\alpha}(\theta) = -S_{\alpha}(\theta)$ by integrating (17) over $[0, \theta]$ we have (15).

Let us note that by (17) the function $S_{\alpha}(\theta)$ is absolutely summable. Then the trigonometric series in (16) is a Fourier series, see Bari[2, Ch. X, Sec. 2], and hence it is term-by-term integrable. The lemma is proved.

Proof (The proof of Theorem 2). In the case $d = 1$ we need to estimate behavior for small θ of the function $\phi(\theta) = (H(1) + H(-1))\, f(\theta)$, where

$$f(\theta) = \sum_{n\geq 1} \frac{1}{n^{1+\alpha}} \left(\cos n\theta - 1\right).$$

The required estimation for the function $\phi(\theta)$ by Lemma 3 follows from (12), the equality $f(\theta) = -S_{1,\alpha}(\theta)$ and (14), see details in Yarovaya[12].

Now, consider the case $d = 2$. Denote by $\|z\| = \max\{|z_1|, |z_2|\}$ the norm of vector $z = (z_1, z_2)$. Then, by using obvious inequalities $\|z\| \leq |z| \leq 2\|z\|$ where $|\cdot|$ denotes the Euclidean norm, from (13) and (12) we obtain

$$H^0 f(\theta) \leq \phi(\theta) \leq 2^{-(2+\alpha)} H_0 f(\theta),$$

with

$$f(\theta) = \sum_{z\in\mathbf{Z}^2, z\neq 0} \|z\|^{-(2+\alpha)} \left(\cos\langle z,\theta\rangle - 1\right).$$

Thus it remains to estimate the function $f(\theta)$ for small θ. To do it, in what follows it would be convenient to represent $f(\theta)$ in the form

$$f(\theta) = f(\theta_1,\theta_2) := \sum_{(n,m)\in\mathbf{Z}^2,(n,m)\neq 0} \frac{1}{\|(n,m)\|^{2+\alpha}} \left(\cos(n\theta_1 + m\theta_2) - 1\right).$$

Let us notice that the function $f(\theta_1,\theta_2)$ is symmetric with respect to the axes of coordinates and the diagonals in $\mathbf{R}^2$: $f(\theta_1,\theta_2) \equiv f(\pm\theta_1,\pm\theta_2)$ and $f(\theta_1,\theta_2) \equiv f(\pm\theta_2,\pm\theta_1)$. Therefore without loss of generality it suffices to consider the case $\theta_1 \geq \theta_2 \geq 0$.

Let us represent the lattice $\mathbf{Z}^2$ as the union of the squares

$$S_k := \left\{(n,m)\in\mathbf{Z}^2 :\ \|(n,m)\| = k\right\}, \quad k = 0,1,\ldots .$$

Then the function $f(\theta_1,\theta_2)$ will take the form

$$f(\theta_1,\theta_2) = \sum_{k\geq 1} \frac{1}{k^{2+\alpha}} \sum_{(n,m)\in S_k} \left(\cos(n\theta_1 + m\theta_2) - 1\right). \tag{18}$$

Let us calculate for any k the inner sum in (18):

$$f_k(\theta_1,\theta_2) = \sum_{(n,m)\in S_k} \left(\cos(n\theta_1 + m\theta_2) - 1\right).$$

For it we introduce the auxiliary function $\tilde{f}_k(t,s) = \sum_{j=-k}^{k-1} \cos(kt + js)$ and note that

$$f_k(\theta_1,\theta_2) = \tilde{f}_k(\theta_1,\theta_2) + \tilde{f}_k(\theta_2,-\theta_1) + \tilde{f}_k(-\theta_1,-\theta_2) + \tilde{f}_k(-\theta_2,\theta_1) - 8k, \tag{19}$$

where every term $\tilde{f}_k$ corresponds to the summation of the quantities $\cos(n\theta_1 + m\theta_2)$ over one of the four "sides" $S_{k,1}$, $S_{k,2}$, $S_{k,3}$ or $S_{k,4}$ of the square S_k (see Fig. 1).

Now, using the representation

$$\tilde{f}_k(t,s) = \cot\frac{s}{2}\sin(ks)\cos(kt) + \sin(ks)\sin(kt),$$

from (19) we get the equality

$$f_k(\theta_1,\theta_2) = 2\cot\frac{\theta_1}{2}\sin(k\theta_1)\cos(k\theta_2) + 2\cot\frac{\theta_2}{2}\sin(k\theta_2)\cos(k\theta_1) - 8k$$

which is equivalent to

$$f_k(\theta_1,\theta_2) = \left(\cot\frac{\theta_1}{2} + \cot\frac{\theta_2}{2}\right)\sin k(\theta_1+\theta_2) + \left(\cot\frac{\theta_1}{2} - \cot\frac{\theta_2}{2}\right)\sin k(\theta_1-\theta_2) - 8k.$$

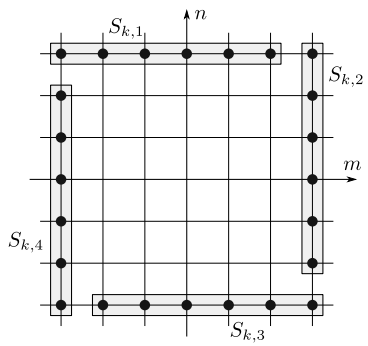


Fig. 1. Partition of the square S_k on the sets $S_{k,1}$, $S_{k,2}$, $S_{k,3}$, $S_{k,4}$.

Hence by (18) we get

$$f(\theta_1,\theta_2)=\left(\cot\frac{\theta_1}{2}+\cot\frac{\theta_2}{2}\right)\sum_{k\geq 1}\frac{1}{k^{2+\alpha}}\sin k(\theta_1+\theta_2)+$$
$$+\left(\cot\frac{\theta_1}{2}-\cot\frac{\theta_2}{2}\right)\sum_{k\geq 1}\frac{1}{k^{2+\alpha}}\sin k(\theta_1-\theta_2)-8\sum_{k\geq 1}\frac{1}{k^{1+\alpha}}.$$

Then in virtue of the asymptotic equality (15) from Lemma 3 we obtain

$$f(\theta_1,\theta_2)\sim\left(\cot\frac{\theta_1}{2}+\cot\frac{\theta_2}{2}\right)\left(N_\alpha(\theta_1+\theta_2)-M_\alpha(\theta_1+\theta_2)^{\alpha+1}\right)+$$
$$+\left(\cot\frac{\theta_1}{2}-\cot\frac{\theta_2}{2}\right)\left(N_\alpha(\theta_1-\theta_2)-M_\alpha(\theta_1-\theta_2)^{\alpha+1}\right)-8N_\alpha,\quad(20)$$

as $|\theta|\to 0$, where

$$N_\alpha=\sum_{n\geq 1}\frac{1}{n^{1+\alpha}},\quad M_\alpha=\frac{1}{\alpha(\alpha+1)}\Gamma(1-\alpha)\cos\left(\frac{1}{2}\pi\alpha\right).$$

Let us rewrite (20) as follows:

$$f(\theta_1,\theta_2)\sim 2N_\alpha\left(\theta_1\cot\frac{\theta_1}{2}+\theta_2\cot\frac{\theta_2}{2}-4\right)-M_\alpha F_\alpha(\theta_1,\theta_2),\qquad(21)$$

where

$$F_\alpha(\theta_1,\theta_2):=\cot\frac{\theta_1}{2}\left((\theta_1+\theta_2)^{\alpha+1}+(\theta_1-\theta_2)^{\alpha+1}\right)+$$
$$+\cot\frac{\theta_2}{2}\left((\theta_1+\theta_2)^{\alpha+1}-(\theta_1-\theta_2)^{\alpha+1}\right).$$

Since $x\cot x = 1-\frac{1}{3}x^2+o(x^2)$ then the first term in (21) is asymptotically equal to $-\frac{1}{6}|\theta|^2$ for $\theta \to 0$. We have assumed that $\theta_1 \geq \theta_2 \geq 0$. Hence for $F_\alpha(\theta_1,\theta_2)$ the following estimations are valid:

$$F_\alpha(\theta_1,\theta_2) \geq \cot\frac{\theta_1}{2}\left((\theta_1+\theta_2)^{\alpha+1}+(\theta_1-\theta_2)^{\alpha+1}\right) \geq$$
$$\geq \cot\frac{\theta_1}{2}\cdot(\theta_1+\theta_2)^{\alpha+1} \geq \cot\frac{\theta_1}{2}\cdot\theta_1^{\alpha+1} \gtrsim 2\theta_1^\alpha \geq \sqrt{2}|\theta|^\alpha.$$

From (21) and the obtained estimations it follows, for $0<\alpha<2$, that

$$f(\theta) = f(\theta_1,\theta_2) \leq -\sqrt{2}M_\alpha|\theta|^\alpha \qquad \theta \to 0.$$

The theorem is proved for the case $d=2$.

Remark 1. The method of proving Theorem 2 allows to estimate the rate of growth of the function $\phi(\theta)$ in the case $d \geq 3$. However, the evaluation of the related trigonometric sums in this case becomes more complicated. To the best of the author's knowledge this approach in not implemented by anyone. Besides, in dimensions $d \geq 3$ repudiation of assumption about finiteness of variance of jumps brings no new phenomena in BRW.

Theorem 3. *Let $X=(X_t)_{t\geq 0}$ be a random walk on $\mathbf{Z}^d$. If A_4' holds for an irreducible function $a(\cdot)$, then the transition probabilities $p(t;x,y)$ satisfy (8) with the generator A. Moreover, the random walk X is transient for $d=1$ and any $0<\alpha<1$, and for $d\geq 2$ and any $1\leq\alpha<2$.*

Proof. Put $p(t)=p(t;0,0)$ and $G_0=G_0(0,0)$. The proof is based on Theorem 2 and the study for $\lambda=0$ of the convergence of the integral (10):

$$G_0 = \int_0^\infty p(t)\,dt = \frac{1}{(2\pi)^d}\int\limits_{[-\pi,\pi]^d}\frac{d\theta}{(-\phi(\theta))} \leq \frac{1}{(2\pi)^d}\int\limits_{[-\pi,\pi]^d}\frac{d\theta}{C|\theta|^\alpha}, \quad C>0.$$

As is known, the latter integral is convergent for $\alpha < d$ from which the assertion of Theorem 3 follows.

4 Branching random walks with heavy tails

Let us consider a branching random walk with a generator A. We assume that the branching takes place only at the origin and is determined by the infinitesimal generating function $f(u) := \sum_{n=0}^\infty b_n u^n$, where $b_n \geq 0$ for $n \neq 1$, $b_1 < 0$ and $\sum_n b_n = 0$. We assume also that $\beta_r = f^{(r)}(1) < \infty$, $r \in \mathbf{N}$, $\beta = \beta_1 = f'(1)$, where $f^{(r)}$ stands for the rth derivative of f. Therefore, if $\mu_t(0) > 0$ particles were at time t at the origin, then, independently of the rest of particles, each particle in the time interval $[t, t+h)$ can either jump

with the probability $p(h,0,y) = a(y)h + o(h)$ to a point $y \neq 0$, or produce n particles ($n \neq 1$) including itself, or die (the case $n = 0$) with the probability $p_*(h,n) = b_n h + o(h)$, or survive (no changes) with the probability $1 - \sum_{y\neq 0} a(y)h - \sum_{n\neq 1} b_n h + o(h)$. By standard argumentation see, e.g., Feller[5], the sojourn time of a particle at the origin is exponentially distributed with the parameter $-(a(0) + b_1)$.

As shown in Albeverio *et al.*[1], Bogachev and Yarovaya[3] and Yarovaya[10], the number of particles at an arbitrary lattice point, as well as the total number of particles, grow exponentially if and only if

$$\beta > \beta_c = G_0^{-1}. \tag{22}$$

In this sense, it is natural to call the value β_c *critical*. It turned out that the same result holds for BRW with heavy tails. Let us define the operator

$$H := A + \beta\Delta_0. \tag{23}$$

As was proved in Yarovaya[11], H, as well as A, acts in each of the spaces $l^p(\mathbf{Z}^d)$, $1 \leq p \leq \infty$. But for the simplicity the next theorem is given only for the case $p = 2$.

Theorem 4. *Let the operator H be defined by (23), where the generator A satisfies A1–A3 and A4′. Then the spectrum of $H : l^2(\mathbf{Z}^d) \to l^2(\mathbf{Z}^d)$ belongs to the real line and all its positive points (provided that such points exist) are isolated eigenvalues.*

Proof. Recall first some information from the spectral theory. Given a linear bounded self-adjoint operator in a Hilbert space its spectrum is not confined just to its eigenvalues. Part of the spectrum of this operator left after removal of the isolated eigenvalus of finite multiplicity is called the *essential spectrum* see, e.g., Kato[7]. As was proved by H. Weyl, the essential spectrum of operator is "stable" with respect to compact perturbations.

The statement of the theorem now follows from Theorem 1 obtained in Sec. 2 and Weyl's criterion for the essential spectrum Kato[7]. Since A is a self-adjoint operator and $\beta\Delta_0$ is a compact self-adjoint operator then by Weyl's criterion the essential spectra of the operators H and A coincide. Theorem 4 is proved.

5 Conclusions

As was demonstrated in Yarovaya[10], Yarovaya[11] and some other works, there is variety of results about exponential growth of the numbers of particles for BRWs under the assumption (1) which depend on the existence of positive isolated eigenvalues of the operator H. If the operator H is defined by (23) with A satisfying A4′ then under the condition (22) H has a unique positive eigenvalue, see Yarovaya[12]. In this situation we can conjecture that

for supercritical BRWs with heavy tails similar theorems about exponential growth of the numbers of particles for supercritical BRWs with the finite variance of jumps might take place.

The research is supported by RFBR grant 13-01-00653a.

References

1. Albeverio, S., Bogachev, L., and Yarovaya, E., "Asymptotics of branching symmetric random walk on the lattice with a single source", *Comptes Rendus de l'Academie des Sciences, Paris, series I.*, 326, 975–980 (1998).
2. Bari, N.K., *Trigonometricheskie rjady*, Fizmatgiz, Moscow, (1961).
3. Bogachev, L., and Yarovaya, E., "Moment analysis of a branching random walk on a lattice with a single source", *Doklady Akademii Nauk*, 363, 439–442 (1998).
4. Borovkov, A., and Borovkov, K., *Asymptotic Analysis of Random Walks. Heavy-Tailed Distributions*, Cambridge University Press, Cambridge (2008).
5. Feller, W., *An Introduction to Probability Theory and Its Applications. Vol. 2*, 2nd ed., Wiley, New York (1971).
6. Gradshteyn, I., and Ryzhik, I., *Table of Integrals, Series, and Products*, 5th ed., Academic Press, Inc. (1994).
7. Kato, T., *Perturbation Theory for Linear Operators*, 2nd ed., Springer-Verlag, Berlin, Heidelberg, New York (1980).
8. Shubin, M., "Pseudodifference operators and their Green function",*Izv. Akad. Nauk SSSR Ser. Mat.*, 49, 652–671 (1985).
9. Spitzer, F., *Principles of Random Walk*, 2nd ed., Springer-Verlag, New York (1976).
10. Yarovaya, E., *Branching Random Walks in Nonhomogeneous Medium*, Moscow State University Press, Moscow (2007).
11. Yarovaya, E., "Supercritical Branching Random Walks with a Single Source", *Communications in Statistics — Theory and Methods*, 40:16, 2926–2945 (2011).
12. Yarovaya E., "Branching Random Walks with Heavy Tails" *Communications in Statistics — Theory and Methods*, 42:16, 2301–2310 (2013).
13. Zygmund, A., *Trygomometric Series. Vol. I*, 2nd ed., Cambridge University Press, Cambridge (1959).

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Assessing a Weibull Approximation to a Competing Risks Model With Two Independent Weibull Distributed Components

Sergio Yáñez[1], Luis A. Escobar[2], and Nelfi González[3]

[1] Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín, Medellín, Colombia (E-mail: syanez@unal.edu.co)
[2] Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA 70803. (E-mail: luis@lsu.edu)
[3] Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín, Medellín, Colombia (E-mail: ngonzale@unal.edu.co)

Abstract. In different application areas, e.g., medicine, biological sciences, and engineering reliability, competing risks models are used to model life of systems or processes with multiple failure modes. The observed failure is the minimum of the possible individual failure times. We discuss the distribution of the minimum for a competing risks model with two independent failure modes and Weibull distributed. We develop criteria and guidelines to assess a Weibull approximation to the distribution of the minimum. We present a relation between a Kullback-Leibler procedure to obtain the best Weibull approximation to the distribution of the minimum and the maximum likelihood Weibull fit ignoring the mode of failure information. The study considers this relationship in small and large samples.
Keywords: Competing risks, Kullback-Leibler information, Ignoring mode of failure, Weibull distribution, Multiple failure modes, Maximum likelihood.

1 Introduction

Competing risks models are used to model life systems or processes with multiple failure modes. In the literature such models are also called multi-risk, compound, or series-system models. Each risk is like a component in a series system, when one component fails, the system fails. This is, the observed failure is the minimum of the possible individual failure times.

Crowder [3,4] show the actual importance of the subject and its broad scope of application areas, e.g., medicine, biological sciences, and engineering reliability. Pintilie [15] gives many medicine applications. In reliability, Lawless [8], Kalbfleisch and Prentice [6] devoted chapters to study the topic. Meeker and Escobar [10] remark about the practical uses of the subject. Seal [16] and David and Moeschberger [5] give details about the history of competing risks.

Nelson [13, chapter 7] describes situations with three specific causes of failure. Bedford [1] reviews applications in maintenance area. Bunea and Mazzuchi [2] relate accelerated lifetest (ALT) with competing risks and copulas. They also mentioned Bayesian ALT under competing risks. Meeker et al. [11] deal with the topic in a warranty data and ALT context. Suzuki et al. [17] give maximum likelihood properties of a two competing risks model with warranty

data. Meeker [12] suggests that more research is needed in the multiple failure modes area.

Meeker and Escobar [10, chapter 6] present the probability plotting as a technique for assessing how well a distribution describes a data set with emphasis on location-scale and log-location-scale distributions. Pascual and Gast [14] discuss probability plotting for independent competing risks models with right censoring; they illustrate their proposal using lognormal and Weibull distributions.

Here we consider a tool to asses a Weibull approximation to competing risks process with two independent Weibull distributed risks. The development is for complete data (i.e., no censoring) and is based on a relationship between (a) the Kullback-Leibler divergence criterion to obtain the best Weibull approximation to the distribution minimum and (b) maximum likelihood estimation to fit a Weibull to the data ignoring the mode of failure. The results are illustrated in Weibull probability plots and they provide guidelines to help the practitioners. The rest of this paper is organized as follows. Section 2 presents the plotting of the minimum on a Weibull probability paper. Section 3 gives details of the relation between the Kullback-Leibler procedure and the ML estimation to fit a Weibull to the data ignoring the mode of failure. Section 4 discusses the guidelines to help the practitioners. Section 5 contains the conclusions and describes topics for future work.

2 Graph of an independent competing risks model in a Weibull probability plot

In reliability and survival data analysis (see, e.g., Meeker and Escobar [10, chapter 6]) the assessment of distributional adequacy is usually done with probability plots. Here we follow their approach. We present most of our results in Weibull probability plots and make comments that help to interpret those plots. This approach simplifies the application of the results.

The Weibull probability scales are $x = \ln(t), y = \ln[-\ln(1-p)]$ were $t > 0$ and $0 < p < 1$. Now, we consider the plot of the cdf for the independent competing risks model with two independent failure modes which are Weibull distributed. Suppose $T_i \sim \text{WEI}(\eta_i, \beta_i), i = 1, 2$ with T_1 and T_2 independent. Define $T = \min\{T_1, T_2\}$. The Weibull cdf for T_i is

$$F_i(t) = 1 - \exp\left[-\left(\frac{t}{\eta_i}\right)^{\beta_i}\right], t > 0, i = 1, 2,$$

where β_i is the shape parameter and η_i is the scale parameter. Then

$$S(t) = P(T > t) = \exp\left[-\left(\frac{t}{\eta_1}\right)^{\beta_1}\right] \exp\left[-\left(\frac{t}{\eta_2}\right)^{\beta_2}\right]. \quad (1)$$

Result 1 *The $F(t)$ curve on a Weibull probability plot is given by*

$$y(t) = \ln[-\ln(S(t))] = \beta_1 \ln(t) - \beta_1 \ln(\eta_1) + \ln\left[1 + \frac{\eta_1^{\beta_1}}{\eta_2^{\beta_2}} t^{(\beta_2 - \beta_1)}\right]. \quad (2)$$

3 Kullback-Leibler procedure versus the Weibull fit ignoring the mode of failure information

Here an important objective is to characterize a best approximation to the distribution of $T = \min\{T_1, T_2\}$. For this, we will use the Kullback-Leibler procedure. This approach avoids sampling errors since this methodology gives an exact Weibull over the whole range of T.

Alternatively, given a set of independent competing risks simulated data one can obtain the Weibull fit ignoring the cause of failure information. We call this approach, the ignoring mode of failure model (IG). We compare this (IG) with the best Kullback-Leibler Weibull approximation to the true independent competing risk model.

Now we give details of the Kullback-Leibler procedure. Suppose that $F(x)$ is the true distribution for the data. In contrast, let $G(x)$ be an arbitrary specified model. Denote by $f(x)$ and $g(x)$ the pdf corresponding to $F(x)$ and $G(x)$, respectively. The quality of the model $g(x)$, to approximate the true model $f(x)$, is measured by the Kullback-Leibler information (K-L) defined as follows

$$I(f;g) = E_F\left\{\ln\left[\frac{f(X)}{g(X)}\right]\right\} = E_F[\ln f(X)] - E_F[\ln g(X)], \tag{3}$$

where E_F represents the expectation with respect to $F(t)$. Small values of $I(f;g)$ indicate a good approximation. The basic ideas used here are taken from Konishi and Kitagawa [7, chapter 3].

The K-L information has the following properties:

(a) $I(f;g) \geq 0$.
(b) $I(f;g) = 0 \Leftrightarrow f(x) = g(x)$.
(c) $I(f;g) \neq I(g;f)$.

Because of the K-L properties (a) and (b) above, the smaller the quantity of K-L, the closer the model $g(x)$ is to $f(x)$. Property (c) shows that the K-L information is not a proper metric or distance. Note that in (3), the term $E_F[\ln f(X)]$ is a constant that depend solely on the true model $f(x)$. Then to compare different models, it is sufficient to consider the second term on the right-hand side of (3).

In the independent competing risks setting, $F(t)$ is the cdf of $T = \min\{T_1, T_2\}$, where $T_i \sim \text{WEI}(\eta_i, \beta_i), i = 1, 2$ with T_1 and T_2 independent. Thus, $F(t)$ is obtained from equation (1) as $F(t) = 1 - S(t)$. This is the correct cdf for the independent competing risks model (ICR). In a simulation study, $F(t)$ is known. When displayed in a plot, $F(t)$ is identified as the ICR model, see, Figure 1.

In this case $g(t)$ is a univariate Weibull model with unknown parameters (η, β). To find the best Weibull approximation to the ICR using the Kullback-Leibler information, we minimize $I(f;g)$ with respect to (η, β). That is

$$\widetilde{\theta} = (\widetilde{\eta}, \widetilde{\beta}) = \arg\min_{(\eta,\beta)} I(f;g) = \arg\min_{(\eta,\beta)}\{E_F[\ln f(T)] - E_F[\ln g(T)]\}. \tag{4}$$

Specifically, since in the simulation process the parameters of $F(t)$ are known, the minimization in (4) with respect to (η, β) yields the quantities $\widetilde{\theta} = (\widetilde{\eta}, \widetilde{\beta})$ which are the best Weibull approximation $g(t)$ to $f(t)$. Thus, when $I(f;g)$ is near to zero, the ICR is well approximated by a univariate Weibull. The rule is that smaller $I(f;g)$ values are better.

It is of interest to study the goodness of the IG model as an approximation of the ICR model. The comparison depends on sample size and here we consider the case of large sample. To fit the IG model, since T_1 and T_2 are independent Weibull variables, we generate 100,000 pairs (t_{1i}, t_{2i}) and take the minimum of each pair to obtain T, say $t_i = \min\{t_{1i}, t_{2i}\}, i = 1, ..., 100{,}000$. Using $t_1, ..., t_{100,000}$ a Weibull univariate model is fitted using ML to obtain $\widehat{\theta} = (\widehat{\eta}, \widehat{\beta})$. Note that the ICR simulated data used here are exact failures times from T_1 y T_2. Thus the data are pairs (t_i, δ_i) where δ_i is a discrete variable for the mode of failure: $\delta_i = 1$ if the failure comes from T_1 and $\delta_i = 0$ if comes from T_2. We can estimate the marginal cdf of $T_i, i = 1, 2$. For T_1, the data are (t_{1i}, δ_i) with $\delta_i = 1$ if $t_{1i} < t_{2i}$ and $\delta_i = 0$ otherwise.

The expectation $E_F[\ln g(T)]|\widetilde{\theta}]$ [i.e., $E_F[\ln g(T)]$ evaluated at $(\widetilde{\eta}, \widetilde{\beta})$] is compared with the estimator of expected log-likelihood $E_F[\ln g(T)|\widehat{\theta}_n]$ which is $\ell_n(\widehat{\theta}_n)/n$, where $\ell_n(\widehat{\theta}_n)$ is the estimated log-likelihood of the ignoring mode of failure model (IG). The closer to one the ratio between these quantities, the better is the IG model as an estimation of the best Weibull approximation to the ICR.

Now we discuss in detail the relation between a Kullback-Leibler procedure to obtain the best Weibull approximation to the distribution of the minimum and the maximum likelihood Weibull fit ignoring the mode of failure information. The log-likelihood for the univariate Weibull model $g(t|\theta)$ with $\theta = (\eta, \beta)$ is

$$\ell_n(\theta) = \sum_{i}^{n} \ln g(t_i|\theta). \tag{5}$$

The maximum likelihood estimate of θ, say $\widehat{\theta}_n$, is given by

$$\widehat{\theta}_n = (\widehat{\eta}, \widehat{\beta}) = \arg\max_{(\eta,\beta)} \ell_n(\theta), \tag{6}$$

where $\ell_n(\widehat{\theta}_n)$ is the estimated maximum log-likelihood. Because of the large n in our simulation scheme, $\widehat{\theta}_n$ should provide the "best" Weibull description of the data when the failure mode is ignored.

From Konishi and Kitagawa [7, chapter 3, page 36], we can state the following theorem, for our simulation scheme where we assume the parameters of $F(t)$ as known.

Result 2 *Suppose* $T_i \sim \text{WEI}(\eta_i, \beta_i), i = 1, 2$ *with* T_1 *and* T_2 *independent and the data are only exact failures times. Define* $T = \min\{T_1, T_2\}$ *and suppose that* $F(t)$ *is known. Then, when* $n \to \infty$

$$\frac{1}{n}\ell_n(\widehat{\theta}_n) \xrightarrow{p} E_F[\ln g(T)|\widetilde{\theta}]. \tag{7}$$

where p indicates convergence is probability.

Proof.

$$E_F[\ln g(T)] = \int \ln g(T) dF(t).$$

The empirical distribution $\widehat{F}$ is the distribution function for the probability function $\widehat{f}(t_i) = (1/n), i = 1, 2, ..., n$ that assigns probability $1/n$ to each of the n failure times. Thus, we obtain

$$E_{\widehat{F}}[\ln g(T)] = \int \ln g(t) d\widehat{F}(t) = \sum_{i=1}^{n} \widehat{f}(t_i) \ln g(t_i) = \frac{1}{n} \sum_{i=1}^{n} \ln g(t_i).$$

According to the law of large numbers, when $n \to \infty$

$$\frac{1}{n} \sum_{i=1}^{n} \ln g(t_i) \xrightarrow{p} E_F[\ln g(T)].$$

Thus, when $n \to \infty$

$$\frac{1}{n} \ell_n(\theta) \xrightarrow{p} E_F[\ln g(T)].$$

Also, when $n \to \infty$

$$\frac{1}{n} \max_{\theta} \ell_n(\theta) = \frac{1}{n} \ell_n(\widehat{\theta}_n) \xrightarrow{p} \max_{\theta} E_F[\ln g(T)] \tag{8}$$

Because F is known, from (4) and the fact that $E_F[\ln f(X)]$ is a constant, it follows that

$$\max_{\theta} E_F[\ln g(T)] = E_F[\ln g(T)|\widetilde{\theta}]. \tag{9}$$

Therefore, as $n \to \infty$

$$\frac{1}{n} \ell_n(\widehat{\theta}_n) \xrightarrow{p} E_F[\ln g(T)|\widetilde{\theta}].$$

Corollary 1. *Suppose the conditions of Result 2 and $\theta_0 = (\eta_0, \beta_0)$ is the true parameter of the Weibull $g(t|\theta)$. Then, when $n \to \infty$*

$$\widehat{\theta}_n \xrightarrow{p} \widetilde{\theta}. \tag{10}$$

4 Guideline

From Result 2 and Corollary 1 we state the following criterion.

If the independent competing risks model is very well approximated for a Weibull distribution it must be close to a $WEI(\hat{\eta}, \hat{\beta})$, where $(\hat{\eta}, \hat{\beta})$ are the ML estimators of the ignoring mode of failure model (IG) for large samples.

Thus we will study on Weibull probability plots how close should be the ICR model and the IG model to asses the Weibull approximation.

The distribution of the minimum of two independent Weibull random variables with the same shape parameter is also a Weibull with the same shape parameter (see, e.g., Lee [9]). Now we discuss for what values of $|\beta_1 - \beta_2|$, it can be obtained a good Weibull approximation for the ICR model, independent of the values of η_1 and η_2. This is directly related to the closeness of the ICR model and the IG model.

We have done a exhaustive and careful simulation study which can be obtained from the first author. In summary, without loss of generality, we had established that the analysis can be restricted to models with $\eta_1 = \eta_2$ and depending only on $|\beta_1 - \beta_2|$. In that context, using Weibull probability plots, we state guidelines to identify a good Weibull approximation to the ICR model.

To state a rule of thumb of how close the slopes should be for the criterion to hold, we analyze the following Weibull probability plots. From Figure 1 , after the late discussion, we can infer general recommendations. It is seen on Figure 1, plots with $\beta_1 = 2, 2.2, 2.4, 2.5, 2.6$, a very good approximation to a Weibull through the closeness between IG and ICR. This is when $|\beta_1 - \beta_2| \leq 0.6$. While, plots with $\beta_1 = 2.7, 2.9, 3$ show the behavior of the ICR model which is departing from the IG straight line as $|\beta_1 - \beta_2| > 0.6$.

As a conclusion of the above discussion, we recommend the following simple practical rule.

Guideline:

When $|\beta_1 - \beta_2| \leq 0.6$, the independent competing risks model (ICR) is very well approximated by a WEI$(\widehat{\eta}, \widehat{\beta})$, where $(\widehat{\eta}, \widehat{\beta})$ are the ML estimators of the ignoring mode of failure model (IG) for large samples.

5 Conclusions and future work

We have shown that the Kullback-Leibler procedure to obtain the best Weibull approximation to the distribution of the minimum for a competing risks model with two independent failure modes and Weibull distributed is very well approximated by the maximum likelihood Weibull fit ignoring the mode of failure information for large samples. Based upon that, and in the frame of the Weibull probability plot we develop an heuristic guideline which depends only on the slope difference $|\beta_1 - \beta_2|$.

The future work include the following: (a) A theoretical support for the results derived from the simulation design; (b) the behavior of Result 1 and Corollary 1 in small samples; (c) extending the results to a general log-location scale family of distributions.

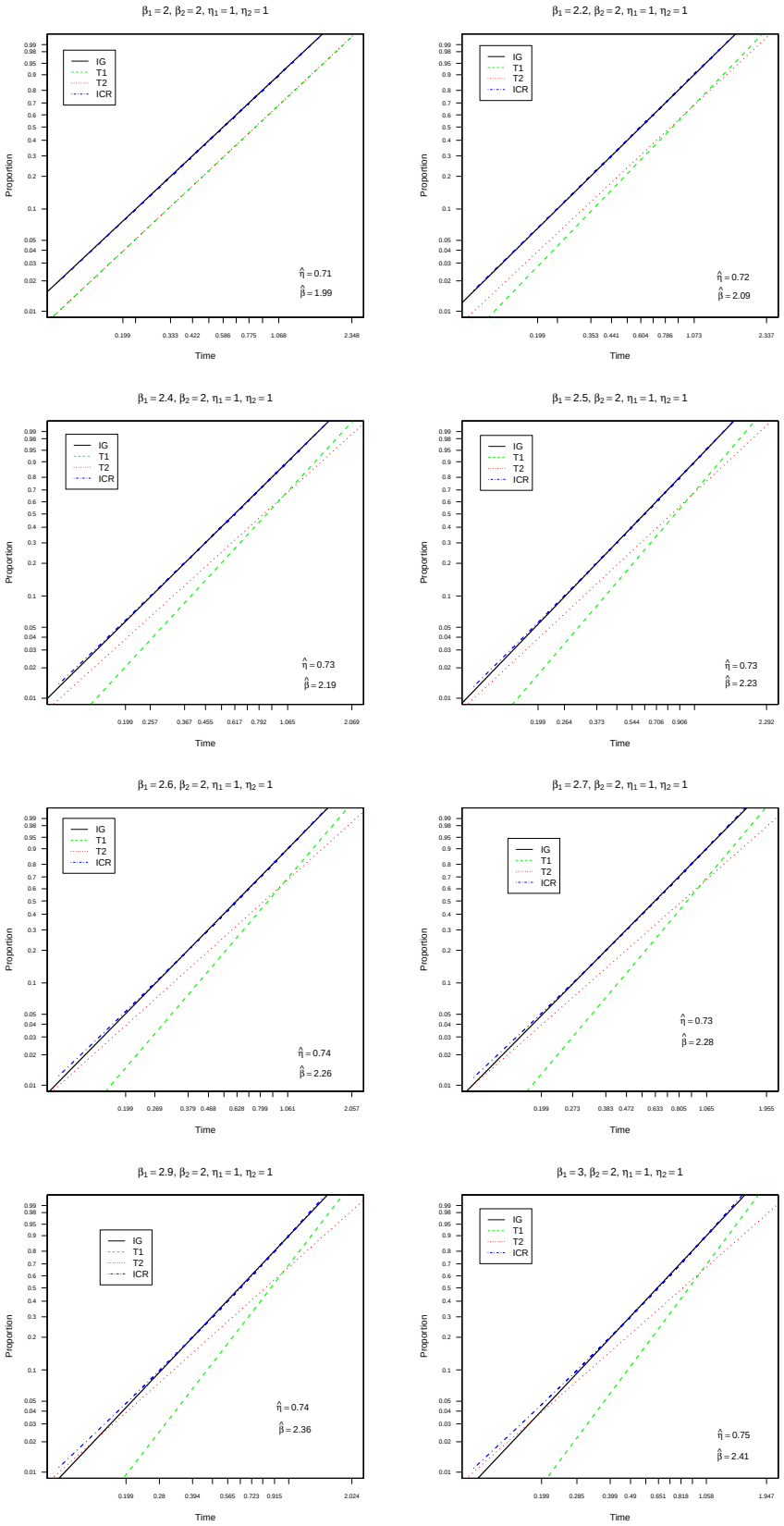


Fig. 1. Weibull probability plots showing the cdf of T_1, the cdf of T_2, the cdf of the ignoring mode of failure model (IG), and the cdf of the true independent competing risks model (ICR) for $\beta_2 = 2$, $\eta_1 = \eta_2 = 1$ and different values of β_1.

6 Acknowledgments

This work was supported by Colciencias, Colombia. This is part of the research project "Characterization of Dependence in Competing Risks Models in Industrial Reliability Data" code 110152128913. This project is directed by Sergio Yáñez, Luis Escobar, and Nelfi González.

References

1. T. Bedford. *Competing Risk Modelling in Reliability. In Moderm Statistical and Mathematical Methods in Reliability*, World ScientificBooks, United Kingdom, pp. 1–16, 2005.
2. C. Bunea and T. A. Mazzuchi. *Sensitivity Analysis of ALT with Competing Failure Modes. In Moderm Statistical and Mathematical Methods in Reliability*, World ScientificBooks, United Kingdom, 2005.
3. M. Crowder. *Classical Competing Risks*, Chapman & Hall/CRC Press, 2001.
4. M. Crowder. *Multivariate Survival Analysis and Competing Risks*, CRC Press, 2012.
5. H. A. David and M. L. Moeschberger. *The Theory of Competing Risks*, Griffin, London, 1978.
6. J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, New York, 2002.
7. S. Konishi and G. Kitagawa. *Information Criteria and Statistical Modelling*, Springer, New York, 2008.
8. J. F. Lawless. *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York, 2003.
9. L. Lee. Multivariate distributions having Weibull properties. *Journal of Multivariate Analysis*, 9, 267–277, 1979.
10. W. Q. Meeker and L. A. Escobar. *Statistical Methods for Reliability Data*,John Wiley & Sons, New York, 1998.
11. W. Q. Meeker, L. A. Escobar, and Y. Hong. Using accelerated life tests results to predict product field reliability, *Technometrics* , 51,146–161, 2009
12. W. Q. Meeker. *Trends in the statistical assessment of reliability. In advances in Degradation Modeling*, pp. 3–16, 2010.
13. W. Nelson. *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*, John Wiley & Sons, New York, 1990.
14. F. G. Pascual and C. Gast. *Probability plotting with independent competing risks. In advances in Degradation Modeling*, pp. 397–413, 2010.
15. M. Pintilie. *Competing Risks. A Practical Perspective*, John Wiley & Sons, 2006.
16. H. K. Seal. Studies in the history of probability and statistics. XXXV. Multiple decrements or competing risks, *Biometrika*, 64, 429–430, 1997.
17. K. Suzuki, W. Yamamoto, T. Hara, and M. Alam. *Properties of lifetime estimator based on warranty data consisting only of failures. In advances in Degradation Modeling*, Birkhäusser, Boston, pp. 39–55, 2010.

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

On the Choice of Runs Rules for efficient Process Monitoring

Babar Zaman[1] and Muhammad Riaz[2,3]

[1]Research Department, King Khalid Eye Specialist Hospital, Riyadh, KSA

[2]Department of Mathematics and Statistics, King Fahad University of Petroleum Minerals, Dhahran 31261, Saudi Arabia.

[3]Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan

Corresponding author's email: riaz76qau@yahoo.com

Abstract

Statistical Process Control (SPC) charts are used to monitor the performance of a process by differentiating natural and unnatural variations. The Shewhart type charts are meant to address larger process shifts in the parameters of the quality characteristic of interest. These charts are not very efficient at detecting shifts of smaller magnitude, for which implementation of the runs rules schemes is an attractive option. In this study we have suggested the use of runs rules schemes for $\overline{X}$,*R* and S charts for Bur distributed process environments with different amounts of skewness and kurtosis (normal distribution is also considered as a special case). We have also investigated the choice of an appropriate runs rule for a given situation to monitor location and dispersion parameters. For performance evaluations and comparisons we have used probability to signal as performance measure. We have also given an application with a real dataset to elaborate and highlight the practical importance of the proposals of the study.

Key words: Average Run Length (ARL); Burr Distribution; Runs Rules Schemes, Shewhart Charts, Skewness and Kurtosis.

1. Introduction:

Statistical Process Control (SPC) is a collection of very powerful tools to control the undesirable variations either in mean or dispersion parameters. The control charts are very effective process monitoring tools introduced by Shewhart in 1920s. The main function of these charts is detection of out-of-control scenarios in the product characteristics. The commonly used charts include $\bar{X}$, R and S charts to monitor the variations in mean and dispersion (cf. Montgomery)[1]. The limitation of these charts is that they are not very efficient for detecting small shifts. There are other types of charts like EWMA and CUSUM charts designed particularly to address small to moderate shifts.

In order to improve the quality of the products for different amounts of shifts runs rules schemes are imposed with Shewhart, EWMA and CUSUM control charts. There is variety of literature available in this

direction. To refer but a few of these: Western Electric Statistical Quality Control Handbook (cf.Western electric Company [2] introduced runs rules for different process patterns; Nelson[3] has suggested eight runs rules for different types of patterns to improve the quality of products; Page[4] used the idea of warning limits; Champ and Woodall[5] made the power comparisons of runs rules for $\bar{X}$ chart by introducing the T (k-m, k, a, b) formula for runs rules, which means that k-m out of k sample statistics fall between a and b limits; Klein[6] adjusted the control limits coefficient for each 2/2 run rule; Koutras et al. [7] and Acosta-Mejia[8] made improvements with two sets of runs rules for $\bar{X}$ chart and k of k run rules for D charts respectively; Antzoulakos and Rakitzis [9,10] modified *r* out of *m* runs rules schemes for monitoring process variability in variance based on the sample standard deviation with one-sided *S* charts; Khoo and Ariffin [11,12] used the runs rules in order to increase the sensitivity of a Shewhart control chart; Riaz et al.[13] used two runs rules schemes for the CUSUM charts and compared with other counterparts.

This research will focus on modified and extended runs rules in order to figure out the efficient runs rule(s) among the set of runs rules for particular shifts for both location and dispersion parameters. We will be dealing with the Burr distributed process behaviors in the form of normal and non-normal data with different combinations of skewness and kurtosis. The organization of the rest of the study is as: Section 2 discusses the runs rules schemes and the methods to find out the efficient runs rules for location and dispersion parameters; Section 3 evaluates the performance of runs rules schemes for Burr distribution with different skewness/kurtosis levels and explains about the choice of a better runs rule scheme and section 4 furnishes the findings and gives some future recommendations.

2. Runs Rules Based Monitoring of Burr Distributed Processes

The industrial environments may encounter different types of data like symmetric, skewed, peaked etc. and the process behaviors may model more appropriately by non-normal distributions. Burr distribution is one such case in industrial applications. In literature may researchers used burr distribution to model the process data for their study investigations. The following is the cumulative distribution function $F_x(x)$ of the Burr Distribution.

$$F_x(x) = \begin{cases} 0 & for\ x < 0 \\ 1 - \dfrac{1}{(1+x^c)^k} & for\ x \geq 0 \end{cases}$$

where c *and* k are the parameters of the distribution.

The skewness and kurtosis will be used as parameters to measure the degree of departure from normality of underlying distribution. Table 1 contains the information of nine types of distributions linked with Burr distribution for our study purposes with the different combination of *c* and *k*. The results are reported in the form of nine cases for nine different distributions in Table 1:

Table.1: Special Cases of Burr Distribution

	Description	Mean	STDV	Skewness	Kurtosis	*c*	*k*
Case 1	Almost uniform Dist.	0.5336	0.3000	0.00	2.000	-18.1484	0.0629
Case 2	Normal Dist.	0.6485	0.2887	0.00	3.000	-11.2519	0.1463
Case 3	Symmetric and peaked Dist.	0.9799	0.0599	0.00	4.0097	27.0689	1.3258
Case 4	Positively skewed Dist.	0.2887	0.3021	1.00	2.9943	-13.068	0.0301
Case 5	Positively skewed and peaked Dist.	0.3750	0.3391	1.00	3.9915	-7.176	0.0787
Case 6	Positively skewed and peaked Dist.	0.5101	0.2654	1.00	5.0262	2.3471	4.4286
Case 7	Highly skewed, highly peaked Dist.	0.1396	0.2402	2.01	6.2233	-21.4163	0.0074
Case 8	Highly skewed, severely peaked Dist.	0.1706	0.2719	2.00	7.1668	-7.5077	0.027
Case 9	Almost exponential Dist.	0.2197	0.3100	2.00	8.9158	-5.8643	0.0442

2.1 Implementation of Runs Rules Schemes

For an efficient monitoring of processes, runs rules are famous choices to be implemented with control charting structures of different quality characteristics of interests. Riaz et al.[14] suggested different runs rules with their independent capacities and checked their performance with some well-known charts for normally processes. They introduced three categories of the runs rules of type $k - m\ out\ of\ k$, given as follows:

$$d_0 = 0, \left(\text{i.e.} 1/1, 2/2, 3/3, 4/4, 5/5, 6/6, 7/7, 8/8, 9/9\right),$$
$$d_1 = 1, \left(\text{i.e.} 1/2, 2/3, 3/4, 4/5, 5/6, 6/7, 7/8, 8/9\right),$$
$$d_2 = 2, \left(\text{i.e. } 1/3, 2/4, 3/5, 4/6, 5/7, 6/8, 7/9\right).$$

They have implemented these runs rules schemes with the design structures of $\bar{X}$, R, and S charts. They have shown through power curve analysis that the efficiency keeps improving with the increase in the value of d_i under normality assumption. This study focuses on the role of d_i for a more generalized class of distribution namely the Burr distribution (for which normal is a special case (cf: table.1)). We define six more categories for d_i s given as:

$d_3 = 3, (\text{i.e.} 1/4, 2/5, 3/6, 4/7, 5/8, 6/9),$
$d_4 = 4, (\text{i.e.} 1/5, 2/6, 3/7, 4/8, 5/9),$
$d_5 = 5, (\text{i.e. } 1/6, 2/7, 3/8, 4/9),$
$d_6 = 6, (\text{i.e. } 1/7, 2/8, 3/9),$
$d_7 = 7, (\text{i.e. } 1/8, 2/9),$
$d_8 = 8, (\text{i.e. } 1/9).$

Let $k - m = r$, the runs rules of r *out of* k *for* any choice of category expression can be written as: $d_{(i=j-l)} = (\eta_l = l)/(k_j = j))$, *where* $i = 0,1,\ldots,8, \eta \& k_j \geq 1,\ l = 1,\ldots,9, j = 1,\ldots 9$ *and* $j \geq l$. The minimum value of l is one and maximum depends on j. It is to be mentioned that the above mentioned are only few runs rules with their selective categories but we can easily extend the whole setup by using more generalized conditions e.g. $l = 1,2,\ldots,n$ *and* $j = 1,2\ldots,m$ *with* $m \geq n$. The categories depend on the choice of n *and* m, for the larger values of n *and* m the number of categories will also increase by imposing the above stated conditions and vice versa. We may generalize the suggested categories using $i = 0,1,\ldots..n_1$. The maximum number of runs rules and their categories are: $d_{(n_1)} = n/m$, where $n_1 = n - m$. We will use these stated runs rules schemes and the relevant categories for our research purposes in order to explore the choice of appropriate schemes in a given situation. We elaborate the setup for some selective options here: for $j = 3,5,6,7$ the possible values of l are $l = 1,2,3,4,5,6,7$. When $j = 3$ and possible values of $l = 1,2,3$ then the possible runs rules are $(1/3, 2/3, 3/3)$ and these belong to (d_2, d_1, d_0) categories. Similarly, when $j = 4,5,6,7$ the possible runs rules are given as: $(1/4, 2/4, 3/4, 4/4)$, $(1/5, 2/5, 3/5, 4/5, 5/5)$, $(1/6, 2/6, 3/6, 4/6, 5/6, 6/6)$, $(1/7, 2/7, 3/7, 4/7, 5/7, 6/7, 7/7)$ along with their respective categories given as: $(d_3, d_2, d_1, d_0), (d_4, d_3, d_2, d_1, d_0), (d_5, d_4, d_3, d_2, d_1, d_0), (d_6, d_5, d_4, d_3, d_2, d_1, d_0)$.

The selection of control limits is very important for every runs rule/category in order to fix the pre-specified false alarm rate α or in- control Average run Length (ARL). Junsub et al. [15] used 3 sigma limits for process monitoring and encountered the problems of fluctuated false alarms due to different skewness and kurtosis levels for Burr distributed process. They used cases # 2, 4 and 6 (cf. Table 1) and did not fixe the ARL_0 for process monitoring and evaluation. We define the appropriate limits for the monitoring of Burr distributed

process using varying runs rules schemes by fixing the pre-fixed pre-specified false alarm rate α. We have mainly followed Riaz et al. [13] which covers the case of normal distribution with runs rules schemes showing superior performance with the increase in the value of d_i. In this study we generalize the idea for Burr distributed processes and explore the role of d_i in appropriately choosing a runs rule scheme for efficient performance of the control charting structures. The description of the runs rules schemes to monitor certain process parameter .are listed below (cf. Table 2):

Table.2: Description of Runs Rules

Upper Side: $T(\eta_l = l, k_j = j, h_{R_l,k}, \infty)$**:** An out-of-control signal is received if at least η_l points of k_j connective points fall above the control line $h_{R,k}$ of the sampling distribution of the control charting statistic.
Lower Side: $T(\eta_l = l, k_j = j, -\infty, h_{R_l,k})$**:** An out-of-control signal is received if at least η_l points of k_j connective points fall below the control line $h_{R,k}$ of the sampling distribution of the control charting statistic**.**
Both Sides: $T(\eta_l = l, k_j = j, h_{R_l,k}, h_{R,k})$**:** An out of control signal is received if at least η_l points of k_j connective points fall below or above the control lines $h_{R,k}$ and $h_{R_l,k}$ **respectively** of the sampling distribution of the charting statistic.
Here $k_j \geq \eta_l$ for all the above-mentioned structures of limits.

The above mentioned limits are defined with a pre specified probability of false alarm rate α defined and evaluated using the following expression (c.f Riaz et al. [24]).

$$\alpha = \sum_{k_j \geq r_l} \frac{k_j!}{(k_j - \eta_l)!\ \eta_l!} p^{r_l} (1-p)^{k_j - r_l}$$

Where p is the probability of a single point η_l *out of* k_j falling outside their respective limits.

3. Performance Evaluations and the Choice of Appropriate Runs Rule Scheme

In this section we evaluate the performance of the runs rule schemes for different variants of Burr distributed processes with varying effects of the skewness and kurtosis. For this purpose we have chosen Shewhart's $\bar{X}$, S and R charts (the others may considered on the similar lines). The shifts are considered in terms of $\mu + \delta\sigma$ for $\bar{X}$ chart and $\gamma^2\sigma^2$ for S and R charts. When $\delta = 0$ *and* $\gamma^2 = 1$ the process is considered to be in-control otherwise out- of-control. For the said of shifts of different amounts, the power computations are carried out

under varying runs rules schemes (given in Table 2) for the different cases of Burr distributed process as mentioned in Table 1. We have considered different choices of n, α, δ, k and r to compute discriminatory powers of $\bar{X}$, S and R charts for different runs rules schemes. The resulting outcomes are presented in the form of figures containing power curves/bar charts. For discussion and comparison purposes, we provide here some selective figures (cf. Figures 1-4) for the sake of brevity. The results advocate the following for different runs schemes (given in Table 2):

Figure.1: *Efficient Runs Rules ($\bar{X}$Chart)*

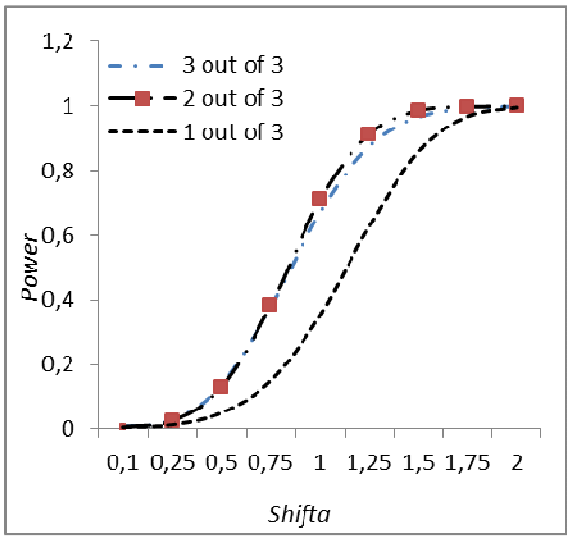


Figure: a2*Runs Rules power performance at different Shifts ($\bar{X}$Chart)*

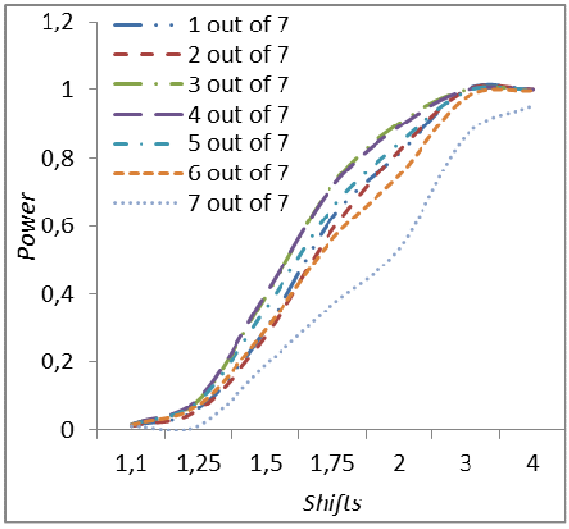
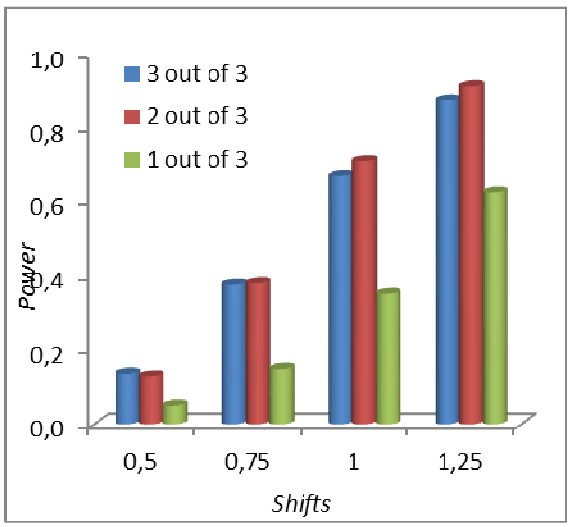


Figure:4: *Runs Rules power performance at different Shifts(R-Chart)*



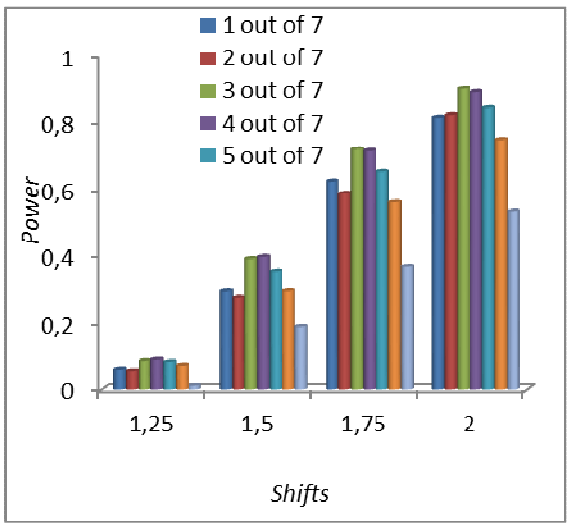


Figure.3: *Efficient Run Rules(R-Chart)*

The above-mentioned findings are based on the well known $\bar{X}$ and R charts for normal distribution which is special case of (table.1) of Burr distributed processes. The similar behaviors may be expected from the other charts as well for different distributional behaviors. In figures 1and 3 in which powers are plotted against to different shifts for location and dispersion parameters for process monitoring. The power curves of $\bar{X}$and R charts showed that the 2/3 and 3/7 run rules is more efficient to detect the shifts for location and dispersion in the comparison of others runs rules. The powers (Figures.1 and 3) at different shifts showed that 2/3 and 3/7 runs rule is also performing more efficient as compared to others Runs Rules for both location and dispersion parameters respectively. The overall power cures and power (figures 2 and 4) at individual shift has the same results to find out the efficient Run Rules for both parameters

R Chart:

We selected few runs rules with few cases; each single Run Rule for the set of different distributions is behaving significant differently for R Chart.

Figure.R1:*(1/1) Run Rule for different Cases(R-Chart*)

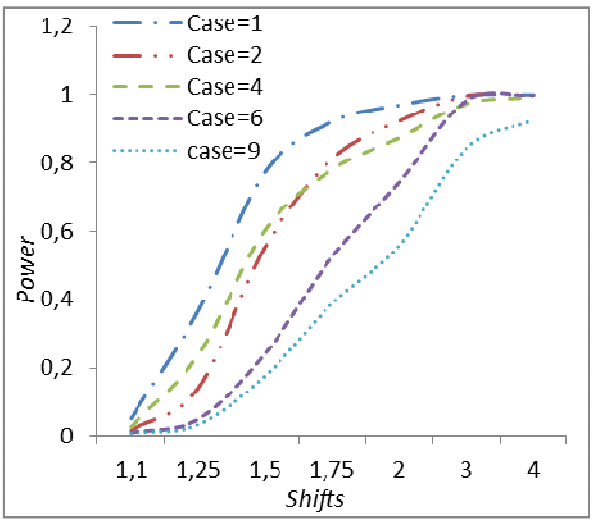
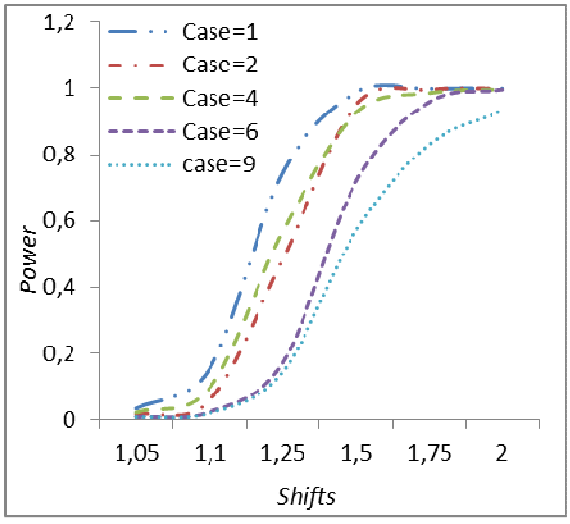


Figure.R4:*(2/4 Runs Rules for different cases (R Chart))*



X-bar Chart:

$\bar{X}$ chart perforance is poor as compare to R chart, beacase $\bar{X}$ chart is used to monitor the location parameter and effected by skweness and in our cases only small deviation in skweness from 0 to 2 that is not sufficient.

Figure.X1:*(1/1) Run Rule for different cases ($\bar{X}$ Chart)*

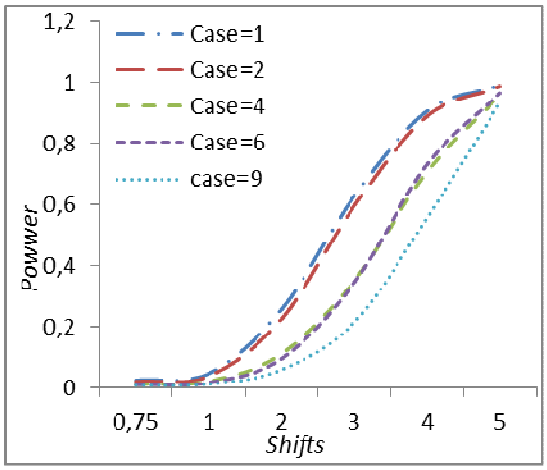
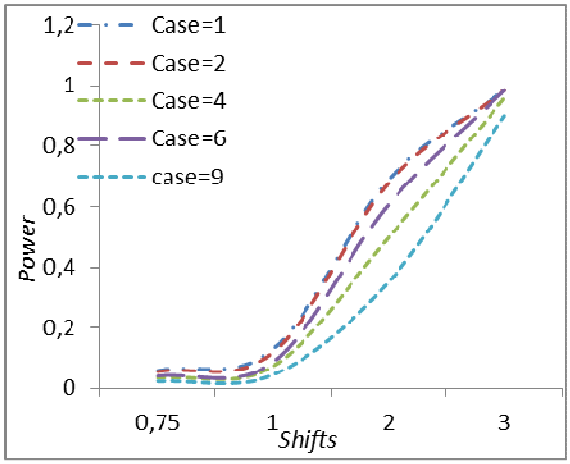


Figure.X3:*(2/4) Runs Rule for different cases($\bar{X}$ Chart)*



S Charts :

Performance of run rules under different Skewness and Kurtosis when Burr function is used.

Figure.S1:*(1/1) Run Rule for different cases(S Chart)*

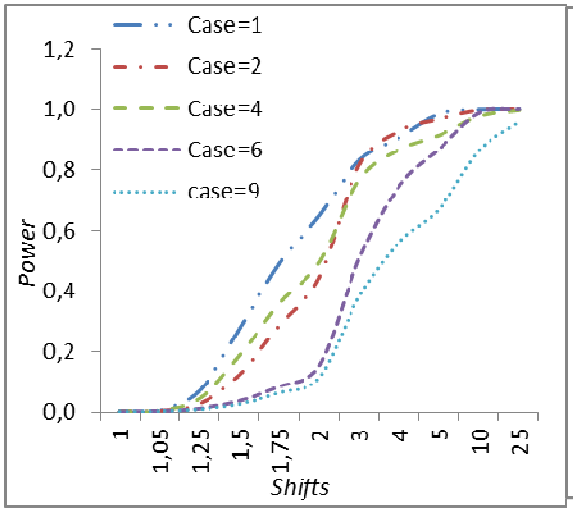
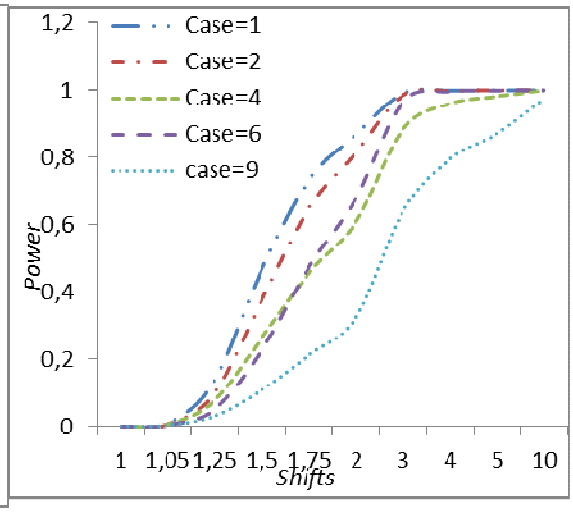


Figure.S3:*5 /6 Run Rule for different cases(S-Chart)*



4 Conclusion and Recomandations.

The Run Rules have been recomanded by the researchers as tool to monitor process behavior for location and dispersion parameters in industrial sector. We did not find any method or tool in literature that help us to find the efficient runs rules for assignable causes under normal process and effect of skewness and kurtosis. Here, we defined the method to find out efficient run rules and found there is need to identify the best one runs rule for location and dispersion process for minimu false alarm rate and also it is not neccesary that same order ot Run Rule is efficient for both parameters at the same time. The Run Rule has different performance when the process is away from normality. $\bar{X}$ and R charts are effected by the Skewness and kurtosis repectively.

References :

[1]Montgomery, D. C (2009). Introduction to Statistical Control.6th ed, Wiley: New York.

[2]Western Electric Company (1956). *Statistical Quality Control Handbook*. Western Electric Company, Indianapolis: IN.

[3]Nelson, L.S (1984). The Shewhart control chart-Test for special causes. *Journal of Quality Technology*; 16:229-245.

[4]Page, E.S (1955) Control charts with warning limits. *Biometrika*; 42:243-257.

[5]Champ, C.W and Woodall, W.H (1987). Exact results for Shewhart control with supplementary runs rules. *Technimetrics*; 29:393-399.

[6]Klein, M (2000). Two Alternatives to the Shewhart X control chart. *Journal of Quality Technology*, 32, 427-431.

[7]Koutras, M.V, Bersimis, S. and Maravelakis, P.E (2007). Statistical process control using Shewhart control chart with supplementary runs rules. Methodology and Computing in Applied Probability; 9:207–224.

[8]Acosta-Mejia, C.A (2007). Two sets of runs rules for the X-bar chart. *Quality Engineering*; 19:129-136.

[9]Antzoulakos, D. L. and Rakitzis, A. C (2008). The modified *r* out of *m*control chart. *Communication in Statistics Simulations and Computations*; 37, 396-408.

[10]Antzoulakos, D. Land Rakitzis, A. C (2010). Runs rules schemes for monitoring process variability. *Journal of Applied Statistics*; 37, 1231–1247.

[11]Khoo, M.B.C (2007). Design of runs rules schemes. *Quality Engineering*; 16, 27-34.

[12]Khoo, M.B.C. and Ariffin K.N (2006). Two improved runs rules for the Shewhart $\bar{X}$ chart. *Quality Engineering*; 18:173-178.

[13]Riaz.M, Abbas, N. and Does, R. J. M. M (2011). Improving the Performance of CUSUM Charts.*Quality and Reliability Engineering International*; 27 (4), 415-424.

[14]Riaz, M, Mahmoud, R and Does, R. J. M. M (2011). On the Performance of Different Control Charting Rules.*Quality and Reliability Engineering International*; 27(8), 1059-1067.

[15]Junsub, Y., Victor, R. P. And Howard R, C (2001). *ARL* Comparisons between Neural Network Models and $\bar{X}$-Control Charts for Quality Characteristics that are non-normally distributed. *Economic Quality Control;* 16, 1, 5 – 15.