

Queuing Systems with Two Service Operations as Mathematical Models of Reliability and Survivability

Revaz Kakubava¹, Tinatin Kaishauri², and Otar Shonia³

- ¹ Department of Applied Mathematics, Georgian Technical University, 77 Kostava str., 0175, Tbilisi, Georgia (E-mail: r.kakubava@gtu.ge)
- ² Management Information Systems, Georgian Technical University, 77 Kostava str., 0175, Tbilisi, Georgia (E-mail: kaishauri@gtu.ge)
- ³ Management Information Systems, Georgian Technical University, 77 Kostava str., 0175, Tbilisi, Georgia (E-mail: o.shonia@yahoo.com)

Abstract: The given paper deals with the redundancy and maintenance problem for a wide class of any territorially distributed standby systems consisting of unreliable repairable elements. Mathematical models for interaction of degradation and its compensation processes in the above mentioned systems are proposed and their possible applications are partially analysed. These models represent mixed type queuing systems for two parallel maintenance operations – replacements and repairs. The problem for optimization of said system by economic criterion is stated. The possible ways of its solution are discussed.

Keywords: queuing models; structural control; maintenance; replacement; renewal (repair).

1 Introduction

During last decades, in reliability theory and practice (as well as in survivability theory and practice), the problems of redundancy, maintainability and supply of large scale systems, including terrestrial ones, are becoming the main directions.

This is strongly attested by leading experts, among them, distinguished scientist Igor Ushakov [1]. Other works of Igor Ushakov himself on the subject, as well as the works of other authors are referenced in [2]. Should also refer to [3, 4].

At the same time, traditional maintenance models in many cases proved to be unsuitable, and there was an urgent need for the construction and investigation of entirely new types of models for the mathematical description of the mentioned technical systems.

These models, as a rule, are distinctive symbiosis of reliability and queuing theories, along with inventory control and other parts of operations research (management science) [1-2].

One of the main reasons determining the above described statement is that in practical cases of redundancy, the main and standby elements, as a rule, were territorially concentrated and failed main element's replacement with the standby one meant the latter's switch, which was often automatically performed and the duration of operation was insignificantly small.

But in modern networks of above mentioned type, standby elements are not directly attached (linked) to main elements. They are located at specific storage

locations and may be tens, hundreds and sometimes thousands of kilometers away from the main elements. Therefore, the delivery duration of stand by elements to the place of the failed main ones is quite substantial. At the same time, in practical cases, due to various reasons, before the beginning of stand by element's delivery operation, passes quite some time, which is often many times greater than the delivery time. Moreover, replacement operation, as a rule, is undertaken not by repair unit, but by special replacement channel. Therefore, replacement of failed main element by stand by one quite naturally becomes an independent maintenance operation.

In addition, the replacement process, apart from the standby element's delivery to the main element's place, includes other sub-operations, whose execution is necessary in order for the standby element to continue main element's functions. In such circumstances, the replacement operation's average duration is not insignificant, and it often reaches 20-40 % of repair operations' average duration.

2 Object of study and its initial mathematical description

The investigation object of this article is a multi-element redundant system with repairable elements.

The system consists of identical m main and n standby elements. Standby elements are designated for permanent replacement of main elements in case of their failure. It is supposed that for the normal operation of the system, the serviceability of all m main elements is desired. However, if their number is less than m, then the system continues to function but with lower economic effectiveness.

The main elements fail with intensity α and the standby ones – with intensity β . A failed main element is replaced by a serviceable standby one if there is available standby element in the system. In the opposite case the replacement will be carried out after standby element's availability. The failed elements, both the main and the standby ones, are repaired and become identical with the new ones.

There are k replacement and l repair units in the system. The durations of replacement and repair operations are random values with distribution functions F(x) and G(x), respectively. When maintenance units are busy, requests for replacement or repairs are queued. Service discipline is FCFS (first come, first served).

As we see, in a natural way we have a queuing systems with two types of maintenance operations – replacement and repair (renewal). We examine here the case, when m is large number (in practice it might be tens, hundreds, thousands and more), and we will suppose that we have both infinite ($m=\infty$) and finite sources of requests and will get mixed type queuing systems.

In this system, the infinite source of requests for services is the set of main elements and finite source is the set of standby elements and service channels are replacement and repair units. At the same time, one event in a flow of



homogenous events – failure of the main element – generates requests for two parallel maintenance operations. First – the replacement of the failed main element with standby one, and the second – the repair of the very failed element.

Request for replacement occurs due to failure of the main element. The same event, coupled with standby element's failure generates a request for repair (renewal).

To this day, both in reliability theory and queuing theory, the above problems have not been investigated in general case. At the same time, modern research methods of Markov and semi Markov processes allow us to construct and analyse such models in the framework of the mathematical theory of reliability and queuing theory [5, 6].

Only a few special cases of the described system have so far been investigated. 1) m=1, n=1; 2) m=1, n=2; 3) M/M/N - i.e. the repair time length has an exponential distribution, while the replacement time length equals zero (instant replacement); 4) some similar cases have also been investigated. One of such cases (m=1; n=1) is examined in [7].

In the last 7-8 years the specialists of Georgian Technical University (GTU) have succeeded in making considerable progress in the investigation along these lines. In particular, the models have been constructed and partly investigated for the following cases [8 10]:

1) m, n, k, l are arbitrary; the functions F(x) and G(x) are exponential;

2) m, n and the function F(x) are arbitrary; k = l = 1 and the function G(x) is exponential;

3) m, n and the function G(x) are arbitrary; k = 1 = 1 and function F(x) is exponential;

4) some similar statements have also been considered.

Now the investigation for other cases is under way. We call interested in it colleagues to join this work.

Conclusions

In modern large scale territorially dispersed networks standby elements (spare components) are not directly attached to main elements. They are located at specific storage locations and may be tens, hundreds and sometimes thousands of kilometers away from the main elements. Therefore, the duration of standby elements delivery to the place of the failed mains' is quite substantial.

Also taking into account other circumstances, the replacement operation's mean duration often reaches 20-40 % of repair operations' mean duration and therefore, the replacement of the failed element with standby one is quite naturally becoming an independent maintenance operation.

Exactly the novel type queuing systems proposed in this work, in many cases, are the most adequate maintenance models for such networks.

References



- 1. I. Ushakov. Reliability: Past, Present, Future. RTA Journal 1, 10 27 (2006).
- I. Ushakov. Reliability Theory: History & Current State in Bibliographies. RTA Journal 1, 8 35 (2012).
- 3. Mohamed Ben-Daya et al. Handbook of Maintenance Management and Engineering. London: Springer (2009).
- 4. D. Achermann. Modelling, Simulation and Optimization of Maintenance Strategies under Consideration of Logistic Processes. Swiss Federal Institute of Technology, Zurich (2008).
- N. Limnios, G. Oprisan. Semi-Markov Processes and Reliability. Boston: Birkhauser (2001).
- 6. G. Levitin. The Universal Generating Function in Reliability Analysis and Optimization. London: Springer (2005).
- 7. R. Kakubava, R. Khurodze. Probabilistic analysis of the downtime of a duplicated system with recovery and switching. Automation and Remote Control 61, 9, Part 1 (2000).
- 8. R. Kakubava. Multi-line Markov Closed Queuing System for Two Maintenance Operations. RTA Journal 1, 15 22 (2010).
- 9. R. Kakubava, J. Sztrik. Queuing Models with Two Types of Service: Applications for Dependability Planning of Complex Systems. Proc. MMR2011. Beijing (2011).
- R. Kakubava. Reliability Model of Standby System with Replacement Delays of Failed Elements. Automatic Control and Computer Sciences 2, 54 59 (2013).



Linear On/Off Inventory Control

Anna V. Kitaeva, and Natalya V. Stepanova

National Research Tomsk Polytechnic University, Tomsk, Russia (E-mail: kit1157@ yandex.ru)

Abstract. Single-product inventory management model with both random and controllable demand and continuous input product flow with fixed uncontrolled rate under finite storage capacity is considered. We consider the stock level process as asymptotically diffusion process and obtain its stationary distribution. The result permits us to solve the problem of minimizing the variance of the stock level process under linear on/off control and control the probabilities of the stock-out and overflow.

Keywords: On/Off Control, Stochastic Demand, Diffusion Approximation, Inventory Management.

1 The problem statement

A systematic study of inventory models incorporated uncertainly and dynamics began in the early 50s from the works by Arrow *et al.* [1] and Dvoretzky *et al.* [6]. Nowadays a set of stochastic models are available to solve the inventory control problem under various conditions encountered in practice, for example, see Ross [10], Chopra and Meindl [5], and Beyer *et al.* [3].

The aim of the paper is to stabilize the performance of the system under consideration. The feature of the system is exogenous (i.e., outside our control) input product flow.

Let Q(t) be the stock level in the moment t, the input product flow be continuous with fixed rate v_0 , the demands be a Poisson process with constant intensity λ , the values of purchases be i.i.d. random variables having a distribution $F(\cdot)$ with finite the first and second moments equals respectively a_1 and a_2 . The storage capacity let be bounded by Q_{max} . Under certain conditions (for example, the threat of overflow) the product is delivered to outlets and the output flow is assumed to be continuous with a rate $v^*(Q)$.

The aim is to stabilize the stationary process $Q(\cdot)$ in the sense of minimum of its variance and to try to avoid the overflow and stock-out. We use the diffusion approximation of Marcovian process $Q(\cdot)$.

Diffusion methods have been applied in a variety of domains, see Janssen *et al.* [8]; as to application to inventory models, see, for example, Bather [3], Harrison [7], and Puterman [9].

The paper consists of two parts: the first part devotes to the approximation and in the second part we solve the optimization problem.



2 Diffusion approximation

Denote the density

$$P(Q,t) = \frac{\Pr\left[Q \le Q(t) \le Q + dQ\right]}{dQ},$$

and the rate of the product's movement due to non-random factors $v_0 - v^*(Q) = v(Q)$.

Let P(Q, t) be a differentiable function of t, v(Q)P(Q, t) be a differentiable function of Q, and $\int_{0}^{\infty} P(Q+u, t)dF(u) < \infty$.

Derive the Kolmogorov backward equation for a functional of Markov process $Q(\cdot)$

$$\varphi(Q,t) = E\left\{H(Q(\tau))|Q(t) = Q\right\},\$$

and write the adjoint equation, which is the Kolmogorov forward equation for density function P(Q, t) as shown by Barucha-Reid [2]. Consider

$$\begin{split} \varphi(Q, t - \Delta t) &= E\left\{H\left(Q(\tau)\right)\middle|Q(t - \Delta t) = Q\right\} = \\ &= (1 - \lambda\Delta t)E\left\{H\left(Q(\tau)\right)\middle|Q(t) = Q + v(Q)\Delta t\right\} + \\ &+ \lambda\Delta t\int_{0}^{Q} E\left\{H\left(Q(\tau)\right)\middle|Q(t) = Q - u\right\}dF(u) + o(\Delta t) = \\ &= (1 - \lambda\Delta t)\varphi(Q + v(Q)\Delta t, t) + \lambda\Delta t\int_{0}^{Q}\varphi(Q - u, t)dF(u) + o(\Delta t) = \\ &= \varphi(Q, t) + v(Q)\Delta t\frac{\partial\varphi(Q, t)}{\partial Q} - \lambda\Delta t\varphi(Q, t) + \lambda\Delta t\int_{0}^{Q}\varphi(Q - u, t)dF(u) + o(\Delta t), \end{split}$$

which implies

$$-\frac{\partial \varphi(Q,t)}{\partial t} = v(Q) \frac{\partial \varphi(Q,t)}{\partial Q} - \lambda \varphi(Q,t) + \lambda \int_{0}^{Q} \varphi(Q-u,t) dF(u).$$

So the adjoint equation is

$$\frac{\partial P(Q,t)}{\partial t} = -\frac{\partial \left\{ v(Q) P(Q,t) \right\}}{\partial Q} - \lambda P(Q,t) + \lambda \int_{0}^{\infty} P(Q+u,t) dF(u).$$
(1)



To solve (1) suppose that the values of $Q(\cdot)$ are large enough. The idea is to consider some infinitesimal parameter ε so that the process $\varepsilon^2 Q(\cdot)$ is not degenerate. Denote

$$v_1(Q) = v(\varepsilon^2 Q), \ t\varepsilon^2 = \tau, \ Q\varepsilon^2 = x(\tau) + \varepsilon y, \ P(Q, t) = \Pi(y, \tau, \varepsilon),$$
(2)

where $x(\cdot)$ is a differentiable function.

Let the limit $\lim_{\epsilon \to 0} \Pi(y, \tau, \epsilon) = \Pi(y, \tau)$ exists. Let $v_1(\cdot)$ be a differentiable function, $\Pi(y, \tau, \epsilon)$ be a differentiable function with respect to τ and twice differentiable with respect to y.

By substituting (2) into (1) we obtain the equation

$$\varepsilon^{2} \frac{\partial \Pi(y, \tau, \varepsilon)}{\partial \tau} - \varepsilon x'(\tau) \frac{\partial \Pi(y, \tau, \varepsilon)}{\partial y} = -\varepsilon \frac{\partial}{\partial y} \{ v_{1}(x(\tau) + \varepsilon y) \Pi(y, \tau, \varepsilon) \} - \lambda \Pi(y, \tau, \varepsilon) + \lambda \int_{0}^{\infty} \Pi(y + \varepsilon u, \tau, \varepsilon) dF(u).$$
(3)

Rewrite (3)

$$\varepsilon^{2} \frac{\partial \Pi(y,\tau,\varepsilon)}{\partial \tau} - \varepsilon x'(\tau) \frac{\partial \Pi(y,\tau,\varepsilon)}{\partial y} =$$
$$= -\varepsilon \frac{\partial}{\partial y} \left\{ \left[v_{1}(x(\tau)) + \varepsilon y v_{1}'(x(\tau)) \right] \Pi(y,\tau,\varepsilon) \right\} - \lambda \Pi(y,\tau,\varepsilon) + \varepsilon y v_{1}'(x(\tau)) \right\} = -\varepsilon \frac{\partial}{\partial y} \left\{ \left[v_{1}(x(\tau)) + \varepsilon y v_{1}'(x(\tau)) \right] \left[u_{1}(y,\tau,\varepsilon) \right] \right\} - \varepsilon \frac{\partial}{\partial y} \left\{ \left[v_{1}(y,\tau,\varepsilon) + \varepsilon y v_{1}'(x(\tau)) \right] \left[u_{1}(y,\tau,\varepsilon) \right] \right\} - \varepsilon \frac{\partial}{\partial y} \left\{ \left[v_{1}(y,\tau,\varepsilon) + \varepsilon y v_{1}'(x(\tau)) \right] \left[u_{1}(y,\tau,\varepsilon) \right] \right\} - \varepsilon \frac{\partial}{\partial y} \left\{ \left[v_{1}(y,\tau,\varepsilon) + \varepsilon y v_{1}'(x(\tau)) \right] \left[u_{1}(y,\tau,\varepsilon) \right] \right\} - \varepsilon \frac{\partial}{\partial y} \left\{ \left[v_{1}(y,\tau,\varepsilon) + \varepsilon y v_{1}'(x(\tau)) \right] \left[u_{1}(y,\tau,\varepsilon) \right] \right\} - \varepsilon \frac{\partial}{\partial y} \left\{ \left[v_{1}(y,\tau,\varepsilon) + \varepsilon y v_{1}'(x(\tau)) \right] \left[u_{1}(y,\tau,\varepsilon) + \varepsilon y v_{1}'(x(\tau)) \right] \right\} + \varepsilon \frac{\partial}{\partial y} \left\{ \left[v_{1}(y,\tau,\varepsilon) + \varepsilon y v_{1}'(x(\tau)) \right] \left[u_{1}(y,\tau,\varepsilon) + \varepsilon y v_{1}'(x(\tau)) \right] \right\} + \varepsilon \frac{\partial}{\partial y} \left\{ \left[v_{1}(y,\tau,\varepsilon) + \varepsilon y v_{1}'(x(\tau)) \right] \left[u_{1}(y,\tau,\varepsilon) + \varepsilon y v_{1}'(y,\tau,\varepsilon) \right] \right\} + \varepsilon \frac{\partial}{\partial y} \left\{ \left[v_{1}(y,\tau,\varepsilon) + \varepsilon y v_{1}'(x(\tau)) \right] \left[u_{1}(y,\tau,\varepsilon) + \varepsilon y v_{1}'(y,\tau,\varepsilon) \right] \right\} + \varepsilon \frac{\partial}{\partial y} \left[v_{1}(y,\tau,\varepsilon) + \varepsilon y v_{1}'(y,\tau,\varepsilon) \right] \right\}$$

$$+\lambda \int_{0}^{\infty} \left[\Pi\left(y,\tau,\varepsilon\right) + \varepsilon u \frac{\partial \Pi\left(y,\tau,\varepsilon\right)}{\partial y} + \frac{\varepsilon^{2} u^{2}}{2} \frac{\partial^{2} \Pi\left(y,\tau,\varepsilon\right)}{\partial y^{2}} \right] dF\left(u\right) + o\left(\varepsilon^{2}\right).$$

It follows

$$\varepsilon^{2} \frac{\partial \Pi(y, \tau, \varepsilon)}{\partial \tau} = \varepsilon \Big[x'(\tau) - v_{1}(x(\tau)) + \lambda a_{1} \Big] \frac{\partial \Pi(y, \tau, \varepsilon)}{\partial y} -$$

$$-\varepsilon^{2}v_{1}'(x(\tau))\frac{\partial\left\{y\Pi(y,\tau,\varepsilon)\right\}}{\partial y}+\varepsilon^{2}\frac{\lambda a_{2}}{2}\frac{\partial^{2}\Pi(y,\tau,\varepsilon)}{\partial y^{2}}+o(\varepsilon^{2}).$$
 (4)

Let function $x(\cdot)$ be a solution of the equation



$$\frac{dx(\tau)}{d\tau} = v_1(x(\tau)) - \lambda a_1.$$
(5)

Then function $\Pi(\cdot, \cdot)$ satisfies the Fokker-Planck equation

$$\frac{\partial \Pi(y,\tau)}{\partial \tau} = -v_1'(x(\tau))\frac{\partial \{y\Pi(y,\tau)\}}{\partial y} + \frac{\lambda a_2}{2}\frac{\partial^2 \Pi(y,\tau)}{\partial y^2}.$$

Consequently the process $y(\tau, \varepsilon) = \frac{\varepsilon^2 Q(t) - x(\tau)}{\varepsilon}$ converges in distribution to the Ornstein–Uhlenbeck process $y(\cdot)$ as $\varepsilon \to 0$ satisfying the following stochastic differential equation

$$dy(\tau) = v_1'(x(\tau)) y d\tau + \sqrt{\lambda a_2} dw(\tau).$$
(6)

Let $v_1(\cdot)$ be a twice differentiable function. From (5) and (6) we get that the process

$$z(\tau) = x(\tau) + \varepsilon y(\tau) \tag{7}$$

satisfies

$$dz(\tau) = \left(\nu_1(z) - \lambda a_1\right) d\tau + \varepsilon \sqrt{\lambda a_2} dw(\tau) + \frac{\varepsilon^2}{2} R_2 d\tau, \qquad (8)$$

where $R_2 = -y^2 v_1''(\epsilon y \theta)$, $0 \le \theta \le 1$. Indeed it is clear

$$dz(\tau) = dx(\tau) + \varepsilon dy(\tau) = \left(v_1(x(\tau)) + \varepsilon y v_1'(x(\tau)) - \lambda a_1\right) d\tau + \varepsilon \sqrt{\lambda a_2} dw(\tau).$$
(9)

By Taylor expansion with Lagrange remainder we obtain

$$v_1(z) = v_1(x + \varepsilon y) = v_1(x) + \varepsilon y v_1'(x) + \frac{\varepsilon^2}{2} y^2 v_1''(x + \theta \varepsilon y).$$

By substituting $v_1(x) + \varepsilon y v_1'(x) = v_1(z) - \frac{\varepsilon^2}{2} y^2 v_1''(x + \theta \varepsilon y)$ into (9) we

obtain (8).

We use (2) and (7) to get asymptotic equation

$$\varepsilon^2 Q(t) = x(\tau) + \varepsilon y(\tau) = z(\tau).$$

From (8) we get



$$\varepsilon^2 dQ(t) = \left(\nu_1\left(\varepsilon^2 Q\right) - \lambda a_1\right) d\tau + \varepsilon \sqrt{\lambda a_2} dw(\tau) - \frac{\varepsilon^2}{2} R_2 d\tau$$

which implies using $v(Q) = v_1(\varepsilon^2 Q)$

$$dQ(t) = \left(v(Q(t)) - \lambda a_1\right) \frac{d\tau}{\varepsilon^2} + \sqrt{\lambda a_2} \frac{dw(\tau)}{\varepsilon} - \frac{1}{2}R_2 d\tau.$$

Since $t\epsilon^2 = \tau$, we have the equation

$$dQ(t) = \left(v(Q(t)) - \lambda a_1\right)dt + \sqrt{\lambda a_2}dw(t) - \frac{1}{2}R_2d\tau.$$

So approximately the equation holds

$$dQ(t) = (v(Q) - a_1\lambda)dt + \sqrt{a_2\lambda}dw(t)$$

Because of the boundedness of $Q(\cdot)$ the stationary distribution exists

$$p(s) = C \cdot \exp\left(\frac{2}{a_2\lambda} \int (v(s) - a_1\lambda) ds\right),$$
(10)

where C is the normalization constant.

3 Linear on/off control

If the inventory level $Q(\cdot)$ is above the some breakdown (base-stock) level $Q_{\max} - Q_0$ we begin to deliver the product to outlets with a rate proportional to the difference $Q - (Q_{\max} - Q_0)$ to prevent the stock's overflow. Thus

$$\nu(Q) = \begin{cases} \nu_0, \text{ if } Q < Q_{\max} - Q_0, \\ \nu_0 - \beta \left(Q - \left(Q_{\max} - Q_0 \right) \right), \text{ if } Q > Q_{\max} - Q_0, \end{cases}$$
(11)

and $v_0 > a_1 \lambda$, $\beta > 0$.

The condition $v_0 > a_1 \lambda$ means that if the inventory level is below the base-stock level, then the stock level is replenished in the mean, i.e., the warehouse accumulates goods.

We use (10) and (11) to get



$$p(x) = C \exp\left(\frac{2}{a_2\lambda}(v_0 - a_1\lambda)(x - (Q_{\max} - Q_0)))\right), \text{ if } x < Q_{\max} - Q_0;$$

and

$$p(x) = C \exp\left(\frac{2}{a_2\lambda}\left((v_0 - a_1\lambda)\left(x - \left(Q_{\max} - Q_0\right)\right) - \frac{\beta\left(x - \left(Q_{\max} - Q_0\right)\right)^2}{2}\right)\right), \text{ if } x > Q_{\max} - Q_0,$$

where

$$C^{-1} == \frac{1-2b\Phi(b)\exp(b^2)}{2d},$$

$$d = \frac{v_0 - a_1 \lambda}{a_2 \lambda} > 0, \ b = -d \sqrt{\frac{a_2 \lambda}{\beta}} < 0, \ \Phi(b) = \int_b^\infty \exp\left(-t^2\right) dt.$$

The expectation of the inventory level process is

$$E(Q) = \frac{1}{1 - 2b\Phi(b)\exp(b^2)} (Q_{\text{max}} - b^2)$$

$$-Q_0 - \frac{1}{2d} - 2b\left(\left(Q_{\max} - Q_0 - \frac{b^2}{d}\right)\Phi(b)\exp\left(b^2\right) - \frac{b}{2d}\right)\right).$$

It's variance is

$$Var(Q) = \frac{1}{2d^2} \left(b^2 + \frac{1+b^2}{1-2b\Phi(b)\exp(b^2)} \right) = \frac{g(b)}{2d^2}.$$
 (12)

The value b giving the minimal value of the variance (12) is

/

$$b_0 \approx -0.563, \quad g(b_0) \approx 0.734.$$
 (13)

`

The probability of the overflow is

$$\alpha = P(Q(t) > Q_{\max}) = \frac{2b\Phi\left(b - \frac{d}{b}Q_0\right)\exp(b^2)}{2b\Phi(b)\exp(b^2) - 1}.$$
 (14)

The probability (14) takes the maximal value under $Q_0 = 0$



$$\alpha_{\max} = 1 + \frac{1}{2b_0 \Phi(b_0) \exp(b_0^2) - 1}$$

We use (13) to compute $\alpha_{max} \approx 0,68$. Minimal value of α under optimal control is

$$\alpha_{\min} = -2b_0^3 \Phi\left(b_0 - \frac{d}{b_0}Q_{\max}\right) \exp\left(b_0^2\right) \approx 0.48\Phi\left(-0.563 + \sqrt{\frac{\beta_0}{a_2\lambda}}Q_{\max}\right).$$

Storage capacity $Q_{max}^{0\alpha}$ given desirable α_{min}^0 is

$$Q_{\max}^{0\alpha} \approx \left(\Psi\left(2.08\alpha_{\min}^{0}\right) + 0.563\right) \sqrt{\frac{a_2\lambda}{\beta_0}}, \qquad (15)$$

where $\Psi(\cdot)$ is the inverse function of $\Phi(\cdot)$.

The probability of the stock-out (the warehouse is empty) is

$$\gamma = P(Q < 0) == \frac{\exp\left(-2d\left(Q_{\max} - Q_0\right)\right)}{1 - 2b\Phi(b)\exp\left(b^2\right)}.$$

The optimal base-stock level given γ_0 is

$$Q_{\max} - Q_0 = \frac{2\ln b_0 - \ln \gamma_0}{2d}.$$

Maximal value of $\boldsymbol{\gamma}$ is

$$\gamma_{\max} = \frac{1}{1 - 2b_0 \Phi(b_0) \exp(b_0^2)} \approx 0.316$$

under $Q_0 = Q_{\text{max}}$. Minimal value of γ is

$$\gamma_{\min} = \frac{\exp\left(-2dQ_{\max}\right)}{1 - 2b_0\Phi(b_0)\exp\left(b_0^2\right)} \approx 0.316\exp\left(-2dQ_{\max}\right).$$

The optimal stock capacity ${\it Q}^{0\gamma}_{max}$ given desirable γ^0_{min} is

$$Q_{\max}^{0\gamma} \approx -\frac{\ln\left(3.16\gamma_{\min}^{0}\right)}{2d}.$$
 (16)



Combining (15) and (16) we receive the storage capacity Q_{max}^0 making it possible to choose $\alpha \in [\alpha_{\min}; 0.68]$ and $\gamma \in [\gamma_{\min}; 0.316]$

$$Q_{\max}^{0} = \max\left(\left(\Psi\left(2.08\alpha_{\min}\right) + 0.563\right)\sqrt{\frac{a_{2}\lambda}{\beta_{0}}}, -\frac{\ln\left(3.16\gamma_{\min}\right)}{2d}\right).$$

4 Conclusions

Trade-off between the probabilities of overflow and stock-out under linear on/off control can be overcome only by increasing of stock's capacity assuming that the input flow is beyond our power.

Note also that the overflow's probability $\alpha_{max} = 0.68$ more than twice as big as the stock-out's probability $\gamma_{max} = 0.316$ under optimal control. So presumably we need to consider more complicated, nonlinear models of controlled output flow.

References

- K.J. Arrow, Th.E. Harris and J. Marschak. Optimal Inventory Policy, Econometrica, 19, 3, 205–272, 1951.
- 2.A.T. Barucha-Reid. *Elements of the Theory of Markov Processes and Their Applications*. McGraw-Hill, New York, 1960.
- 3.J.A. Bather. A Continuous Time Inventory Model, *Journal of Applied Probability*, 3, 538–549, 1966.
- 4.D. Beyer, F. Cheng, S.P. Sethi and M. Taksar. *Markovian demand inventory models*. Springer, New York, 2010.
- 5.S. Chopra and P. Meindl. Supply chain management. Prentice Hall, London, 2001.
- 6.A. Dvoretzky, J. Kiefer and J. Wolfowitz. On the optimal character of the (S, s) policy in inventory theory, *Econometrica*, 21, 586–596, 1953.
- 7.J.M. Harrison. *Brownian Motion and Stochastic Flow Systems*. John Wiley and Sons, New York, 1985.
- 8.J. Janssen, O. Manca and R. Manca. *Applied Diffusion Processes from Engineering to Finance*. John Wiley and Sons, London, 2013.
- 9.M. Puterman. A diffusion process model for a storage system. In Geisler, M.A. (ed.), *Studies in the Management Sciences*, *Logistics*, I, 143–159, North-Holland Press, Amsterdam, 1975.
- 10.Sh.M. Ross. *Applied probability models with optimization applications*. Dover Publications, New York, 1992.



Modelling structural changes in relations between returns of selected REIT indexes

Jozef Komorník¹, Magda Komorníková², and Cuong Nguyen³

1	Faculty of Management, Comenius University, Odbojárov 10, P.O.BOX 95, 820 05
	Bratislava, Slovakia
	(F-mail: Jozef Komornik@fm uniba sk)

- ² Faculty of Civil Engineering, Slovak University of Technology, Radlinského 11, 813 68 Bratislava, Slovakia
- (E-mail: Magdalena.Komornikova@stuba.sk)
- ³ Faculty of Commerce, Lincoln University NZ, Canterbury, New Zealand (E-mail: Cuong.Nguyen@lincoln.ac.nz)

Abstract. We have investigated the relations between 8 selected countries' (USA, Canada, Japan, Australia, Hongkong, Singapore, UK and France) daily returns of the REIT (Real Estate Investment Trust) indexes in the time period January 3, 2000 – May 8, 2012, divided in 3 subperiods bounded by the recent global financial market crises (July 1, 2008 – April 30, 2009). We have observed that in the postcrisis subperiod the influence of the delayed values (by 1 day) of the returns of the US REIT index on the REIT indexes of Japan and Australia greatly increased, while the same effect did not take place for the remaining (even East Asian) REIT indexes. We used the copula approach for fitting the optimal models for the investigated relations. **Keywords:** Real Estate Investment Trust (REIT), Returns of REIT indices, Copula, Archimedean copula, Reflection of copulas .

1 Introduction

The aim of this paper is to analyse the relations between the above mentioned 8 selected countries daily returns of the REIT (Real Estate Investment Trust) indexes in different time periods, determined by the recent global financial markets crises (July 1, 2008 – April 30, 2009). Our aim was to investigate the influence of the (expectedly) dominant US Real Estate (RE) market (where the crises has been initiated) on the remaining RE markets in the group, as well as its development in the different considered time subperiods. From this point of view, the most interesting change has been observed in the third (postcrisis) subperiod with considerably stronger relations of the delayed (by 1 day) returns of the US REIT index to the couple of returns of the Japan's and Australia's REIT indexes in comparison to the returns of the remaining 6 REIT indexes.

The paper is organized as follows. The second section contains the results of nonparametric correlation analyses (based on the Kendall coefficients) of the returns of the considered group of REIT indexes that have been filtered (in order to avoid a possible violation of the i.i.d. property) by ARMA–GARCH models (separately in the individual subperiods of time). The third section is devoted to a brief overview of the theory of copulas including the methodology of their fitting to two–dimensional time series. The fourth section contains



an overview of the best copula models for different time subperiods and selected significantly correlated pairs of returns of REIT indexes. Finally, some conclusions are presented.

2 The results of nonparametric correlation analyses of the returns of the REIT indexes filtered by ARMA-GARCH models

The results for different time subperiods are contained in the following Tables 1, 2 and 3. The considered countries are represented by the following consecutive number:

- 1. USA
- 2. Canada
- 3. Australia
- 4. Japan
- 5. Hongkong
- 6. Singapore
- 7. France
- 8. UK.

countries	1	2	3	4	5	6	7	8
1	х	0.937	0.920	0.508	0.868	0.686	0.679	0.567
2	0.937	х	0.880	0.519	0.885	0.696	0.689	0.559
3	0.920	0.880	х	0.498	0.835	0.662	0.658	0.571
4	0.508	0.519	0.498	х	0.527	0.496	0.464	0.399
5	0.868	0.885	0.835	0.527	х	0.714	0.681	0.554
6	0.686	0.696	0.662	0.0.496	0.714	х	0.587	0.516
7	0.679	0.689	0.658	0.464	0.681	0.587	х	0.565
8	0.567	0.559	0.571	0.399	0.554	0.516	0.565	х

Table 1. The values of the Kendall's correlation coefficient τ for the first (precrisis) time period

We can observe that all values of the Kendall correlation coefficients are (quite surprisingly) high during the first two subperiods. They also remain quite high in the third period for the group comprising Canada (2), Hongkong (5), Singapore (6), France (7) and UK (8) as well as for the couple Australia (3) and Japan (4). However, their values for the remaining couple of countries are considerably diminished.

Realizing that there exists a time lag between the (presumably) most influential US RE market and another considered RE markets (except for Canada), we also calculated the values of the Kendall correlation coefficients for the delayed (by 1 day) values of the returns of the US REIT index with the other returns of REIT indexes that are presented in the Table 4. We see that their

countries	1	2	3	4	5	6	7	8
1	х	0.377	0.301	0.267	0.313	0.328	0.350	0.306
2	0.377	х	0.692	0.549	0.868	0.731	0.515	0.391
3	0.301	0.692	х	0.535	0.667	0.716	0.486	0.397
4	0.267	0.549	0.535	х	0.586	0.556	0.488	0.378
5	0.313	0.868	0.667	0.586	х	0.707	0.503	0.365
6	0.328	0.731	0.716	0.556	0.707	х	0.496	0.419
7	0.350	0.515	0.486	0.488	0.503	0.496	х	0.626
8	0.306	0.391	0.397	0.378	0.365	0.419	0.626	х

Table 2. The values of the Kendall's correlation coefficient τ for the second (crisis) time period

countries	1	2	3	4	5	6	7	8
1	x	0.047	0.111	0.061	0.057	0.078	0.220	0.221
2	0.047	x	-0.005	-0.004	0.898	0.828	0.613	0.584
3	0.111	-0.005	x	0.222	0.013	0.052	0.059	0.087
4	0.061	-0.004	0.222	x	0.018	0.064	0.062	0.073
5	0.057	0.898	0.013	0.018	x	0.840	0.622	0.594
6	0.078	0.828	0.052	0.064	0.840	х	0.643	0.617
7	0.220	0.613	0.059	0.062	0.622	0.643	x	0.736
8	0.221	0.584	0.087	0.073	0.594	0.617	0.736	x

Table 3. The values of the Kendall's correlation coefficient τ for the third (postcrisis) time period

values are very low in comparison with the corresponding values of the Kendall correlation coefficients of nondelayed values of the returns of the US REIT index with the returns of the another considered REIT indexes for the first 2 subperiods (while reaching considerably higher levels in the third subperiod only for the pairs with Australia (3) and Japan (4)).

	03/01/2000 - 31/07/2008	01/08/2008 - 30/04/2009	01/05/2009 – 08/05/2012
2	0.011	0.004	-0.002
3	0.009	0.091	0.241
4	0.009	0.191	0.276
5	0.008	0.023	0.012
6	0.009	0.065	0.053
$\overline{7}$	0.003	0.154	0.059
8	-0.016	0.175	0.080

Table 4. The values of the Kendall's correlation coefficient τ for the delayed values (by 1 day) of the returns of the US REIT index with those of the other considered countries in different subperiods of time



3 Theoretical concepts

Let (X, Y) be a 2-dimensional random vector with a joint distribution F_{XY} and marginal distribution functions F_X , F_Y . We will use the standard definition of a copula (see e.g., Joe[4], Nelsen[7]) $C(u, v) : [0, 1]^2 \to [0, 1]$ satisfying

$$F_{XY}(x,y) = C(F_X(x), F_Y(y)) \tag{1}$$

and corresponding density function (if C is absolutely continuous)

$$c(u,v) = \frac{\partial^2}{\partial u \,\partial v} C(u,v). \tag{2}$$

In our subsequent investigations, we mainly utilize four copula families of Archimedean class: Gumbel, strict Clayton, Frank and Joe BB1 (see e.g., Embrechts *et al.*[2], Joe[4], Nelsen[7]). Recall that their are given by the following expressions.

• Gumbel family

$$C^G_\theta(u,v) = \exp^{-\left((-\ln u)^\theta + (-\ln v)^\theta\right)^{\frac{1}{\theta}}} \tag{3}$$

where $\theta > 1$. Let us note that $C_1^G(u, v) = \Pi(u, v) = u \cdot v$. • Strict Clayton family (Kimeldorf and Sampson)

$$C^{C}_{\theta}(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{\frac{-1}{\theta}}$$

$$\tag{4}$$

for $\theta > 0$, $C_0^C(u, v) = \Pi(u, v) = u \cdot v$.

To enrich the classes of considered models, we also consider the classes of Frank and Joe BB1 copulas.

• Frank family

$$C_{\theta}^{F}(u,v) = -\frac{1}{\theta} \log \left(1 - \frac{(1 - e^{-\theta u})(1 - e^{-\theta v})}{(1 - e^{-\theta})} \right)$$
(5)

for $\theta > 0$, $C_0^F(u, v) = \Pi(u, v) = u \cdot v$.

• Joe's family

$$C^{J}_{\theta}(u,v) = 1 - \left((1-u)^{\theta} + (1-v)^{\theta} - (1-u)^{\theta} (1-v)^{\theta} \right)^{1/\theta}$$
(6)

for $\theta \geq 1$.

A rich overview of Archimedean copulas is presented in Embrechts *et al.*[2], Genest and Favre[3], Joe[4] and Nelsen[7].

Let us recall that for a given copula C(u, v) the lower (left) and upper (right) tail dependence coefficients are defined by

$$\lambda_L(C) = \lim_{\delta \to 0} \Pr\left(F_Y(y) \le \delta \mid F_X(x) \le \delta\right) = \lim_{\delta \to 0} \frac{C(\delta, \delta)}{\delta} =$$
(7)
$$= \lim_{\delta \to 0} \Pr\left(F_X(x) \le \delta \mid F_Y(y) \le \delta\right)$$



and

$$\lambda_R(C) = \lim_{\delta \to 0} \Pr\left(F_Y(y) \ge 1 - \delta \mid F_X(x) \ge 1 - \delta\right) = \lim_{\delta \to 0} \frac{2\delta - 1 + C(1 - \delta, 1 - \delta)}{\delta} = \lim_{\delta \to 0} \Pr\left(F_X(x) \ge 1 - \delta \mid F_Y(y) \ge 1 - \delta\right).$$
(8)

It is well known (see e.g. Joe[4] or Nelsen[7]) that the Gumbel copula C^G_{θ} , Clayton copula C^C_{θ} and Joe copula C^J_{θ} satisfy the relation

$$\lambda_L(C_{\theta}^G) = 0, \quad \lambda_R(C_{\theta}^G) = 2 - 2^{\frac{1}{\theta}},$$
$$\lambda_L(C_{\theta}^C) = 2^{-\frac{1}{\theta}}, \quad \lambda_R(C_{\theta}^C) = 0,$$

and

$$\lambda_L(C^J_{\theta}) = 0, \quad \lambda_R(C^J_{\theta}) = 2 - 2^{\frac{1}{\theta}}.$$

It is also well known (see Embrechts et al[2]) that the values of λ_R and λ_L for Frank copulas are equal to 0.

We follow the approach of Patton[10] and consider a so–called *survival cop-ula* derived from a given copula C(u, v) corresponding to the couple (X, Y) by

$$SC(u,v) = u + v - 1 + C(1 - u, 1 - v)$$
(9)

which is the copula related to the couple (-X, -Y) with the marginal distribution functions

$$F_{-X}(x) = 1 - F_X(-x^+)$$
 and $F_{-Y}(y) = 1 - F_Y(-y^+)$. (10)

Obviously, if a copula C represents the right or left tail dependence, its survival copula SC represents the opposite one.

Convex combinations of copulas and corresponding survival copulas has been successfully applied for modelling of exchange rates dependences (e.g. in Patton [10] and Ning[8], [9]).

Applying reflections of copulas (left, right and composed), we can construct new copulas that exhibit interesting properties concerning additional coefficients of tail dependencies (see Komorník and Komorníková[5,6]).

It is well known that for the convex sums of copulas, the corresponding density function is the convex sum (with the same weights) of incoming density functions. The same kind of mixing property holds for the above mentioned coefficients of tail dependencies.

3.1 Fitting of copulas

In practical fitting of the data we have utilized the maximum pseudolikelihood method (MPL) of parameter estimation (with initial parameters estimates received by the minimalization of the mean square distance to the empirical copula C_n presented e.g. in Genest and Favre[3]). It requires that the copula $C_{\theta}(u, v)$ is absolutely continuous with density $c_{\theta}(u, v) = \frac{\partial^2}{\partial u \partial v} C_{\theta}(u, v)$. This



method (described e.g. in Genest and Favre[3]) involves maximizing a rankbased log-likelihood of the form

$$L(\theta) = \sum_{i=1}^{n} \ln\left(c_{\theta}\left(\frac{R_i}{n+1}; \frac{S_i}{n+1}\right)\right)$$

where *n* is the sample size, R_i stands for the rank of X_i among X_1, \ldots, X_n, S_i stands for the rank of Y_i among Y_1, \ldots, Y_n and θ is vector of parameters in the model. Note that arguments $\frac{R_i}{n+1}$, $\frac{S_i}{n+1}$ equal to the corresponding values of the empirical marginal distributional functions of random variables X and Y.

4 Application to real data modelling by copula models

We considered models from Frank (C_{θ}^{F}) , Joe (C_{θ}^{J}) , strict Clayton (C_{θ}^{C}) and Gumbel (C_{θ}^{G}) families and their pairwise convex combinations (we will call them *mixed 2 families copulas* as well as the convex combinations with their survival copulas (we will call them *mixed 1 family copulas*).

For selecting the optimal models we applied the Kolmogorov – Smirnov Anderson–Darling (KSAD, for which we use the abbreviation AD) test statistic defined e.g. in Berg and Bakken[1] (for which we also constructed a GoF simulation based test), when comparing models with their submodels and different families of models.

Surprisingly, quite successful in this selection process (especially for the crisis period) have been symmetric Gumbel model of the type

$$0.5 \left(C_{\theta}^{G} + S C_{\theta}^{G} \right)$$

(that have equal values of the left and right tail dependence coefficients λ_L and λ_R).

The best fitted copulas for selected couples for the first subperiods are presented in Table 5 (with the returns of the nondelayed US REIT index) and Table 6 contains the results for the third subperiod (with the returns of the US REIT index delayed by 1 day).

4.1 Best models for the first period.

In this period, we identified only one symmetric Gumbel type model (for the couple Japan & Hongkong). Several optimal models have the form of mixed 1 family copulas. However, quite many optimal models for this period have the form of convex combinations of strict Clayton copulas with Gumbel copulas.

4.2 Best models for the second period.

For the crisis period we can observe mainly optimal models in the form of mixed Gumbel copulas, many of them symmetric. Exceptionally, for the couple of Canada & Hongkong, the optimal copula is of the Gumbel type.

Couple	03/01/2000 - 31/07/2008	01/08/2008 - 30/04/2009
1 & 2	$\alpha \cdot C_{\theta_1}^G + (1 - \alpha) \cdot SC_{\theta_2}^G$	$\alpha \cdot C_{\theta_1}^G + (1 - \alpha) \cdot SC_{\theta_2}^G$
1 & 3	$\alpha \cdot C_{\theta_1}^C + (1 - \alpha) \cdot C_{\theta_2}^G$	$0.5\left(C_{\theta}^{G} + SC_{\theta}^{G}\right)$
1 & 4	$\alpha \cdot C_{\theta_1}^C + (1 - \alpha) \cdot C_{\theta_2}^G$	$0.5\left(C_{\theta}^{G} + SC_{\theta}^{G}\right)$
1 & 5	$\alpha \cdot C^J_{\theta_1} + (1 - \alpha) \cdot SC^J_{\theta_2}$	$0.5\left(C_{\theta}^{G} + SC_{\theta}^{G}\right)$
2& 3	$\alpha \cdot C^J_{\theta_1} + (1 - \alpha) \cdot SC^J_{\theta_2}$	$0.5\left(C_{\theta}^{G} + SC_{\theta}^{G}\right)$
2 & 5	$\alpha \cdot C_{\theta_1}^C + (1 - \alpha) \cdot C_{\theta_2}^G$	$C^G_{ heta}$
3 & 5	$\alpha \cdot C_{\theta_1}^C + (1 - \alpha) \cdot SC_{\theta_2}^C$	$\alpha \cdot C_{\theta_1}^G + (1 - \alpha) \cdot SC_{\theta_2}^G$
4 & 5	$0.5\left(C_{\theta}^{G} + SC_{\theta}^{G}\right)$	$0.5\left(C_{\theta}^{G} + SC_{\theta}^{G}\right)$
5&6	$\alpha \cdot C_{\theta_1}^C + (1 - \alpha) \cdot C_{\theta_2}^G$	$\alpha \cdot C_{\theta_1}^C + (1 - \alpha) \cdot SC_{\theta_2}^C$
7 & 8	$\alpha \cdot C_{\theta_1}^C + (1-\alpha) \cdot C_{\theta_2}^G$	$\alpha \cdot C_{\theta_1}^G + (1-\alpha) \cdot SC_{\theta_2}^G$

Table 5. The overview of optimal types of copulas for selected pairs of the (filtered) returns of REIT indexes for precrisis and crisis subperiods (with the returns of the nondelayed US REIT index)

Couple	01/05/2009 – 08/05/2012
1 & 3	$\alpha \cdot C^J_{\theta_1} + (1 - \alpha) \cdot SC^J_{\theta_2}$
1 & 4	$\alpha \cdot C_{\theta_1}^G + (1 - \alpha) \cdot SC_{\theta_2}^G$
5&6	$\alpha \cdot C_{\theta_1}^C + (1 - \alpha) \cdot C_{\theta_2}^G$
5&7	$\alpha \cdot C_{\theta_1}^C + (1 - \alpha) \cdot SC_{\theta_2}^C$
5&8	$\alpha \cdot C_{\theta_1}^C + (1 - \alpha) \cdot SC_{\theta_2}^C$
6&7	$\alpha \cdot C_{\theta_1}^C + (1 - \alpha) \cdot C_{\theta_2}^G$
6&8	$\alpha \cdot C_{\theta_1}^C + (1 - \alpha) \cdot C_{\theta_2}^G$
7&8	$0.5 (C_{\theta_1}^G + SC_{\theta_2}^G)$

Table 6. The overview of optimal types of copulas for selected pairs of the (filtered) returns of REIT indexes for postcrisis subperiod (with the returns of the US REIT index delayed by 1 day)

4.3 Best models for the third period.

Here we can observe just one symmetric Gumbel optimal model (for the couple France & UK), several mixed 2 families models (for Joe BB1, Clayton and Gumbel families) as well as convex combinations of Clayton and Gumbel copulas.

5 Concluding remarks

This paper uses copula models to examine the dependence structure between returns of international RE markets in 3 different time periods (separated by the recent financial crisis). The paper employs Archimedean and mixed copula models of Clayton, Gumbel, Frank and Joe families as well as their survival copulas. The flexibility of using mixed copulas is that it allows us to capture different structures among variables, including both right and/or left tail dependence structure at the same time.



Our results prove that using copula, especially mixed copula, to capture dependency is a useful and flexible approach. There are several opportunities for future research, such as extending the number of REIT indices and/or examining the relationship between different investment in commodity markets, bond and stock markets, using the proposed method in this paper (possibly also with an extended range of copula families, for example asymmetric logistic model copula (ALM), and mixed asymmetric logistic model copula (MALM)).

Acknowledgement

This work was partially supported by Slovak Research and Development Agency under contract No. APVV-0073–10 and by VEGA 1/0143/11.

References

- 1.D. Berg and H. Bakken. *Copula Goodness-of-fit Tests: A Comparative Study.* Working paper, University of Oslo and Norwegian Computing Center, 2006.
- 2.P. Embrechts, F. Lindskog and A. McNail. Modeling dependence with copulas and applications to risk management. Rachev, S. (Ed.) Handbook of Heavy Tailed Distributions in Finance. Elsevier, Chapter 8, 329–384, 2001.
- 3.C. Genest and A.C. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering* 12, 347–368, 2007.
- 4.H. Joe. Models and Dependence Concepts. Chapman and Hall, London, UK, 1997.
- 5.J. Komorník and M.Komorníková: Reflections of copulas and their applications in modelling of financial data. Forum Statisticum Slovacum, 1, 12–19, 2012.
- 6.J. Komorník and M.Komorníková. Modelling financial time series using reflections of copulas. *Kybernetika*, 49 (3), 487–497, 2013.
- 7.R.B. Nelsen. An introduction to copulas. Lecture Notes in Statistics 139, Springer-Verlag, New York, 1999.
- 8.C. Ning. *Extreme dependence of international stock market*. Working paper, Ryerson University, 2008.
- 9.C. Ning. Dependence structure between the equity market and the foreign exchange market - a copula approach. *Journal of International Money and Finance*, 29, 5, 743–759, 2010.
- 10.A.J. Patton. Modelling Asymmetric Exchange Rate Dependence. International Economic Review, 47, 2, 527–556, 2006.



Reliability Evaluation of Multi-Camera Motion Detector by using Monte-Carlo Simulator

Samuel Kosolapov

Signal and Image Processing Laboratory, ORT Braude Academic College of Engineering, Karmiel, Israel E-mail: ksamuel@braude.ac.il

Abstract. Motion Detectors (MD) utilizing one digital camera, are well-known and widely used for detection of physical objects intrusion into protected zone. However, one-camera MD operation is limited to the protection of 2D region of fixed size, which significantly limit practical usage of MD of this kind. Stereo Motion Detectors (SMD utilizing two video cameras, can detect physical violation of the user specified 3D volume, but, as it was shown by earlier research, two-camera setup has low reliability for some motion paths, which lower total SMD reliability. In this research reliability of Multi-Camera Motion Detector (MCMD was evaluated by using Monte-Carlo software simulator (implemented by using MAPLE script). Reliability of a number of practically interested setups was analyzed.

Keywords: Image Processing, 3D Imaging, Stereo Camera, Motion Detector, Monte-Carlo simulation, MAPLE

1 Introduction

Motion Detectors (MD) utilizing one digital camera, are well-known and widely used for detection of physical objects intrusion into protected zone [1]. For most MD any significant change in the content of the frame grabbed by digital camera is treated as "security violation event". This means that MD operation is limited to the protection of 2D region of fixed size, which significantly limits practical usage of MD. Stereo Motion Detectors (SMD) utilizing two video cameras, can detect physical violation of the user specified 3D volume [2,3], but, as it was shown by earlier research, two-camera setup has low reliability for some motion paths, which lower total SMD reliability [4]. Constantly dropping prices on high-resolution digital cameras makes implementation of Multi-Camera Motion Detector (MCMD) practical, at least for the case of 3 or 4 digital cameras [5]. It seems obvious that increase in the number of cameras increases MCMD reliability, however, not every multi-camera setup is practically feasible because of camera(s) calibration need. Operation of SMD and MCMD in most cases requires some kind of calibration, which, in some cases, is problematic in the real-life conditions. Hence, it would be preferable to utilize setups that can be aligned during assembly, thus, effectively eliminating need for "after-assembly" calibration. In order to evaluate accuracy and reliability of the selected "aligned" MCMD setup, software simulator was designed and implemented by using MAPLE.





2 SMD and MCMD Exemplary Setups



On the Fig. 2 exemplary two-camera (stereo) SMD setup is presented. This setup was analyzed in the previous work [3, 4]. Important that main optical axes of both cameras pass point "O" (origin) selected by operator during alignment step. This alignment can be easily achieved with adequate accuracy by using motorized cameras (as shown in Fig. 2).

Fig. 3 presents geometry of SMD setup in the XY plane (created by points {"A", "B", "O"}). From this figure relations between physical coordinates {X,Y} of the exemplary point "T" and columns of the image of the point "T" on the sensors of the Left and Right cameras {ColR, ColR} can be derived (Fig.4). Camera parameters {W, FL, FR, ps} are known from camera specifications. In the frames of "Alignment Instead of Calibration" approach, geometric distances of the setup (like AC, BC) are not measured, but assumed as known with some tolerance.





Geometry of the setup in the Z plane and relevant equations are trivial (standard lens equations) and thus not shown here.

By using equations presented on the Fig. 4 and equations for the Z plane, we can evaluate physical coordinates of the exemplary point "T" $\{X,Y,Z\}$ by row and columns of the image of this point on the sensors of both cameras

{RowL, ColL} and {RowR, ColR}.

When point "T" is moving, {RowL, ColL} and {RowR, ColR} are changing, and thus, we can detect if motion in the user specified region was happen. However, from the analysis of Fig.2 and Fig. 3 can be seen that if point T is moving in the direction close to the direction optical axis of the (say) Left camera, changes of {RowL, ColL} and very small. In this case, 3D motion detection become non-reliable: actually, 3D two-camera setup operates as 2D one-camera setup.

In attempt to eliminate the problem of this directional sensitivity, three-camera MCMD setup (see Fig. 5) was tested. It is clear, that three cameras operates like three SMD: SMD#1 (Left and Right Cameras), SMD#2 (Left and Central Cameras), SMD#3 (Central and Right Cameras). In this situation, motion in the direction of the optical axis of (say) Left camera makes operation of the SMD#1 and SMD#2 non-reliable, but SMD#3 operates in the reliable way. So, in case one of tree SMD raises "alarm", violation is considered as detected.

In order to validate this statement for the selected scene and for the selected setup in the quantitative manner, Monte-Carlo software simulation was executed.





3 Monte-Carlo Simulation Procedure

In order to evaluate the feasibility of the selected "Alignment Instead of Calibration" approach, for the MCMD configuration in test, MAPLE-based software simulations were performed. On the first stage, developed simulator calculates series of digital images of the objects of the scene including "violating object" for all three cameras. By using calculated coordinates of the "violating object" in accordance with selected "violation path", pseudo-video of the violation is generated for all three cameras as a sequence of digital images (frames) in accordance with equations presented on Fig. 4. The simplest onecamera motion detector detects motion of the "violating object" by processing images created as pixel-by-pixel differences of the consequent frames. In case no changes in the scene happened, difference image is "black". Practically, camera noise is present. This noise is one of the factors leading to lower of reliability the camera-based motion detector [1]. On the next step, positions of "non-black" regions of two cameras images are used to evaluate {row-column} pairs and their correspondent {X, Y, Z} coordinates. In case any of $\{X, Y, Z\}$ is inside the user defined protection zone, "alarm" must be raised. Simulator takes into account camera' parameters and noise, effect of digitization and assembly errors of MCMD setup. Well-known mathematical models of different 3D setups cannot be used directly to evaluate accuracy and reliability of MCMD, because a number of parameters cannot be measured exactly for non-calibrated setup. Thus, classical Monte-Carlo approach was used to evaluate accuracy and reliability of the MCMD configuration in test. Simulator is operated a number of times, while, for every simulation run, values of the selected set of setup parameters are modified in a pseudo-random way. User can select parameters and assembly tolerances of the selected MCMD setup, geometry of 3D volume to be protected, intrusion object size and its motion path. As stated before, first unit of this simulator generates plurality of digital images in the situation when 3D object of specified size and shape is moving over specified 3D path. First unit utilizes a pseudo-random set of parameters - emulating assembly errors. "Restoration" is executed by operating the second simulator unit using "exact" setup parameters - as if assembly was ideal. This organization enables to estimate as accuracy as reliability of the "Alignment Instead of Calibration" approach. Simulation results in evaluation of the "true positive", "true negative", "false positive" and "false negative" of the selected setup. Additionally, results of simulations can be organized as a set of 2D and 3D plots enabling to recommend customer-tailored MCMD configuration for the specified volume to be protected.



4. Simulation Results

It is clear that absence of calibration leads to "Fast Positive Error": protected volume was not violated, but "alarm" was raised. One of the goals of the Monte-Carlo simulator was to evaluate "False Positive Rate" (FPR). In the following example FPR was evaluated for the distances AC, BC, OC 5m with 2% tolerance (10 cm assembly error). Camera with 10 mm (10% tolerance) lens having VGA resolution was used. Error in point "O" alignment was set as 10 pixels; error of motion detector was set to 10 pixels. Simulator was operated 5000 runs. In case "dangerous zone margin" was set to 1cm, FPR value was 0.42 (that is in 2088 of 5000 cases motion detector generated false "alarm" – which is unacceptable). However, for the "dangerous zone margin" 10 cm, FPR was 0 (for all 5000 runs result was correct). For the specific example (see exemplary scene of Fig. 1) 10 cm accuracy can be considered as more than adequate.

Conclusions

Monte-Carlo simulator enables to evaluate accuracy and reliability of the threecamera MCMD, thus, eliminating the need for the tedious field tests. Accuracy and reliability of the three-camera MCMD in the selected exemplary configuration was estimated as adequate for the exemplary practical scene presented at Fig. 1. Usage of modern high-resolution cameras will additionally increase accuracy and reliability of MCMD.

Acknowledgment

This work was supported by a grant from research committee of ORT Braude Academic College of Engineering, Karmiel, Israel

References

[1] L.R. Kasturi, R. Schnuck, "Machine Vision", McGraw-Hill Inc., 289-297, 1995.

[2] Hee-Sung Kim., G. Kurillo G, R. Bajesy, "Hand Tracking and Motion Detection from the Sequence of Stereo Color Image Frames", IEEE International Conference of Industrial Technology, 1-6, 2003.

[3] S. Kosolapov., A. Lomes, "Dead-Zone and 3D Accuracy Evaluation of Modified Two-Camera Stereo Setup by using Monte-Carlo Simulation". Proceedings ASMDA 2011, 744, 2011.

[4] S. Kosolapov, Reliability Evaluation of 3D Motion Detector by using Monte-Carlo Simulation, Proceeding SMTDA 2012, 419-423, 2012.

[5] S. Kawabata. "Intrusion Detection System using Contour-based Multi-Planar Visual Hull Method", Doctoral work, 2003.



BPSO Versions with Chi-squared Distribution for MKP Resolution

Raida KTARI^{1,2}, Habib CHABCHOUB¹

¹ Sfax University, L.O.G.I.Q, Tunisia
 ² Aix-Marseille Université, CNRS LIF UMR 7279, France
 (E-mail : raidaktari@yahoo.com, habib.chabchoub@fsegs.rnu.tn)

Abstract. This paper deals with the application of BPSO (Binary Particle Swarm Optimization), EPSO (Essential Particle Swarm Optimization) and EPSOq (Essential Particle Swarm Optimization queen) to the Multidimensional Knapsack Problem (MKP) which is well-known to be NP-hard Combinatorial Optimization problem. The particularity of this paper consists in proposing a novel random number generator based on the Chi-squared distribution for the particles' positions and velocities in the initialization step. A repair operator is also utilized to change an unfeasible solution to a feasible one. The performance assessment of the BPSO, EPSO and EPSOq is a critical point in this paper. That is why, we experiment these approaches on a variety of MKP instances from OR-Library. Then, we compare our results with those of other previous works and with the best known results in the literature.

Keywords: MKP, PSO, Chi-squared distribution, randomness, repair operator, performance.

1 Introduction

Particle Swarm Optimization (PSO) is one of the evolutionary optimization methods inspired by nature which include evolutionary strategy (ES), evolutionary programming (EP), genetic algorithm (GA), and genetic programming (GP). PSO was originally designed and introduced by Eberhart and Kennedy [7], [9] in 1995. It is a population based search algorithm based on the simulation of the social behaviour of birds, bees or a school of fishes.

The PSO algorithm is based on the exchange of information between individuals, so called particles, of the population, so called swarm. Indeed, each particle adjusts its own position towards its previous experience and towards the best previous position obtained in the swarm. Memorizing its best own position establishes the particle's experience implying a local search along with global search emerging from the neighboring experience or the experience of the whole swarm.

PSO has gained widespread appeal amongst researchers and has been shown to offer good performance and efficiency in a variety of application domains such as power and voltage control (Abido [1]), mass-spring system (Brandstatter and Baumgartner [4]), and task assignment (Salman et al. [13]). The comprehensive survey of the PSO algorithms and applications can be found in Kennedy and Eberhart [11].

Particle swarm optimization was first introduced as an optimization method for solving continuous problem. Later, Kennedy and Eberhart [10], proposed a

binary version of PSO (BPSO) to accommodate discrete binary variables and allow it to operate in a binary problem space. A particle moves in a search space restricted to 0 or 1 on each dimension.

Since the performance ability of BPSO is not good enough, several improved versions of BPSO have been proposed to modify the velocity vector when binary variables are involved. Recently, Chen et al. [5] proposed a modified version of BPSO, so called Essential Particle Swarm Optimization (EPSO). In this research work, authors dismantle the BPSO algorithm qualitatively, breaking it into its essential components and then reinterpreting it in another ways as new program. After, as pheromone array in Ant Colony Optimization (Alaya et al. [2]), they introduce the queen informant particle in EPSO. They identify this implementation of this idea with the acronym EPSOq that is considered as another improved version of BPSO.

This paper deals with the application of BPSO, EPSO and EPSOq in the field of Combinatorial Optimization (CO) problems, which is a quite rare field tackled by PSO. The constrained problem discussed in this paper is the well-known to be NP-hard CO problem. The 0-1 multidimensional knapsack problem (MKP), which consists in selecting a subset of n given objects (or items) in such a way that the total profit of the selected objects is maximized while a set of knapsack constraints are satisfied. More formally, the MKP01 can be stated as follows:

$$Max \sum_{j=1}^{n} c_j x_j$$
 (1)

(MKP)
$$\begin{cases} \text{s.s} \sum_{j=1}^{m} a_{ij} x_j \leq b_i ; \forall i \in M = \{1, ..., m\} \\ x_i \in \{0, 1\} \forall j \in N = \{1, ..., n\} \end{cases}$$
 (2)

Equation (1) describes the objective function for the MKP. Each of the m constraints described in condition (2) is called a knapsack constraint, so the MKP is also called the m-dimensional knapsack problem. Let $M = \{1, 2, ..., m\}$ and $N = \{1, 2, ..., n\}$, with $b_i > 0$ for all $i \in M$ and $a_{ij} \ge 0$ for all $i \in M$, $j \in N$, a

and N = {1, 2, ..., n}, with $\mathbf{b}_i > 0$ for all $i \in M$ and $\mathbf{a}_{ij} \ge 0$ for all $i \in M$, $j \in N$, a well-stated MKP assumes that $\mathbf{c}_j > 0$ and $\mathbf{a}_{ij} \le \mathbf{b}_i < \sum_{j=1}^n \mathbf{a}_{ij}$ for all $i \in M$, $j \in N$.

MKP is one of the most intensively studied discrete programming problems, mainly because its simple structure which can be seen as a general model for any kind of binary problems with positive coefficients.

In this paper, we propose an algorithm which applies the BPSO, EPSO and EPSOq approaches for solving the 0-1MKP. This is achieved by using a generator based on the Chi-squared distribution in order to generate particles well-varied in the research space during the initialization phase.

This algorithm has also the advantage of using a repair operator based on the pseudo-utility ratios derived from the surrogate duality approach to guarantee generating feasible solutions.

The performance assessment of the BPSO, EPSO and EPSOq is a critical point in this paper. That is why, we experiment these approaches on a very large variety of bigger size MKP01 instances from OR-Library, which are considered



to be rather difficult for optimization approaches. Then, we compare our results with the best known ones in the literature (Angelelli et al. [3], Chu and Beasley [6], Vasquez and Vimont [14]).

2 BPSO Methods for MKP Resolution

Since we have found any literature concerning neither the EPSO nor EPSOq algorithm applied to the 0-1 MKP problems, we select some bigger size MKP instances from OR-Library to assess their performance, and we compare their results not only with that of the BPSO but also with the best known profits in the literature. To make a fair comparison, we set the same ring topology as the neighborhood structure with number of neighbors set to 2 for the three PSO improved versions. The parameters given in (Chen et al. [5]) are regarded as optimal for our algorithm. That is why; we conserve their values even for BPSO. We also use the same initialization generator for the particles position and the same reparation method for the unfeasible solutions. These methods are described as follows:

Initialization method: Monte Carlo methods use the computer together with the generation of random numbers and mathematical models to generate statistical results that be able to simulate and experiment with the behavior of various business, engineering and scientific systems. Monte Carlo simulations usually use the application of random numbers that are uniformly distributed over the interval [0, 1]. These uniformly distributed random numbers are employed in order to obtain stochastic variables from various probability distributions. These stochastic variables can then be useful to approximate the behavior of worthwhile system variables.

Indeed, most computer languages have form of random number generator that generates uniformly distributed random numbers between 0 and 1. Almost these random number generators use modulo-arithmetic so as to generate numbers that appear to be uniformly distributed. As a result, the random number generators are called pseudorandom number generators since they are not really random. They only simulate the behavior of a uniformly distributed random number on the interval [0, 1].

As an example, we apply in the present paper the chi-square goodness of fit test to the random number generator associated with JAVA our computer programming language in order to initialize the particles positions. We try to build a computer program to generate, for instance, 1000 random numbers between 0 and 1. We can then divided the interval 0 to 1 into 10 classes using the intervals (0, .1), (.1, .2), ..., (.9, 1.0) and then we sort the 1000 random numbers to precise the number in each class. These values are the observed frequencies designed by the experiment. If the pseudorandom number generator is truly uniform, then the theoretical frequency associated with each class would have a value of 100.

So, for each position bit, a number will be generated thanks to this approach and consequently, it will be rounded either 0 or 1. This method would be very helpful in creating a swarm whose particles are well-scattered in the search



space and even quick not only for the particles positions initialization step but also for the particles velocities start up.

Reparation method: Usually, in PSO, a given particle is dynamically attracted by the social and the cognitive components. Therefore, during evolution towards the best global solution, the algorithm can pass through regions of unfeasible solutions.

Indeed, an unfeasible particle now can be changed to a feasible one later. A particle representing an unfeasible solution is allowed to exist in the swarm. PSO usually makes use of penalty function technique in order to reduce the constrained problem to an unconstrained problem by imposing a penalty to the fitness of such particle (Hu and Eberhart [8]).

Instead of this methodology, we incorporate the heuristic repair operator suggested in (Kong and Tian [12]) specially designed for MKP. The repair operator utilizes the notion of the pseudo-utility ratios u_j that is explained in the following equation:

$$\boldsymbol{u}_{j} = \frac{c_{j}}{\sum_{i=1}^{m} w_{i} a_{ij}}$$
(4)

where $w=(w_1, w_2, ..., w_m)$ is a set of surrogate multipliers (or weights) of some positive real numbers. To obtain reasonably good surrogate weights we can solve the LP relaxation of the original MKP and utilize the values of the dual variables as the weights. Otherwise, w_i is set equal to the shadow price of the *i*th constraint in the LP relaxation of the MKP.

A brief description behind this reparation method is given as follows: The first phase, which is called DROP phase, changes the value of each bit of the solution from one to zero in increasing order of u_j if feasibility is violated. The second phase, so called ADD phase, reverses the process by changing each bit from zero to one in decreasing order of u_j as long as feasibility is not violated. Adopting this procedure, we are able to solve the problem of an unfeasible particle existence in the search space.

3 Computational Results and Discussions

Combining all the ideas described above, we aggregate the BPSO, EPSO and EPSOq in a unique algorithm applied on a variety large MKP instances with maximum number of cycles of n, where n is the objects number. Indeed, we concentrated in MKP instances with $m \in \{5, 10, 30\}$ constraints, $n \in \{100, 250, 500\}$ variables and $\alpha \in \{0.25, 0.5, 0.75\}$ tightness ratios. Experiments were performed in an Intel® Core TM i5 with 4 Gb of RAM and 2.67 GHz CPU. The following table reports just the test results of some instances.

The first column indicates the instance name, the second column is the bestknown solutions and the next 3 columns record the fitness average and its CPU time average in seconds of BPSO, EPSO and EPSOq respectively over 30 runs.

Instance	Best-known	BPSO		EPSO		EPSQq.	
		Fitness	CPU time	Fitness	CPU time	Fitness	CPU time
OR5x100-0.25_1	24381	21774	0.012	24071.5	0.0033	24381	0.0029
OR5x100-0.50_2	42545	41870.5	0.15	42320	0.009	42545.5	0.0099
OR5x100-0.75_3	59802	44294	0.67	59494.5	0.094	58889	0.093
OR10x250-0.25_4	61000	57550	9	60369	0.53	60662.5	0.52
OR10x250-0.50_5	108485	106255	6.23	107314	0.792	107314	0.69
OR10x250-0.75_6	152109	151619	4.012	151417.5	0.871	152109	0.792
OR30x500-0.25_7	114181	109086	327.004	112112	11.91	113895	11021
OR30x500-0.50_8	216542	212662.5	221.95	212088.5	8.58	214150.5	7.95
OR30x500-0.75_9	303199	298667	123.01	300267	7.023	301447.5	6.4

Table.1: Comparison of results obtained by BPSO, EPSO, EPSOq and the optimal known solutions

From this table and for all the 270 instances of MKP that we tested and which are considered to be rather difficult for optimization approaches, Our experiments show that BPSO, EPSO and EPSOq are able to find good solutions. But, it is obvious that EPSOq outperforms BPSO and even EPSO with better solution quality, and with quick convergence to satisfied solution, as the size of the problem increases.

In fact, incorporating the queen informant in the EPSO doesn't increase the number of function evaluations because it was added just a new informer that only offers information to the other particles. Moreover, the initialization generator based on the Chi-squared distribution achieves a higher exploration of solutions at the start of algorithm and the repair operator plays a critical role in a higher exploitation near the global optimum solutions at the end of algorithm.

We also compare the EPSOq fitness with the best known solutions in the literature. Indeed, the average performance of EPSOq for the 270 instances considered was 1.028% of the known optimum. This reveals that PSO can perform well for this class of combinatorial problem, even for large instances. That is, EPSOq seems to be efficient in navigating the hyper-surface of the search space and finding good solutions (and, sometimes, the best solution). It is also as competitive as the existing discrete PSO approaches for the MKP thanks to the initialization generator, the repair operator and the queen informant.

4 Conclusions

This paper presents the application of the BPSO, EPSO and EPSOq on the 0-1 multidimensional knapsack problem.

The particles positions are initialized using a random number generator based on the Chi-squared distribution in the aim to obtain a swarm well-varied. To tackle the problem of the movement to regions at very large distances from the main swarm or even outside the search space, we implement the reparation method using the notion of the pseudo-utility ratios. So as to evaluate the effectiveness and viability of these discrete binary PSO versions, we take many large MKP instances from OR-Library to test its performance. Indeed, its simulation results are compared with the best known solutions. The EPSOq results were close and sometimes equal to the optimal solution known, even considering that the parameters were not optimized. These experiments were not optimal. Little effort was taken to find the best parameters. In future versions, we will do our best in these aspects. Eventually, the EPSOq is seemed as a new path for building a fast and easy discrete PSO by its smart representation.

References

- 1.M. A. Abido. Optimal power flow using particle swarm optimization. *Electrical Power* and Energy Systems, 24, 563-571, 2002.
- 2.I. Alaya, C. Solnon and K. Ghédira. Optimisation par colonies de fourmis pour le problème du sac à dos multidimensionnel. *TSI-Technique et Science Informatiques*, 45-60, 2007.
- 3.E. Angelelli, M. G. Speranza and M. W. P. Savelsbergh. Competitive analysis for dynamic multiperiod uncapacitated routing problems. *Networks*, 49, 4, 308–317, 2007.
- 4.B. Brandstatter and U. Baumgartner. Particle Swarm Optimization—Mass-Spring System Analogon. *IEEE Transactions on Magnetics*, 38, 997-1000, 2002.
- 5.E. Chen, J. Li and X. Liuc. In search of the essential binary discrete particle swarm. *Applied Soft Computing*, 11, 3260–3269, 2011.
- 6.P. Chu and J. Beasley. A genetic algorithm for the multidimensional knapsack problem. *Journal of Heuristics*, 4, 63-86, 1998.
- 7.R. C. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. *Proceedings of the 6th International Symposium on Micro Machine and Human Science*, 39-43, 1995.
- 8.X. Hu and R. Eberhart. Solving Constrained Nonlinear Optimization Problems with Particle Swarm Optimization. *Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics.* 2002.
- 9.J. Kennedy and R.C. Eberhart. Particle Swarm Optimisation. *Proceedings of the IEEE International Conference*, 4, 1942- 1948, 1995.
- 10.J. Kennedy and R.C. Eberhart. A discrete binary version of the particle swarm algorithm. *IEEE International Conference on Systems, Man, and Cybernetics*, 5, 4104-4109, 1997.
- 11.J. Kennedy and R. Eberhart. Swarm Intelligence. *Morgan Kaufmann Publishers, Inc., San Francisco, CA*, 2001.
- Kong and P. Tian. Apply the Particle Swarm Optimization to the multidimensional knapsack. L. ICAISC 2006, LNAI 4029, 1140–1149, 2006.
- 13.A. Salman, I. Ahmad and S. Al-Madani. Particle Swarm Optimization for Task Assignment Problem. *Microprocessors and Microsystems*, *26*, 363-371, 2002.
- 14.M. Vasquez and Y. Vimont. Improved results on the 0-1 multidimensional knapsack problem. *European Journal of Operational Research*, 165, 70–81, 2005.



Generalised Asymmetric Linnik distributions and process

Kuttykrishnan.A.P., Associate Professor, Department of Statistics, Sir Syed College, Taliparamba, Kerala, India-670142.

Abstract

In recent years there has been an increasing interest in developing the theory and applications of geometric stable distributions. The class of geometric stable distributions is a four-parameter family of distributions denoted by $GS_{\alpha}(\sigma,\beta,\mu)$ and conveniently described in terms of characteristic function

$$\Phi(t) = \frac{1}{1 + \sigma^{\alpha} |t|^{\alpha} \varpi_{\alpha,\beta}(t) - i\mu t} \text{ where } \varpi_{\alpha,\beta}(t) = \begin{cases} 1 - i\beta \operatorname{sign}(t) \tan(\pi\alpha/2) & \text{if } \alpha \neq 1 \\ 1 + i\beta \frac{2}{\pi} \operatorname{sign}(t) \log|t| & \text{if } \alpha = 1. \end{cases}$$

The parameter $\alpha \in (0,2]$ is the index of stability and determines the tail of the distribution. These classes of distributions arise as a limiting distribution of geometric random sums of independent and identically distributed random variables. Since the geometric random sums frequently appear in many applied problems in various areas, the geometric stable distributions have wide variety of applications especially in the field of reliability, biology, economics, financial mathematics etc. When $\beta = 0, \mu = 0$, the geometric stable distribution

have the characteristic function
$$\Phi(t) = \frac{1}{1 + \sigma^{\alpha} |t|^{\alpha}}$$
, and the corresponding distribution is

called Linnik distribution and is named after Ju.V.Linnik, who showed that the above function is a bona fide characteristic function of a symmetric distribution for any $0 < \alpha \le 2$. It may be noted that probability density and distribution functions of the Linnik random variable are not in closed form except for $\alpha = 2$, which corresponds to the Laplace



distribution. The Laplace distribution is symmetric, and there were several asymmetric extensions in generalizing the Laplace distribution.

In this paper we introduce and study new classes of distributions, namely Pakes generalized asymmetric Linnik distribution and geometric Pakes generalized asymmetric Linnik distribution. First order autoregressive process with Geometric Pakes generalized asymmetric Linnik distribution as marginal distribution is developed. Higher order extensions are discussed. A bivariate distribution related to geometric Pakes asymmetric Laplace and Linnik distribution is introduced and bivariate time series model corresponding to this distribution is developed.

Key words: Autoregressive process, Geometric infinite divisibility, Geometric exponential distribution, Geometric marginal asymmetric Laplace and Linnik distribution, Geometric Pakes generalized asymmetric Linnik distribution, Geometric Stable distribution.

1. Introduction

In recent years there has been an increasing interest in developing the theory and applications of geometric stable distributions. The class of geometric stable distributions is a four-parameter family of distributions denoted by $GS_{\alpha}(\sigma,\beta,\mu)$ and conveniently described in terms of characteristic function

$$\Phi(t) = \frac{1}{1 + \sigma^{\alpha} |t|^{\alpha} \varpi_{\alpha,\beta}(t) - i\mu t} \text{ where } \varpi_{\alpha,\beta}(t) = \begin{cases} 1 - i\beta \operatorname{sign}(t) \tan(\pi\alpha/2) & \text{if } \alpha \neq 1 \\ 1 + i\beta \frac{2}{\pi} \operatorname{sign}(t) \log|t| & \text{if } \alpha = 1. \end{cases}$$

The parameter $\alpha \in (0,2]$ is the index of stability and determines the tail of the distribution. These classes of distributions arise as a limiting distribution of geometric random sums of independent and identically distributed random variables. Since the geometric random sums frequently appear in many applied problems in various areas (see Gnedenko and



Korolev (1996)), the geometric stable distributions have wide variety of applications especially in the field of reliability, biology, economics, financial mathematics etc.

When $\beta = 0, \mu = 0$, the geometric stable distribution have the characteristic function

$$\Phi(t) = \frac{1}{1 + \sigma^{\alpha} |t|^{\alpha}}$$
, and the corresponding distribution is called Linnik distribution and is

named after Ju.V.Linnik, who showed that the above function is a bona fide characteristic function of a symmetric distribution for any $0 < \alpha \le 2$. The probability density function of

the Linnik random variable when
$$\sigma=1$$
 has the

representation
$$f_{\alpha}(x) = \frac{\sin \frac{\pi \alpha}{2}}{\pi} \int_{0}^{\infty} \frac{\upsilon^{\alpha} \exp(-\upsilon |x|)}{1 + \upsilon^{2\alpha} + 2\upsilon^{\alpha} \cos \frac{\pi \alpha}{2}} d\upsilon$$
 for $x > 0$ and for $x < 0$,

 $f_{\alpha}(x) = f_{\alpha}(-x)$. It may be noted that probability density and distribution functions of the Linnik random variable are not in closed form except for $\alpha = 2$, which corresponds to the Laplace distribution. The Laplace distribution is symmetric, and there were several asymmetric extensions in generalizing the Laplace distribution. Kozubowski and Podgórski (2000) studied asymmetric Laplace (AL) distribution with characteristic function $\Phi(t) = \frac{1}{2} + \frac{1$

$$\Phi(t) = \frac{1}{1 + \sigma^2 t^2 - i\mu t}, \quad \omega < \mu < \omega, 0 \ge 0, \text{ and discussed applications of it in the fields}$$

financial mathematics. Kotz *et al.* (2001) discussed a new class of distributions, namely generalized asymmetric Laplace distributions, with characteristic function

$$\Phi(t) = \left(\frac{1}{1 + \sigma^2 t^2 - i\mu t}\right)^{\tau}, -\infty < \mu < \infty, \sigma, \tau \ge 0. \text{ By specifying } \tau = 1, \sigma = 0 \text{ and } \mu > 0, \text{ we}$$

have an exponential distribution with mean μ and obtain symmetric Laplace distribution if



 $\tau = 1, \mu = 0$ and $\sigma \neq 0$. When $\sigma = 0$, the function reduced to characteristic function of a gamma variable with the scale parameter μ and the shape parameter τ .

In this paper we introduce and study new classes of distributions, namely Pakes generalized asymmetric Linnik distribution and geometric Pakes generalized asymmetric Linnik distribution. In Section 2 we introduce Pakes generalized asymmetric Linnik distribution and obtain the representation of the random variable. In Section 3 we introduce and study geometric Pakes generalized asymmetric Linnik distribution. Time series model equivalent to TEAR (1) model discussed in Lawrance and Lewis (1981) with geometric Pakes generalized asymmetric Linnik distribution is introduced and studied in Section 4.

2. Pakes generalized asymmetric Linnik distribution

Pakes (1998) generalized the Linnik distribution and introduced a symmetric distribution, namely generalized Linnik distribution with characteristic function

$$\Phi(t) = \left(\frac{1}{1 + \sigma^{\alpha} |t|^{\alpha}}\right)^{\tau}, \ \sigma, \tau \ge 0, 0 < \alpha \le 2. \text{ It may be noted that when } \alpha = 2 \text{ this reduces to}$$

the characteristic function of generalized Laplace distribution of Mathai (1993). Similar to generalized asymmetric Laplace distribution we can define an asymmetric distribution with characteristic function

$$\Phi(t) = \left(\frac{1}{1 + \sigma^{\alpha} |t|^{\alpha} - i\mu t}\right)^{\tau}, -\infty < \mu < \infty, \sigma, \tau \ge 0, 0 < \alpha \le 2.$$
(2.1)

We shall refer this distribution as the Pakes generalized asymmetric Linnik distribution and denoted by $PGAL_{\alpha}(\mu, \sigma, \tau)$.


When $\alpha = 2, \tau = 1$, it reduces to the asymmetric Laplace distribution of Kozubowski and Podgórski (2000).

Theorem 2.1 A PGAL_{α}(μ, σ, τ) random variable X with characteristic function (2.1) admits the representation $X \underline{d} \mu W + \sigma W^{1/\alpha} Z$, where Z is symmetric stable with characteristic function $\Psi(t) = \exp(-\sigma^{\alpha} |t|^{\alpha})$ and W is a gamma random variable with probability density function $g(w) = \frac{1}{\Gamma(\tau)} w^{\tau-1} e^{-w}$, $w > 0, \tau > 0$, independent of Z.

Proof:

Conditioning on $\,W$, we obtain the characteristic function $\,\Phi(t)\,$ of $\,\mu\,W + \sigma\,W^{1/\alpha}Z$

as

$$\Phi(t) = E(E(e^{it(\mu W + \sigma W^{1/\alpha} Z} / W))$$
$$= \int_{0}^{\infty} E(e^{it(\mu w + \sigma w^{1/\alpha} Z})g(w)dw$$
$$= \frac{1}{\Gamma(\tau)}\int_{0}^{\infty} w^{\tau - 1}e^{-w(1 + \sigma^{\alpha} |t|^{\alpha} - i\mu t)}dw$$
$$= \left(\frac{1}{1 + \sigma^{\alpha} |t|^{\alpha} - i\mu t}\right)^{\tau}.$$

Hence the theorem.

3. Geometric Pakes generalized asymmetric Linnik distribution

Pillai (1990) introduced geometric exponential distribution and studied the properties of the renewal process with geometric exponential waiting time distribution. Jose and Seetha Lekshmi (1999,2003), studied geometric gamma and geometric Laplace distributions and developed autoregressive time series models. Jayakumar and Ajitha



(2003) introduced geometric Mittag-Leffler distribution and studied its properties including infinite divisibility and attraction to stable laws. Seetha Lekshmi and Jose (2006) introduced and studied geometric Pakes generalized Linnik distribution and developed time series model using this distribution.

Since the distribution with characteristic function (2.1) is infinitely divisible, using the result of Klebanov *et al.* (1984), we can define a geometrically infinitely divisible distribution with characteristic function $\Psi(t)$ such that $\Phi(t) = \exp\left\{1 - \frac{1}{\Psi(t)}\right\}$.

The characteristic function (2.1) can be written as

$$\left(\frac{1}{1+\sigma^{\alpha}|t|^{\alpha}-i\mu t}\right)^{\tau} = \exp\left\{1-\frac{1}{\left(1+\tau\log\left(1+\sigma^{\alpha}|t|^{\alpha}-i\mu t\right)\right)^{-1}}\right\}.$$

Hence $\Psi(t) = \frac{1}{1 + \tau \log(1 + \sigma^{\alpha} |t|^{\alpha} - i\mu t)}$ is a characteristic function of a geometrically

infinitely divisible distribution.

A distribution with characteristic function

$$\Psi(t) = \frac{1}{1 + \tau \log\left(1 + \sigma^{\alpha} \left| t \right|^{\alpha} - i\mu t \right)}, -\infty < \mu < \infty, \sigma, \tau \ge 0, 0 < \alpha \le 2,$$
(3.1)

is called geometric Pakes generalized asymmetric Linnik (GPGAL) distribution with parameters μ, σ, α and τ .

If X is a random variable with characteristic function (3.1), we represent it as $X \underline{d} GPGAL_{\alpha}(\mu, \sigma, \tau)$. It may be noted that when $\tau = 1$ in (3.1), the corresponding distribution is the geometric version of asymmetric Linnik distribution and in such case we call it as geometric asymmetric Linnik distribution ($GPGAL_{\alpha}(\mu, \sigma, \tau = 1)$).

Now we consider the asymptotic behavior of the $\operatorname{GPGAL}_{\alpha}(\mu,\sigma,\tau)$ distribution.



Theorem 3.1. The GPGAL_{α}(μ, σ, τ) distribution is the limit distribution of geometric sums of PGAL_{α}(μ, σ, τ) random variables.

Proof:

Let $\Phi(t)$ be the characteristic function of $PGAL_{\alpha}(\mu,\sigma,\frac{\tau}{n})$ random variable.

Then
$$\Phi(t) = \left(\frac{1}{1+\sigma^{\alpha}|t|^{\alpha}-i\mu t}\right)^{\frac{\tau}{n}}$$
.

Define $\Theta(t) = \frac{1}{\Phi(t)} - 1 = (1 + \sigma^{\alpha} |t|^{\alpha} - i\mu t)^{\tau/n} - 1.$

Hence, using Lemma 3.2 of Pillai (1990), $\Phi_n(t) = \frac{1}{1 + p \Theta(t)}$, where p > 1 is the

characteristic function of a geometric sum of random variables. By choosing p = n, we

have $\Phi_n(t) = \left\{ 1 + n \left[\left(1 + \sigma^{\alpha} \mid t \mid^{\alpha} - i \mu t \right)^{\tau/n} - 1 \right] \right\}^{-1}$. So $\Phi_n(t)$ is the characteristic function of

a geometric sum of $PGAL_{\alpha}(\mu,\sigma,\frac{\tau}{n})$ random variables.

Consider
$$\lim_{n \to \infty} \Phi_n(t) = \frac{1}{1 + \lim_{n \to \infty} n \left[\left(1 + \sigma^{\alpha} \mid t \mid^{\alpha} - i \mu t \right)^{\tau/n} - 1 \right]}$$

 $=\frac{1}{1+\tau \log (1+\sigma ^{\alpha }\left| \,t\,\right| ^{\alpha }-i\,\mu \,t)}\text{, which is the characteristic function of }$

 $GPGAL_{\alpha}(\mu,\sigma,\tau)$ random variables.

Hence $GPGAL_{\alpha}(\mu, \sigma, \tau)$ distribution is the limit distribution of geometric sum of $PGAL_{\alpha}(\mu, \sigma, \frac{\tau}{n})$ random variables. Hence the theorem.



Now we prove stability property of $GPGAL_{\alpha}(\mu,\sigma,\tau)$ random variables with respect geometric summation.

Theorem 3.2 Let $\{X_n\}$ be a sequence of independent and identically distributed random variables and let N_p be a geometric random with mean 1/p. Further, assume that N_p is

independent of the X_i 's .If $U_N = \sum_{i=1}^{N_p} X_i$ then the random variables U_{N_p} and X_i are identically distributed if X_i follows $GPGAL_{\alpha}(\mu, \sigma, \tau)$ distribution.

Proof:

Let $\Phi(t)$ and $\Theta(t)$ be the characteristic functions of $X_{i}^{}$ and $U_{N_{p}^{}}^{}$ respectively.

Then
$$\Theta(t) = \frac{p \Phi(t)}{1 - (1 - p) \Phi(t)}$$
. (3.2)

Suppose $X_i \stackrel{d}{=} GPGAL_{\alpha}(\mu, \sigma, \tau)$, then by (3.2) we have

$$\Theta(t) = \frac{p}{p + \tau \log(1 + \sigma^{\alpha} |t|^{\alpha} - i \mu t)}$$
$$= \frac{1}{1 + \frac{\tau}{p} \log(1 + \sigma^{\alpha} |t|^{\alpha} - i \mu t)}$$

 $\mbox{Hence} \quad U_{N_p} \stackrel{d}{=} GPGAL_{\alpha}(\mu,\sigma,\frac{\tau}{p}) \,.$

Hence the theorem. .

4. Autoregressive model with $GPGAL_{\alpha}(\mu,\sigma,\tau)$ marginal distribution

Time series in which observations are of clearly non-Gaussian nature are very common in many areas. A number of literatures have developed in recent years in modeling time series data with non-Gaussian, and more generally asymmetric marginal distributions. Gaver and Lewis (1980) discussed and studied conventional first order linear autoregressive model $X_n = \rho X_{n-1} + \varepsilon_n$ with exponential marginal distribution.



Subsequently, Lawrance and Lewis (1981), Dewald and Lewis (1985), and Jayakumar and Pillai (1993) developed autoregressive models with different marginal distributions such as Laplace, gamma, Mittag-Leffler distributions. These first order autoregressive models are developed using the self-decomposability property of the corresponding marginal distributions. Now we develop a time series model using GPGAL_{α}(μ , σ , τ) marginal distribution on the basis of geometric infinitely divisible property of the distribution. This model is equivalent to the one-parameter TEAR (1) model discussed in Lawrance and Lewis (1981).

Theorem 4.1

Let
$$\{X_n, n \ge 1\}$$
 be defined as

$$X_n = \begin{cases} \varepsilon_n & \text{w.p. } \theta \\ X_{n-1} + \varepsilon_n & \text{w.p. } 1 - \theta \end{cases}$$
(4.1)

where $0 \le \theta \le 1$ and $\{\epsilon_n\}$ is a sequence of independent and identically distributed random variables. A necessary and sufficient condition that $\{X_n\}$ is a stationary process with $GPGAL_{\alpha}(\mu, \sigma, \tau)$ marginal is that $\{\epsilon_n\}$ is distributed as $GPGAL_{\alpha}(\mu, \sigma, \theta \tau)$.

Proof

Let $\Phi_{X_n}(t)$ be the characteristic function of $\{X_n\}$. Then from (4.1), we get $\Phi_{X_n}(t) = \theta \Phi_{\varepsilon_n}(t) + (1-\theta) \Phi_{X_{n-1}}(t) \Phi_{\varepsilon_n}(t).$ (4.2)

Assuming stationarity, we have

$$\Phi_{X}(t) = \theta \ \Phi_{\varepsilon}(t) + (1-\theta) \ \Phi_{X}(t) \Phi_{\varepsilon}(t).$$

Hence

$$\Phi_{\varepsilon}(t) = \frac{\Phi_{X}(t)}{\theta + (1 - \theta) \Phi_{X}(t)}.$$
(4.3)



Suppose
$$X_n \stackrel{d}{=} GPGAL_{\alpha}(\mu, \sigma, \tau)$$
 then $\Phi_X(t) = \frac{1}{1 + \tau \log(1 + \sigma^{\alpha} |t|^{\alpha} - i\mu t)}$.

Substituting this in (4.3) and simplifying, we get

$$\Phi_{\varepsilon}(t) = \frac{1}{1 + \theta \tau \log(1 + \sigma^{\alpha} |t|^{\alpha} - i\mu t)}$$

Hence $\epsilon_n \stackrel{d}{=} GPGAL_{\alpha}(\mu, \sigma, \theta \tau).$

Conversely, if $\{\epsilon_n\}$ is a sequence of independent and identically distributed $GPGAL_{\alpha}(\mu, \sigma, \theta \tau)$ random variables and $X_0 \stackrel{d}{=} GPGAL_{\alpha}(\mu, \sigma, \tau)$.

Then from (4.2), when n=1, we have
$$\Phi_{X_1}(t) = \frac{1}{1 + \tau \log(1 + \sigma^{\alpha} |t|^{\alpha} - i\mu t)}$$
.

 $\text{If } X_{n-1} \underline{\underline{d}} \, GPGAL_{\alpha}(\mu,\sigma,\tau) \, \, \text{then we get } \, X_n \, \, \underline{\underline{d}} \, \, GPGAL_{\alpha}(\mu,\sigma,\tau) \, .$

Thus, using inductive argument $\left\{X_n\right\}$ is a stationary process with $GPGAL_\alpha(\mu,\sigma,\tau) \text{ marginal distribution}.$

Hence the theorem.

We call the process defined by (4.1) with $X_0 \stackrel{d}{=} GPGAL_{\alpha}(\mu, \sigma, \tau)$ and $\left\{\epsilon_n\right\}$ is a sequence of independent and identically distributed $GPGAL_{\alpha}(\mu, \sigma, \theta \tau)$ random variables, as the first order autoregressive process with $GPGAL_{\alpha}(\mu, \sigma, \tau)$ marginal distribution.

From the definition (4.1) of the model it is easily verified that

$$\Phi_{X_n}(t) = \theta \Phi_{\varepsilon_n}(t) \frac{1 - (1 - \theta)^n \Phi_{\varepsilon_n}^n(t)}{1 - (1 - \theta) \Phi_{\varepsilon_n}(t)} + (1 - \theta)^n \Phi_{X_0}(t) \Phi_{\varepsilon_n}^n(t).$$

When $n \to \infty$, $\Phi_{X_n}(t) = \theta \Phi_{\varepsilon_n}(t) \frac{1}{1 - (1 - \theta) \Phi_{\varepsilon_n}(t)}$.



Let X_0 is distributed arbitrarly and $\{\epsilon_n\}$ is a sequence of independent and identically distributed $\text{GPGAL}_{\alpha}(\mu, \sigma, \theta \tau)$ random variables then the characteristic function

$$\Phi_{X_n}(t) = \frac{1}{1 + \tau \log(1 + \sigma^{\alpha} |t|^{\alpha} - i\mu t)}.$$

Hence if X_0 is distributed arbitrarly, then also the autoregressive process is asymptotically Markovian with $\text{GPGAL}_{\alpha}(\mu, \sigma, \tau)$ marginal distribution.

Now from the joint characteristic function of (X_n, X_{n+1}) of the process it is easily verified that $\Phi_{(X_n, X_{n+1})}(t_1, t_2) \neq \Phi_{(X_n, X_{n+1})}(t_2, t_1)$. Hence the AR (1) process given by (4.1) with GPGAL_{α}(μ, σ, τ) marginal distribution is not time reversible.

5.Conclusion

Most of the data sets in the areas of financial mathematics, reliability, environmental studies etc, often do not follow the normal law but with asymmetric and heavy tailed character. A large number of research journals discussed applications of Laplace and asymmetric Laplace distributions in different fields where data exhibits asymmetric and heavy tailed character. In communication theory, frequently encountered impulsive noise possesses heavy tails, and so Laplace noise has been suggested as a best model. Also it is established that the Laplace distribution is considered as a model for the distribution of speech waves and the distribution is commonly encountered in image and speech compression. Empirical analysis of some important time series data, especially in the field of financial mathematics, environmental studies etc. shows that asymmetric and heavy tailed distributions, related to Laplace distribution, are more suitable for modeling the data. The applications of Laplace distribution in modeling sizes of sand particle, diamonds etc are also well established in many research articles (for more details see Kotz



et al. (2001)). In this paper we examined the distributions related to Laplace and asymmetric Laplace distributions and can be used as an appropriate model in the areas where Laplace and Linnik distributions do not provide a better fit. Much will be learned in future by further analysis related to the applications of this new class of distributions across a range of contexts.

References

Dewald, L.S. and Lewis, P.A.W. (1985) A new Laplace second order autoregressive time series model-NLAR (2), *IEEE Trans.Inform.Theory***31** (5), 645-651.

Gaver, D.P. and Lewis, P.A.W. (1980) First order autoregressive Gamma sequences and point processes, *Adv.Appl.Prob.***12**, 727-745.

Gnedenko, B.V. and Korolev, V.Yu. (1996) *Random summation: Limit Theorems and Application*. CRC Press, Boca Raton.

Jayakumar, K.and Ajitha, B.K (2003) On the geometric Mittag-Leffler Distributions, *Calcutta Statist. Assoc. Bulletin* **54**, 195-208.

Jayakumar, K. and Pillai, R.N. (1993) The first order autoregressive Mittag-Leffler process, *J.Appl.Prob.***30**, *462*-466.

Jose, K.K. and Seetha Lekshmi,V (1999) On geometric exponential distribution and its applications, *J.Indian Statist.Assoc.***37**,51-58.

Jose, K.K. and Seetha Lekshmi,V (2004) An autoregressive process with geometric α -Laplace marginals,*Statist.Papers* **45**,337-350

Klebanov,L.B., Maniya,G.M., and Melamed,I.A. (1984) A problem of Zolotarev and analogs of infinitely divisible and stable distribution in a scheme for summing a random number of random variables. *Theory Probab. Appl.* **29**, 791-794.



Kotz, S. Kozubowski, T.J and Podgórski, K. (2001) *The Laplace distribution and generalizations: A revisit with applications to communications, Economics, Engineering, and finance, Birkhauser, Berlin.*

Kozubowski, T.J., Meerschaert, M. M., Panorska, A.K., and Scheffler, H.P. (2005). Operator geometric stable laws. *J.Multivariate Anal.* **92**, 298-323.

Kozubowski, T.J. and Podgórski, K. (2000). Asymmetric Laplace distributions, *Math.Sci.***25**, 37-46.

Kuttykrishnan,A.P. and Jayakumar, K (2005) Operator geometric stable distributions and processes, Paper presented at the 25th conference of Indian Society for Probability and Statistics and annual meeting of Indian Bayesian society, held at Department of Statistics, Bangalore University, Bangalore,India during 28-30 December 2005.

Lawrance, A.J. and Lewis, P.A.W. (1981). A new autoregressive time series model in exponential variables (NEAR (1)). *Adv.Appl.Prob.***13**, 826-845.

Mathai, A.M. (1993) On non-central generalized Laplacianness of quadratic forms in normal variables. *J. Multivariate Anal.* **45(2)**, 239-246.

Seetaha Lekshmi,V, and Jose, K.K.(2006) Autoregressive processes with Pakes and geometric Pakes generalized Linnik marginals. *Statist.Prob.Letters* **76**,318-326.

537





One dimensional embedding for nonnegative data visualization

Lazhar Labiod and Mohamed Nadif

Department of Computer sciences, LIPADE, University of Paris 5, 45 rue des Saints Pères 75006 Paris, France (E-mail: 1.labiod@yahoo.fr) (E-mail: mohamed.nadif@parisdescartes.fr)

Abstract. Data visualization has attracted more attention this last decade as a powerful tool for a better understanding of data. In this paper we propose a new theoretical framework for data visualization, this framework is based on Rank-one SVD looking up an appropriate leading left and right singular vectors of a rows or/and columns normalized m by n adjacency matrix A. This involved computing of a truncated rank one singular value decomposition of a suitable normalized data matrix, constructing an one dimensional embedding for both rows and columns data. The visualization of A consists in a simple permutation of rows and columns data according the sorted first left and right singular vectors, which involves an optimal data reorganization revealing homogeneous blocks. Finally, we link our approach to spectral co-clustering and show its usefulness in the context of co-clustering.

Keywords: Data visualization, Stochastic data, Power method, SVD, Co-clustering.

1 Introduction

Data visualization has attracted more attention this last decade as a powerful tool for a better understanding of the data. Prominent authors in the discipline of information visualization [1] have identified that the data mining community gives minimal attention to information visualization, but believe that there are hopeful signs that the narrow bridge between data mining and information visualization will be expanded in the coming years. Bertin [7] has described the visualization procedure as *simplifying without destroying* and was convinced that simplification was 'no more than regrouping similar things'. Spath [3] considered such matrix permutation approaches to have a great advantage in contrast to the cluster algorithms, because no information of any kind is lost, and because the number of clusters does not have to be presumed, it is easily and naturally visible. Murtagh [4], Arabie and Hubert [6,5] have referred to similar advantages calling such an approach a non-destructive data analysis, emphasizing the essential property that no transformation or reduction of the data itself takes place.

In certain problems it may be useful to perform co-clustering, where both objects and features are assigned to groups simultaneously. One approach to the co-clustering problem is to view it as the task of partitioning a weighted bipartite graph. Dhillon [8] proposed a spectral approach to approximate the optimal normalised cut of a bipartite graph, which was applied for document clustering. This involved computing a truncated singular value decomposition (SVD) of a suitably normalised term-document matrix, constructing an embedding of both terms and documents, and applying k-means to this embedding to produce a simultaneous k-way partitioning of both documents and terms. Finally, data visualization is obtained by reorganizing rows and columns data according to the co-clustering result. Despite the advantages of co-clustering, all methods require the knowledge of the number of blocks, in this paper we will not tackle co-clustering but we will see how an appropriate visualization of data involves a reorganization into homogeneous blocks, and we will show the usefulness of our approach in the context of co-clustering. We propose a new theoretical framework, specifically we develop an efficient iterative procedure to find a one dimensional embedding of both rows and columns data. This involved an optimal simultaneous rows and columns data reordering. We show that the solution is given by the leading left and right singular vectors of data matrix.

The rest of paper is organized as follows. Section 2 introduces the problem formulation and the aims. Section 3 is devoted to the proposed algorithm for data visualization. In Section 5, we discuss the relationship between our approach and spectral co-clustering. Section 5 presents numerical experiments on real and simulated data. Finally, the conclusion summarizes the advantages of our contribution.

2 Problem formulation

An interesting connection between data matrices and graph theory can be established. Let A be a $m \times n$ data matrix, it can be viewed as a weighted bipartite graph G = (V, E), where V is the set of vertices and E is the set of edges, it is said to be bipartite if its vertices can be partitioned into two sets I and J such that every edge in E has exactly one end in I and the other in J: $V = I \cup J$. The data matrix A can be viewed as a weighted bipartite graph where each node i in I corresponds to a row and each node j in J corresponds to a column. The edge between i and j has weight a_{ij} , denoting the element of the matrix in the intersection between row i and columns j, for convenience of discussion we also call the vertices in I as the the documents (rows) while vertices in J as words (columns). The adjacency matrix of a bipartite graph is:

$$\mathbf{B} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix},\tag{1}$$

from which we define a stochastic data matrix as follow,

$$\mathbf{S} = \mathbf{D}^{-1}\mathbf{B} = \begin{bmatrix} 0 & D_r^{-1}A \\ D_c^{-1}A^T & 0 \end{bmatrix} \text{ where } \mathbf{D} = \begin{bmatrix} D_r & 0 \\ 0 & D_c \end{bmatrix},$$
(2)

where D_r and D_c , are diagonal matrices such that $D_r = diag(A1)$ and $D_c = diag(A^T 1)$. Let us focus our attention on the first leading eigenvectors of **S**, since **S** is nonnegative and stochastic, The Perron-Frobenius theorem tel us that the first vector is also nonnegative and constant. The power method is



the well known technique used to compute the leading eigenvector of S. The power method consists in the following iterative process:

$$\pi^{(t)} = \mathbf{S}^{(t)} \pi^{(0)} \text{ and } \pi^{(t)} = \frac{\pi^{(t)}}{||\pi^{(t)}||}$$
(3)

By Perron-Frobenius theorem [9] all eigenvalues are real and belong to [-1, 1]. Since **S** is stochastic, it is known that for every right eigenvector there is a corresponding left eigenvector that corresponds to the same eigenvalue $\lambda_1 = 1$; the greatest eigenvalue called the Perron root. The right eigenvector corresponding to the uniform distribution $(\frac{1}{m+n}, ..., \frac{1}{m+n}, ..., \frac{1}{m+n})^T$. The corresponding left eigenvector $\pi = 1$ represents the constant left eigenvector of **S** so that $\pi^T 1 = m + n$. In the matrix notation we have $\pi = \mathbf{S}\pi$ and $\mathbf{S}1 = 1$.

At first sight, this process might seem uninteresting since it eventually leads to a vector with all rows and columns coincide for any starting vector. However our practical experience shows that, first the vector π very quickly collaps into rows and columns blocks and these blocks move towards each other relatively slowly. If we stop the power method iteration at this point, the algorithm would have a potential application for data reordering. The structure of $\pi^{(t)}$ during short-run stabilization makes the discovery of data ordering straightforward. The key is to look for values of $\pi^{(t)}$ that are approximately equal and reordering data accordingly.

3 Rank one SVD algorithm for data visualization

3.1 SVD algorithm

Singular value decomposition (SVD) is a widely-used multivariate data analysis technique [11] which has many potential applications in machine learning, document clustering, pattern recognition, and signal processing. Successful applications of SVD is due to its ability to learn parts-based representation through a matrix decomposition. Given a data matrix A of size $m \times n$, SVD algorithms seek to find factors U, A and V such that

$$A = U_{m \times g} \Lambda_{g \times g} V_{n \times g}^T \quad \text{where} \quad U^T U = I_g \text{ and } V^T V = I_g. \tag{4}$$

The steps of Jordan's SVD are algorithm are described in [11].

3.2 Rank one SVD algorithm

Now, let us consider $\pi = \begin{bmatrix} u \\ v \end{bmatrix}$, where $u \in \mathbf{R}^m_+$ and $v \in \mathbf{R}^n_+$. The upper part of π .i.e u is for documents weight and the lower part u is for words weights. Exploiting now the diagonal structure of \mathbf{S} , then we can write

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 & D_r^{-1}A \\ D_c^{-1}A^T & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \Leftrightarrow \begin{cases} u = D_r^{-1}Av & (a) \\ v = D_c^{-1}A^Tu & (b) \end{cases}$$
(5)



For numerical computation of the leading singular vectors, we use rank one SVD algorithm noted R1SVD, which is a variation of the power method adapted to rectangular data matrix. This iterative process starts with arbitrary vector u^0 and repeatedly performs the updates of v and u by alternating between formulas (a) and (b) given in equation 5 until convergence.

Algorithm 1 : R1SVD

0
Input: data $A \in \mathbf{R}^{m \times n}_+$, D_r and D_c
Output: u, λ, v
Initialize: $\tilde{u} = D_r^{-1} A \mathbb{1}, u = \frac{\tilde{u}}{ \tilde{u} }$
repeat
$\tilde{v}^{(t+1)} = D_c^{-1} A^T u^{(t)}$
$v^{(t+1)} = \frac{\tilde{v}^{(t+1)}}{ \tilde{z}^{(t+1)} }$
$\tilde{u}^{(t+1)} = D_r^{(t)} A v^{(t)}$
$u^{(t+1)} = rac{ ilde{u}^{(t+1)}}{ ilde{u}^{(t+1)} }$
$\gamma^{(t+1)} \leftarrow u^{(t+1)'} - u^{(t)} + v^{(t+1)} - v^{(t)} $
until stabilization of $u, v, \gamma^{(t+1)} - \gamma^{(t)} \leq threshold$

Documents and words weight are collected by u, v respectively, the corresponding component value of u and u give document and word weights, respectively. We can sort word and document in decreasing (or increasing) order of theirs weights and reorganize data matrix accordingly to reveal the homogeneous blocks structure of A. We have developed a mutually reinforcing optimization procedure to exploit duality between documents set and words set, if a word is shared by many documents associated with a block, then word has a high weight associated with the block. On the other hand, if a document is shared by many words associated with a block, then the document has a high weight associated with the same block.

3.3 Relationship with spectral co-clustering

R1SVD is related to spectral co-clustering in that it finds a low dimensional embedding of data, and k-means or another clustering techniques is used to produce the final co-clustering. But as a result, in this paper it is not necessary to find any singular vector (as most co-clustering methods do), in order to find a low dimensional embedding for co-clustering, the embedding just needs to be a good linear combination of left singular vectors and of the right singular vectors respectively. In this respect R1SVD is very different approach from.

In spectral co-clustering the embedding is formed by the bottom left and right eigenvectors of a normalized data matrix [8]. In R1SVD, embedding is defined as weighted linear combination of singular vectors, then u is a defined as linear combination of all left singular vectors of $A_r = D_r^{-1}A$ and v as weighted linear combination of all right singular vectors of $A_c = AD_c^{-1}$. The left and right embedding turn out to be very interesting for data reordering and coclustering.



From the start, the first largest left of A_r and right singular vectors of A_c are not very interesting since they move towards the uniform distribution via a long run times. However the intermediate u and v obtained by R1SVD after a short run time are very interesting. The experimental observation suggests that an effective reordering might run R1SVD for a small number of iterations.

Let us define the data matrix ${\bf M}$ as follow

$$\mathbf{M} = \begin{pmatrix} 0 & A_r \\ A_c^T & 0 \end{pmatrix} \tag{6}$$

Assuming that **M** is diagonalizable i.e, that there exists a non singular matrix Q of eigenvectors such that

$$Q^{-1}\mathbf{M}Q = diag(\lambda_1, \lambda, ..., \lambda_{n+m}).$$
⁽⁷⁾

Furthermore, assuming that the eigenvalues are ordered $|\lambda_1| > |\lambda_2| \ge |\lambda_3| \ge \dots \ge |\lambda_{n+m}|$.

and expanding the initial approximation $\pi^{(0)} = \begin{bmatrix} u^{(0)} \\ v^{(0)} \end{bmatrix}$ in terms of the eigenvectors of **M**

$$\pi^{(0)} = c_1 q_1 + c_2 q_2 + \dots + c_{n+m} q_{n+m} \text{ with } q_k = \begin{bmatrix} y_k \\ z_k \end{bmatrix}$$
(8)

where the upper part y_k is for the rows of A_r and the lower part z_k is for the columns of A_c , $c_k \neq 0$ is assumed. We have

$$\pi^{(t)} = \mathbf{M}^{(t)} \pi^{(0)} = \lambda_1^{(t)} (c_1 q_1 + \sum_{k=2}^{n+m} c_k (\frac{\lambda_k}{\lambda_1})^{(t)} q_k)$$
(9)

for k = 2, ..., n + m, we have $|\lambda_i| < |\lambda_1|$, so the second term tends to zero, and the power method converges to the eigenvector $q_1 = \begin{bmatrix} y_1 \\ z_1 \end{bmatrix}$ corresponding to the dominant eigenvalue λ_1 . The rate of convergence is determined by the ratio $|\frac{\lambda_2}{\lambda_1}|$, if this is close to one the convergence is very slow.

In equation (9), we expand the left eigenvector $\pi^{(t)}$ as a linear combination of the eigenvectors of $\mathbf{M}^{(t)}$. It is easy to see that $\pi^{(t)} = \begin{bmatrix} u^{(t)} \\ v^{(t)} \end{bmatrix}$ is the leading eigenvector of $\mathbf{M}^{(t)}$. Mathematically, in equation (10)we show that the leading eigenvector of $\mathbf{M}^{(t)}$ is closely related to the first singular vectors of $\hat{A}^{(t)}$, $\pi_r^{(t)}$ and $\pi_c^{(t)}$ are the left and right singular vector of $\hat{A}^{(t)}$ respectively.

$$\begin{bmatrix} u^{(t)} \\ v^{(t)} \end{bmatrix} = \begin{bmatrix} 0 & A_r^{(t)} \\ A_c^{T(t)} & 0 \end{bmatrix} \begin{bmatrix} u^{(0)} \\ v^{(0)} \end{bmatrix} = \begin{bmatrix} A_r^{(t)} u^{(0)} \\ A_c^{T(t)} v^{(0)} \end{bmatrix}$$
(10)

Instead of constructing \mathbf{M} (like the most spectral co-clustering methods do) which is bigger and sparser than the data matrix A, we provide a way to visualize and co-cluster data, not using \mathbf{M} but directly from A.



Now, by exploiting the block structure of **M** in equation (9), we show in equation (11); $u^{(t)}$ as a linear combination of the left singular vectors of A_r and $v^{(t)}$ as a linear combination of the right singular vectors of A_c .

$$\begin{bmatrix} u^{(t)} \\ v^{(t)} \end{bmatrix} = \begin{bmatrix} \lambda_1^{(t)} (c_1 y_1 + \sum_{i=2}^n c_i (\frac{\lambda_i}{\lambda_1})^{(t)} y_i) \\ \lambda_1^{(t)} (c_1 z_1 + \sum_{j=n+1}^{n+m} c_j (\frac{\lambda_j}{\lambda_1})^{(t)} z_j) \end{bmatrix}$$
(11)

4 Experiments

We now provide experimental results to illustrate the behavior of The R1SVD algorithm. We argue that R1SVD allows us to capture the trends of objects over a subset of attributes and then reorganizes data matrix into homogeneous blocks. We apply our algorithm on different real world and word-document simulated data sets (using a Bernoulli latent block model [10]). Different patterns are considered to show the ability of the R1SVD algorithm to rediscover the hidden blocks in data without fixing any parameters on the rows and columns ordering.

Table 1. 16 Townships Data.

	Α	в	С	D	Е	F	G	Η	Ι	J	Κ	L	Μ	Ν	0	Р
High School	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
Agricult Coop	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0
Rail station	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
One Room School	1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	1
Veterinary	0	1	1	1	0	0	1	0	0	0	0	1	0	0	1	0
No Doctor	1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	1
No Water Supply	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0
Police Station	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
Land Reallocation	0	1	1	1	0	0	1	0	0	0	0	1	0	0	1	0

Table 2. Reorganization of 16 Townships data after co-clustering.

	Н	Κ	в	\mathbf{C}	\mathbf{D}	\mathbf{G}	\mathbf{L}	Ο	Μ	Ν	J	Ι	Α	Р	F	Е
High School	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Railway Station	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Police Station	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Agricult Coop	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0
Veterinary	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0
Land Reallocation	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0
One Room School	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
No Doctor	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
No Water Supply	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0

We now try to visualize the homogeneous block structure that might be discovered by our algorithm. Table 1 shows the original characteristics- townships data matrix and table 2 the reordered matrix obtained by arranging rows and columns based on the sorted u and v. the table 2 reveals the hidden sparsity structure of both characteristics and townships clusters. The three diagonal blocks in this table correspond to the three clusters. It clearly appears that we can characterize each cluster of townships by a cluster of characteristics: {H, K} by {High School, Railway Station,Police Station}, {B, C, D, G, L, O} by {Agricult Coop, Veterinary,Land Reallocation} and {M, N, J, I, A, A, P, F, E} by {One Room School, No Doctor, No Water Supply}.

Figure 1 shows in order from the left to right, the original characteristicstownships data matrix, the reordered data matrix, the reordered rows similarity matrix $S_r = AA^T$, the reordered columns similarity matrix $S_c = A^T A$ and the plots of sorted singular vectors u and v.





Fig. 1. Townships: reordered $Sr = AA^T$ according u, reordered $Sc = A^TA$ according v.

The three diagonal blocks in figures 2 and 3 correspond to the three clusters. The rough block diagonal structure indicates the cluster structure relation between documents and words. Hence by exploiting the duality of the data and features, incorporating the features information in data reordering at each stage, our algorithm yields better reordering solution that one dimensional clustering approaches, especially for high dimensional sparse datasets.



Fig. 2. Data1: reordered $Sr = AA^T$ according u, reordered $Sc = A^TA$ according v.



Fig. 3. Data1: reordered $Sr = AA^T$ according u, reordered $Sc = A^TA$ according v.

The R1SVD framework seems to have the potential to address the question of the number of clusters underlying the data, it detects the suitable number of blocks by analyzing the evolution of the first left u of A_r and right singular vectors v of A_c . A performance study has been conducted to evaluate our method. In this subsection, we try to answer the question; is this reordering meaningful?. In order to be able to answer this question we use confusion matrices to measure the clustering performance of the co-clustering result provided by our method. The co-clustering task is to recover groups of rows and columns. After the learning stage, the clusters indicators are given by the vectors u and v. It can be seen that our method reconstructs efficiently all co-clusters for balanced and unbalanced data sets used in our experiments.

From table 3, we observe that data reordering provided by R1SVD can be useful in a co-clustering context. It is very interesting to underly the fact that R1SVD visualization does not destroy data and, unlike most co-clustering algorithms, it does not require the number of blocks.

Table 5. Comusion matrix evaluation on rows and columns data.
--

data	1 (ro	ows)	data	a2 (r	ows)	data1 (col	umns)	data2	(col	umns)
0	205	0	0	0	795	0	0	40	155	0	0
1614	0	5	626	0	0	397	0	0	0	133	0
0	0	176	0	579	0	0	63	0	0	0	212

5 Conclusion

In this paper we have presented an iterative matrix-vector multiplication procedure called rank-one SVD for data visualization. The procedure is to apply iteratively an appropriate stochastic adjacency data matrix associated to a bipartite graph, and then compute the first leading left singular vector associated to the eigenvalue λ_1 of this matrix. Stopping the algorithm after a few iterations involves a visualization of data matrix into homogeneous blocks. This approach appears therefore very interesting in co-clustering context.

References

- 1.B. B. Bederson and B. Shneiderman, The Craft of Information Visualization: Readings and Reflections, San Francisco, Morgan Kaufmann, 2003.
- 2.B. Shneiderman, Inventing discovery tools: combining information visualization with data mining, Inf Vis 1 (2002), 5-12.
- 3.H. Spath, Cluster Analysis Algorithms for Data Reduction and Classification of Objects, Chichester, UK, Ellis Horwood, 1980.
- 4.F. Murtagh, Book review: W. Gaul and M. Schader, Eds., Data, expert knowledge and decisions, Heidelberg: Springer-Verlag, 1988, viii + 380, J Classification 6 (1989), 129-132.
- 5.P. Arabie and L. J. Hubert, An overview of combinatorial data analysis, In Clustering and Classification, P. Arabie, L. J. Hubert, and G. De Soete, eds. River Edge, World Scientific, 1996, 5-63.
- 6.P. Arabie and L. J. Hubert, Combinatorial data analysis, Annu Rev Psychol 43 (1992), 169-203.
- 7.J. Bertin, Graphics and Graphic Information Processing, Berlin, Walter de Gruyter (Translated by W. J. Berg and P. Scott), 1981.
- 8.I. Dhillon, "Co-clustering documents and words using bipartite spectral graph partitioning," ACM SIGKDD International Conference, San Francisco, USA, pp. 269-274, 2001.
- 9.Horn, R. A., Johnson, C. R. (1986). Matrix analysis. Cambridge, U.K.: Cambridge University Press.
- 10.G. Govaert and M. Nadif, "Block clustering with Bernoulli mixture models: Comparison of different approaches," *Computational Statistics and Data Analysis*, 52, pp. 233-3245, 2008.
- 11.G.W.Stewart, On the history of the simgular value decomposition. SIAM Review, 35, 551-566, 1996.



Approach for Vehicle Routing Problem with Resource Constraints

A. LAMAMRI⁽¹⁾, A. NAGIH⁽²⁾, H. AIT HADDADENE⁽³⁾ ⁽¹⁾University Khemis-Miliana, Algeria, ⁽²⁾Laboratory LITA, University Paul Verlaine – Metz, France. ⁽³⁾Laboratory LITA, Lorraine University Ile-Saulcy-Metz,France. *E-mail* :plamamri@yahoo.fr

Abstract

The vehicle routing problem with resource constraints is a generalization of the classical vehicle routing problem. Given the large size of the problems encountered in practice, these models are solved by an approach based on column generation that can handle implicitly all feasible solutions and a master problem determining the best solution. In this paper, we present the development of a new approach to improve the acceleration of the method of column generation for solving the problem of construction vehicle routing, it is projected in each arc, the resources a vector of size smaller by using a Lagrangean relaxation algorithm to determine the coefficients of the projection arc combined with an algorithm for re-optimization, then generates a subset of solutions to the master problem. Numerical testes on problems from instances of random vehicle routing giving competitive results.

Keywords: Vehicle routing problem, path problems in graphs, decomposition methods, column generation.

1 Introduction

The vehicle routing problem with resource constraints (VRPRC) is given by a set of customers N and a set of K vehicles available in a repository. This problem is to find a set of minimum *cost* route, departing and returning to a single repository, where each customer is visited by one vehicle to satisfy some demand. Each customer must be served during a given time window. A vehicle arriving in advance at a customer waits until the start date of service without additional *cost*, time windows in this case are called 'hard'. Some models penalize the hold with an extra *cost*, these models are called 'soft ', but much research is devoted to the time windows 'hard'. A vehicle arriving late at a customer is not allowed to perform his service. A feasible



tour is a series of visits (conducted by the same vehicle) respecting the time windows, which begins and ends at the same depot.

The (VRPRC) is defined on the network $G^k = (X^k, A^k), X = N \cup \{o^k, d^k\}$, where the depot is represented by the two nodes o^k and d^k , and $N = \{1, \ldots, n\}$ is the set of vertices representing the customer, and A the set of arcs that interconnect the customers and the depot. An arc $(i, j) \in A$ means the possibility of linking the service customers i and j. To write the formulation of this problem, we introduce the following notations:

 $-c_{ij}$ is the cost of the arc $(i, j) \in A$.

 $-t_{ij}$ The arc duration $(i, j) \in A$.

- $[a_i, b_i]$ The time window during which the customer service $i \in N$ must start.

 $-d_i$ Customer demand $i \in N$.

- Q capacity of each vehicle.

Assignment of customers to vehicles is called feasible if:

 The combined demand of customers visited by a vehicle does not exceed its capacity.

- Time constraints are met by each vehicle.

- Each customer is visited by one vehicle.

- Every vehicle that leaves the depot back to depot after completing his route.

The problem is to find a feasible assignment of vehicles to tour the minimum cost.

2 Formulation

The (VRPRC) can then be formally described as the following multi-commodity network flow model with time window and resource constraints:



$$\min\sum_{(k=1)}^{K} \sum_{(i,j) \in A} c_{ij} x_{ij}^k \tag{1}$$

s.t
$$\sum_{(k=1)}^{K} \sum_{(i,j)\in A} x_{ij}^{k} = 1$$
 for $i \in N = \{1, \dots, n\}$ (2)

$$\sum_{(i,j)\in A}^{\infty} d_i \ x_{ij}^k \le Q \quad \text{for } k \in K \tag{2'}$$

$$\sum_{\substack{(i:(o^k,i)\in A)}}^{(i)} x_{o^k i}^k = 1 \quad \text{for} \quad k \in K \tag{3}$$

$$\sum_{\substack{(i,d^k) \in A \\ id^k}} x_{id^k}^k = 1 \quad \text{for } k \in K$$
(4)

$$\sum_{\substack{(i:(i,j)\in A)}}^{(i:(i,d^k)\in A)} x_{ij}^k = \sum_{\substack{(l:(j,l)\in A)}} x_{jl}^k \text{ for } j \in N$$
(5)

$$x_{ij}^k(T_i^{(k,q)} + t_i^{(k,q)} - T_j^{(k,q)} \le 0 \text{ for } (i,j) \in A, k \in K, q \in Q$$
 (6)

$$a_i^{(k,q)} \leq T_i^{(k,q)} \leq b_i^{(k,q)}$$
 for $i \in N, k \in K, q \in Q$ (7)

 $x_{ij}^k \in \{0,1\}, T_i^{(k,q)} \ge 0 \text{ for } (i,j) \in A, k \in K, q \in Q$ (8)

Binary variables x_{ij}^k indicate if the tour takes the arc $(i, j) \in A$, while the variable $T_i^{(k,q)}$ indicates the cumulative consumption of each resource q at each node i.

The objective function (1) minimizes the total travel cost. The constraints (2) ensure that each customer is visited exactly once, and (2') state that a vehicle can only be loaded up to it's capacity. Next, equations (3-5)indicate that each vehicle must leave the depot o; after a vehicle arrives at a customer it has to leave for another destination; and finally, all vehicles must arrive at the depot d. The inequalities (6) establish the relationship between the vehicle departure time from a customer and its immediate successor. Finally constraints (7) affirm that the time windows are observed, and (8) are the integrality constraints. Note that an unused vehicle is modeled by driving the "empty" route (o, d), and the constraints (5) provides the cumulative consumption of resource q at node j, since we have :

$$T_{i}^{(k,q)} = \max(a_{i}^{(k,q)}, T_{i}^{(k,q)} + t_{ij}^{(k,q)}) \quad (9)$$

Note that the constraints (3 - 7) are local constraints valid only for the network G. Only the partitioning constraints (2) are global constraints linking the K sub-networks. The relaxation of these binding constraints and the decomposition of the initial problem by sub-network will be an interesting option for a resolution. Finally, note that resource constraints (6 - 7)make the problem (*VRPRC*) *NP*-hard. Even the problem of realizability is associated *NP*-complete [5].



3 Solving approaches

And as the number of coefficients to adjust will be more important for the approach of projection arcs, finding the optimal multiplier u_{ij}^* require several iterations of DPA-L [2], this method can be expensive. To quickly obtain good heuristic solutions (feasible), our approach applied once DPA-L and then apply DPA-LND [2], using multipliers to find u_{ij} generate feasible columns and negative marginal cost. Specifically, we first choose a sequence of steps (p_k) as the standard $(\sum p_k)$ is divergent and $\lim_{n\to\infty} p_k = 0$, i.e. conditions ensuring convergence of the algorithm sub gradient. It applies primarily DP-L using multipliers $u_{ij}^{(k-1)}$ of the previous iteration, we find the sub gradients Sg_{ij}^k corresponding arc (i, j) then we calculates the new Lagrange multipliers u_{ij}^k . This heuristic is certainly based on the fact that when k is large, the vector C_k reduced costs on the arcs of the network do not change much from one iteration to another of the algorithm for column generation. Thus for k large, we can expect to see u_{ij}^k converge to an optimal value.

The main steps of our approach are summarized below:

Master problem

4 Numerical results

This section presents the preliminary evaluation of our approach to the problem of construction of vehicle routing with a single resource.

Solomon's 100-customer Euclidean (VRPTW) instances are used to test our algorithm. In these instances, the travel time and the Euclidean distance between two customer locations are the same and this value is truncated to two decimal places. There are six different classes of instances depending on the geographic location of the customers (R : random; C : clustered ; RC: mixed) and width of the scheduling horizon (1 : short horizon; 2 : long horizon). In this work, instances of type 1 are discarded due to the short horizon that does not allow a significant number of routes to be sequenced to form a workday. Results are thus reported for R2, C2 and RC2. Due to the limitations of our exact approach, the computational study focuses on instances obtained by taking only the first 25 customers from each original instance.



Solomon's (VRPTW) test instances are modified to fit our problem. In particular, a value tmax to limit route duration is needed. This value was first set to 100 in the case of R2 and RC2, and 200 in the case of C2. The value is larger for C2 because the service or dwell time at each customer is 90, as opposed to 10 for R2 and RC2. Finally, a gain of 1 is associated with each customer and weighted by an arbitrarily large constant to maximize first the number of served customers, and then minimize the total distance.

The results for the instances with reduced time windows are shown in Table1. In the table1, a particular instance is identified by its class and its index followed by a dot and the number of customers considered. For example RC202.25 is the second instance of class RC2, where only the first 25 customers are considered. In these table, column Problem is the identifier of the problem instance, ItrGC is the total number of iteration of (PM) solved by Simplex, Col is the total number of columns generated during the branch-and-price algorithm, T(ssp) is the computation time in seconds and Obj. is the total distance.

Problems	ItrGC	Col	$T(\mathtt{ssp})$	Obj
RC201.25	123	609	0.9	967.9
RC202.25	110	1132	221.0	961.6
RC203.25	713	2589	2566.2	751.3
RC205.25	218	944	5.4	974.9
RC206.25	444	1703	4.6	977.1
RC207.25	3119	13989	418.4	819.6
R201.25	218	577	1.0	772.8
R202.25	108	1030	127.0	694.0
R205.25	1326	4930	60.1	761.2
R210.25	71	9 18	121.4	704.6
R211.25	57	1150	42.9	623.7
C201.25	329	3448	5.1	679.5
C202.25	4023	13860	782.8	677.3

Table1

The comparison between the different methods and our approach has revealed that it has provided good results. These are best when certain conditions are met:

The initialization of the algorithm and the choice of Lagrange multipliers and the displacement step.



Conclusions

In this paper, we proposed an algorithm for vehicle routing problem with resource constraints (VRPRC) which is an extension of the classical vehicle routing problem (VRP) to take into account the more practical problem; we have mainly developed approaches to column generation and decomposition master problem and sub-problem. The difficulty of solving the sub-problem is directly related to the number of resources, we particularly studied the techniques of reduction of space resources, and this notion of reduction is a key element of the effectiveness of the overall resolution of problem.

References

- M. Desrochers, F. Soumis, "AGeneralized Permanent Labeling Algorithm for the Shortest Path Problem with Time Windows", INFOR 26, 191 – 212 (1988).
- [2] A. Nagih, F. Soumis "Nodal aggregation of resource constraints in a shortest path problem", European Journal of Operational Research (2005).
- [3] N. Touati, L. Létocart, and A. Nagih. "Solutions diversification in a column generation scheme". Discrete Optimization (2008).
- [4] N. Touati, L. Létocart, and A. Nagih. "Reoptimization in a column generation scheme". Computers and Operations Research (2008).
- [5] Vangelis Paschos, "Livre, optimisation combinatoire 3: applications", Hermès Science. ch 10 (2005).



SGR modeling of fake ordinal data with correlational structures

Luigi Lombardi¹, Massimiliano Pastore², Massimo Nucci³, and Andrea Bobbio⁴

- ¹ Dipartimento di Psicologia e Scienze Cognitive, Università di Trento, Corso Bettini, 31, I-38068 Rovereto (TN), Italy (E-mail: luigi.lombardi@unitn.it)
- ² Dipatimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova, via Venezia, 8, I-35131 Padova, Italy (E-mail: massimiliano.pastore@unipd.it)
- ³ Dipartimento di Psicologia Generale, Università di Padova, via Venezia, 8, I-35131 Padova, Italy (E-mail: massimo.nucci@unipd.it)
- ⁴ Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia Applicata, Università di Padova, via Venezia, 8, I-35131 Padova, Italy (E-mail: andrea.bobbio@unipd.it)

Abstract. In many psychological inventories (i.e., personnel selection surveys and diagnostic tests) the collected samples often include fraudulent records. This confronts the researcher with the crucial problem of biases yielded by the usage of standard statistical models. In this paper we generalize a recent probabilistic perturbation procedure, called SGR - Sample Generation by Replacements - (Lombardi & Pastore[4]), to simulate fake data with correlational structures. To mimic these more complex faking data we proposed a novel extension of the SGR conditional replacement distribution which is based on a discrete version of the truncated multivariate normal distribution. We also applied the new procedure to real behavioral data on the role of perceived affective self-efficacy in social contexts.

Keywords: Sample Generation by Replacement, Fake-good data, Truncated multi-variate normal distribution.

Many self-report measures of attitudes, beliefs, personality, and pathology include items that can be easily manipulated by respondents. For example, an individual may deliberately attempt to manipulate or distort responses to personality inventories and attitude tests to create positive impressions (e.g., Paulhus[7]; Zickar & Robie[8]). In other circumstances, some individuals may tend to malinger responses on a symptom checklist to simulate grossly exaggerated physical or psychological symptoms in order to reach specific goals such as, for example, obtaining financial compensation, avoiding being charged with a crime, avoiding military duty, or obtaining drugs (e.g., Hall & Hall[2]).

Sample Generation by Replacement (SGR) is a recent data simulation procedure to artificially generate samples of fake ordinal data (Lombardi & Pastore[4]). SGR is based on a two-stage sampling algorithm characterized by two distinct generative models: the model representing the process that generates the data prior to any fake perturbation (data generation process) and the



model representing the faking process to perturb the data (data replacement process). This approach has been recently applied to evaluate the impact of hypothetical faking good manipulations in ordinal data on the performances of a set of widely known and commonly used stand-alone SEM-based fit indices (Lombardi & Pastore[4]). SGR has also been used to study the sensitivity of reliability indices to fake perturbations in dichotomous and ordered data generated by factorial models (Pastore & Lombardi[6]).

Up to now the SGR approach has been limited to the modeling of conditionally independent fake data. In this contribution, we propose a novel generalization of the conditional replacement distribution that accounts for possible correlated structures in the simulated fake data.

1 Main features of the SGR approach

1.1 Data generation process

We think of the original data as being represented by an $n \times m$ matrix **D**, that is to say, n observations (e.g., participants) each containing m elements (e.g., participant's responses). We assume that entry d_{ij} of **D** (i = 1, ..., n; j = 1, ..., m) takes values on a small ordinal range 1, 2, ..., Q. In particular, let \mathbf{d}_i be the $(1 \times m)$ array of **D** denoting the simulated pattern of responses of participant i. The response pattern \mathbf{d}_i is a multidimensional ordinal random variable with probability distribution $p(\mathbf{d}_i|\theta)$, where θ indicates the vector of parameters of the generative probabilistic model of the data. Moreover, we assume that the simulated response patterns are independent and identically distributed (i.i.d.) observations.

1.2 Data replacement process

The basic principle of the SGR approach is to generate a new $n \times m$ ordinal data matrix \mathbf{F} , called the *fake data matrix* of \mathbf{D} , by manipulating each element d_{ij} in **D** according to a replacement probability distribution. Let \mathbf{f}_i be the $(1 \times m)$ array of **F** denoting the hypothetical pattern of fake responses of participant i. The fake response pattern \mathbf{f}_i is a multidimensional ordinal random variable with conditional replacement probability distribution $p(\mathbf{f}_i | \mathbf{d}_i, \theta_F)$ where θ_F indicates the vector of parameters of the probabilistic faking model. In general, θ_F represents hypothetical a priori knowledge about the distribution of faking (e.g., the chance of observing a fake observation in the data) or empirically based knowledge about the process of faking (e.g., the direction of faking - fake good vs. fake bad -). In the SGR framework the replacement distribution $p(\mathbf{f}_i | \mathbf{d}_i, \theta_F)$ is restricted to satisfy a conditional independence assumption (see Lombardi & Pastore[4]; Pastore & Lombardi[6]). More precisely, in the replacement distribution each fake response f_{ij} only depends on the corresponding data observation d_{ij} and the model parameter θ_F . Therefore, because the patterns of fake responses are also i.i.d. observations, the



simulated data array (\mathbf{D}, \mathbf{F}) is drawn from the joint probability distribution

$$p(\mathbf{D}, \mathbf{F}|\theta, \theta_F) = \prod_{i=1}^{n} p(\mathbf{d}_i|\theta) p(\mathbf{f}_i|\mathbf{d}_i, \theta_F)$$
(1)

$$=\prod_{i=1}^{n} p(\mathbf{d}_i|\theta) \prod_{j=1}^{m} p(f_{ij}|d_{ij},\theta_F)$$
(2)

By repeatedly sampling data from the two generative models we can simulate the so called *fake data sample* (FDS). We can then study the distribution of some relevant statistics computed on this FDS.

1.3 The problem of the independence assumption

We recall that the SGR simulation procedure generates fake perturbations that are restricted to satisfy the conditional independence assumption. Unfortunately, this restriction clearly limits the range of empirical faking processes that can be mimicked by the SGR simulation procedure. In particular, because the replacement distribution acts as a perturbation process for the original data, the resulting fake data will always show correlations that are (on average) weaker than the ones observed for the original data, thus showing a sort of residual correlation effect. However, it is known that some empirical contexts may require different model assumptions about the faking process that cannot be captured by this simple framework. For example, different modulations of graded faking such as slight faking and extreme faking (e.g., Zickar & Robie[8]) are not consistent with the simple independence hypothesis.

To fill this gap, in this contribution we propose a novel conditional replacement distribution that allows to modulate different levels of correlational patterns in the simulated fake data. Because our new proposal is based on a discrete version of the truncated multivariate normal distribution, in what follows we present some relevant properties of this important distribution function before introducing the new perturbation model that does not hinge on the independence assumption.

2 A SGR framework for correlated fake-good data

The SGR approach offers an elegant way to simulate faking good scenarios. In general, faking good can be conceptualized as an individual's deliberate attempt to manipulate or distort responses to create a positive impression (Paulhus[7]; Zickar & Robie[8]). Notice that, the faking good (as well as the faking bad) scenario always entails a conditional replacement model in which the conditioning is a function of response polarity. This model represents a perturbation context in which responses are exclusively subject to positive feigning:

$$f_{ij} \ge d_{ij}; \qquad i = 1, \dots, n; j = 1, \dots, m$$



2.1 The truncated multivariate replacement distribution

As a kernel for the conditional replacement distribution we consider the truncated multivariate normal distribution $TN(\mu, \Sigma, \mathbf{a}, \mathbf{b})$ (e.g., Horrace[3]). This distribution can be expressed as

$$f(\mathbf{x}|\mu, \mathbf{\Sigma}, \mathbf{a}, \mathbf{b}) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right\}}{\int_{\mathbf{a}}^{\mathbf{b}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right\} d\mathbf{x}}$$
(3)

for $\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}$ and 0 otherwise. The $(1 \times m)$ vectors \mathbf{a} and \mathbf{b} are the lower and upper truncation points $(a_j < b_j; j = 1, ..., m)$ for the multivariate normal distribution with m dimensions. Finally, μ and Σ are the location parameter vector and the covariance matrix of the (not truncated) multivariate normal distribution.

Now, let $\mathbf{f}_i = (k_1, \ldots, k_m)$ and $\mathbf{d}_i = (h_1, \ldots, h_m)$ be the replaced values and the original values for the i^{th} simulated observation, respectively. According to the Underlying Variable Approach (UVA; Muthén[5]) we can set

$$p(\mathbf{f}_i | \mathbf{d}_i, \theta_F) = \begin{cases} \int_{\alpha_{k_1-1}}^{\alpha_{k_1}} \cdots \int_{\alpha_{k_m-1}}^{\alpha_{k_m}} f(\mathbf{x} | \mathbf{0}, \mathbf{\Sigma}, \mathbf{a}^i, \mathbf{b}^i) d\mathbf{x}, \forall j : 1 \le h_j \le k_j \le Q \\ 0, \qquad \exists j : k_j < h_j \end{cases}$$

In the replacement distribution **0** is the $1 \times m$ array of zeros representing the location parameter (that is to say $\mu = 0$), whereas the pair $(\alpha_{k_j-1}, \alpha_{k_j})$ are the thresholds corresponding to the discrete value k_j (j = 1, ..., m). Following the UVA framework we assume that there exists a continuous data matrix \mathbf{F}^* underlying the fake ordinal data matrix \mathbf{F} . The connection between the ordinal variable f_{ij} and the underlying variable f_{ij}^* in \mathbf{F}^* is given by

$$f_{ij} = k_j \quad \Leftrightarrow \quad \alpha_{k_j-1} < f_{ij}^* < \alpha_{k_j}; \quad i = 1, \dots, n; j = 1, \dots, m.$$

Recall that to represent an ordinal item with Q categories we need Q + 1 thresholds. Finally, the bounds $\mathbf{a}^i = (a_1^i, \ldots, a_m^i)$ and $\mathbf{b}^i = (b_1^i, \ldots, b_m^i)$ are set to

$$a_j^i = \alpha_{h_j-1}$$
 $b_j^i = +\infty, \quad j = 1, \dots, m$

where we recall that $(\alpha_{h_j-1}, \alpha_{h_j})$ is the pair of thresholds corresponding to the value h_j for the original variable d_{ij} in \mathbf{d}_i . Figure 1 shows an example for a one dimensional case. In sum, the parameter array for the faking model is given by

$$\theta_F = (\alpha, \Sigma).$$

with α being the $m \times (Q-1)$ threshold matrix.

2.2 Some relevant properties of the new replacement distribution

It is important to point out two interesting properties of the novel replacement distribution for correlated fake data.





Fig. 1. Example of a truncated normal distribution (one dimensional) with d = 2 and f = 3. The mean μ of the distribution is 0, the bounds a and b are α_1 and $+\infty$, respectively. The conditional probability $p(3|2, \theta_F)$ is the shaded area between α_2 and α_3 .

The first property is related to the probability of replacement π . According to the truncated multivariate replacement distribution the value $1 - \pi_i$ of the conditional probability of non-replacement

$$\mathbf{f}_i = \mathbf{d}_i = (h_1, \dots, h_m)$$

is equivalent to

$$1 - \pi_i = \int_{\alpha_{h_1-1}}^{\alpha_{h_1}} \cdots \int_{\alpha_{h_m-1}}^{\alpha_{h_m}} f(\mathbf{x}|\mathbf{0}, \mathbf{\Sigma}, \mathbf{a}^i, \mathbf{b}^i) d\mathbf{x}$$
(4)

and consequently, the probability of replacement π_i is

$$\pi_i = 1 - \left(\int_{\alpha_{h_1-1}}^{\alpha_{h_1}} \cdots \int_{\alpha_{h_m-1}}^{\alpha_{h_m}} f(\mathbf{x}|\mathbf{0}, \mathbf{\Sigma}, \mathbf{a}^i, \mathbf{b}^i) d\mathbf{x} \right)$$
(5)

Note that in the original model of replacement (Lombardi & Pastore[6]) the probability of replacement π was an explicit parameter in the replacement distribution, whereas in this new proposal π is an implicit parameter that can be derived by taking the average across the *n* distincts π_i :

$$\bar{\pi} = \frac{1}{n} \sum_{i=1}^{n} \pi_i.$$

The second property is related to the correlational structure of the simulated fake data set \mathbf{F} . Unlike the standard model of replacement, in the new configuration we can directly represent correlations between the replaced values for the m ordinal variables. In particular, the correlation matrix \mathbf{R}_f of \mathbf{F} can be modulated by the covariance matrix $\boldsymbol{\Sigma}$ in the replacement model. Note, however, that $\boldsymbol{\Sigma}$ is the covariance matrix of the original (not truncated) multivariate normal distribution. In general, the computation of the first and



second moments is not trivial for the truncated case, since they are obviously not the same as μ and Σ from the parametrization of $TN(\mathbf{0}, \Sigma, \mathbf{a}, \mathbf{b})$. In particular, it can be seen that truncation can significantly reduce the variance and change the covariance between variables. In the next section we will show some examples of how the resulting correlation matrix \mathbf{R}_f can be affected from the interaction between different modulations of faking (represented by different configurations of threshold values α) and the structure of the covariance matrix Σ .

3 Applicative example

The new replacement distribution is illustrated using data from a questionnaire about the role of perceived affective self-efficacy in personality and social psychology (Bandura, Caprara, Barbaranelli, Gerbino, and Pastorelli[1]). Participants were 463 undergraduate students (389 females) at the University of Padua (Italy). Ages ranged from 18 to 48, with a mean of 20.64 and a standard deviation of 2.71. Data consisted of the participants' responses to three of the 12 items of the AEP/A scale (Caprara, 2001) scored on a 4-point agree-disagree scale (value 1 denotes that a participant totally disagrees with the statement, whereas value 4 means total agreement with the statement). However, the 463 participants were divided into two groups. The first group $(n_1 = 231)$ received a neutral set of instructions, whereas the second group $(n_2 = 232)$ received ad lib faking instructions. The resulting responses were collected into two empirical data matrices: \mathbf{D}_e for the neutral group and \mathbf{F}_e for the ad lib faking group. As expected the observed responses for the ad lib faking group were affected by fake good observations. More precisely, the participants deliberately manipulated their responses using larger values of the scale to create better impressions. This hypothesis was partially supported by the moderate ceiling effects observed in the data (see Fig. 2). Because no additional items on social desirability was available in the AEP/A inventory, we decided to perform an SGR analysis on the basis of two hypothetical scenarios: slight faking and uniform faking.

3.1 Comparing faking models

An SGR analysis was used to evaluate the mimicking ability of four different faking models with respect to the empirical fake set \mathbf{F}_e . In particular, because \mathbf{D}_e and \mathbf{F}_e contained responses collected from two different groups of individuals, we decided to evaluate the simulated fake samples against the empirical marginal means of the three items as well as against their empirical correlations in \mathbf{F}_e . To that end, we defined four perturbation models derived by the combination of two factors with two levels each. The first factor defined two structures for the covariance matrix in the truncated replacement distribution: a) the identity matrix \mathbf{I} (denoting an independent model) b) the empirical correlation matrix of \mathbf{R}_e computed on the observed matrix \mathbf{F}_e (representing the correlational structure in \mathbf{F}_e). The second factor represented two different options for the thresholds in the replacement distribution: a) $\alpha_1 = -0.674, \alpha_2 = 0.0, \alpha_3 = 0.674$ denoting a uniform support fake-good distribution (Lombardi and Pastore[4]) b) $\alpha_1 = -1.591, \alpha_2 = 1.596, \alpha_3 = 3.332$ denoting a slight fake-good distribution (e.g., Zickar & Robie[8]). All the other components (parameters' values) were set to identical values in all the four faking models. In particular, the original data set **D** in the conditional replacement distribution was set equal to the empirical data set **D**_e. The latter means that in this application we did not simulate the data **D**, instead we directly used the observed data **D**_e in the replacement distribution equation. The four faking models were then used to simulated new FDSs.

The results of the SGR analysis are shown in Figures 2,3. Figure 2 represents the simulated marginal means of the fake-good data as a function of the two factors. The results showed that the slight faking model with mild dependency was preferred to the other three models. Similarly, figure 3 shows the simulated correlations of the fake-good data as a function of the two factors. Like for the marginal means, also for the correlations the slight faking model with mild dependency showed a better performance as compared to the other three models. In other words, the ad lib faking instructions seem to be more consistent with a mild *positive impression effect* that boosts the correlations between the items.



Fig. 2. Boxplots for the simulated marginal means of the fake-good data for the four models. The solid line denotes the observed pattern for the marginal means in \mathbf{D}_e . The dashed line indicates the observed pattern for the marginal means in \mathbf{F}_e . UF = uniform support fake-good distribution, SL = slight fake-good distribution; SO = independent faking model, Sf = faking model with a correlational structure. a_1, a_2 , and a_3 denote the three selected items of the AEP/A scale. The data represented in each boxplot were derived from 500 FDSs.





Fig. 3. Boxplots for the simulated correlations of the fake-good data for the four models. The solid line denotes the observed correlational pattern in \mathbf{D}_e . The dashed line indicates the observed correlational pattern in \mathbf{F}_e . UF = uniform support fake-good distribution, SL = slight fake-good distribution; S0 = independent faking model, Sf = faking model with a correlational structure. The variable $rf_{jj'}$ denotes the correlation between item a_j and item a'_j of the AEP/A scale. The data shown in each boxplot were derived from 500 FDSs.

References

- A. Bandura, G.V. Caprara, C. Barbaranelli, M. Gerbino and C. Pastorelli. Role of affective and self-regulatory efficacy in diverse spheres of psychosocial functioning. *Child Development*, 74, 3, 769–782, 2003.
- 2.R. C. Hall and R. C. Hall. Detection of malingered PTSD: An overview of clinical, psychometric, and physiological assessment: Where do we stand? *Journal of Forensic Science*, 52, 717–725, 2007.
- 3.W.C. Horrace. Some results on the multivariate truncated normal distribution. Journal of Multivariate Analysis, 94, 209–221, 2005.
- 4.L. Lombardi and M. Pastore. Sensitivity of fit indices to fake perturbation of ordinal data: A sample by replacement approach. *Multivariate Behavioral Research*, 47, 519–546, 2012.
- 5.B. Muthén. A general structural equation model with dichotomous, ordered categorical and continuous latent variables indicators. *Psychometrika*, 49, 115–132, 1984.
- 6.M. Pastore and L. Lombardi. The impact of faking on Cronbach's Alpha for dichotomous and ordered rating scores. submitted paper, 2012.
- 7.D. L. Paulhus. Two-component models of socially desirable responding. Journal of Personality and Social Psychology, 46, 598–609, 1984.
- M. J. Zickar and C. Robie. Modeling faking good on personality items: An item-level analysis. Journal of Applied Psychology, 84, 551–563, 1999.



Robust extrapolation of functionals of stochastic sequences with stationary increments

M. Luz and M. Moklyachuk

Department of Probability Theory, Statistics, and Actuarial Mathematics, National Taras Shevchenko University of Kyiv, Kyiv 01601, Ukraine (E-mail: mmp@univ.kiev.ua)

Abstract. The problem of optimal estimation of the functional $A\xi = \sum_{k=0}^{\infty} a(k)\xi(k)$ depending on the unknown values of a stochastic sequence $\xi(k)$ with stationary increments of order *n* from observations of the sequence for the moments of time $k = -1, -2, \ldots$ is considered. Formulas for calculating mean square error and spectral characteristic of optimal linear estimation of the functional are proposed in the case where spectral density is exactly known. When the set admissible spectral densities is given, formulas that determine the least favorable spectral densities are proposed. **Keywords:** Sequence with stationary increments, minimax-robust estimation, mean square error, least favorable spectral density, minimax-robust spectral characteristic.

1 Introduction

Stochastic processes with *n*th stationary increments were being considered by A.M. Yaglom (see [15]). He described a spectral representation of these processes and a canonical representation of the spectral density, solved an extrapolation problem. An extrapolation problem of the process with stationary increments was defined and solved in term of increments of the process. Stochastic processes with *n*th stationary increments were investigated by M.S. Pinsker [11], A.M. Yaglom and M.S. Pinsker [10] as well.

Traditional methods of solution of the linear extrapolation, interpolation and filtering problems for stationary stochastic processes were proposed by A. N. Kolmogorov [5], N. Wiener [13], A. M. Yaglom [14]. When the complete information on the spectral density is not available but the set of admissible spectral densities is given, minimax estimation methods are applied. These methods consist of finding the estimation which minimizes the value of mean square error for each density in the set. The minimax approach to extrapolation problem for stationary processes was proposed by Ulf Grenander [1].

J. Franke [2], J. Franke and H. V. Poor [3] investigated the minimax extrapolation and filtering problems for stationary sequences with the help of convex optimization methods. In the papers by M. P. Moklyachuk [6] -[9] the minimax approach to extrapolation, interpolation and filtering problems was applied to the functionals that depend on the unknown values of stationary processes and sequences.

In this article we are interested in finding the estimations of the unknown values of the functional $A\xi = \sum_{k=0}^{\infty} (k)\xi(k)$ that depends on the unknown values of a stochastic sequence $\xi(k)$ with stationary increments of order n from



observations of the sequence $\xi(k)$ for $k = -1, -2, \ldots$ The extrapolation problem for the sequences with stationary increments is solved in case where the spectral density is known. In case where the spectral density is not known but the set of admissible spectral density is given, the least favorable spectral density for the optimal extrapolation of the functional $A\xi$ was found.

2 Stationary increments. Spectral representation

Definition 1. For a given stochastic sequence $\{\xi(m), m \in Z\}$ a function

$$\xi^{(n)}(m,\mu) = (1 - B_{\mu})^{n} \xi(m) = \sum_{l=0}^{n} (-1)^{l} C_{n}^{l} \xi(m - l\mu), \qquad (1)$$

where B_{μ} is a backward operator with step $\mu \in Z$, $B_{\mu}\xi(m) = \xi(m-\mu)$, is called stochastic *n*-th increment with step $\mu \in Z$.

Definition 2. Stochastic increment $\xi^{(n)}(m,\mu)$ of some stochastic sequence $\{\xi(m), m \in Z\}$ is called stationary increment in wide sense if mathematical expectations

$$\mathbf{E}\xi^{(n)}(m_0,\mu) = c^{(n)}(\mu),$$
$$\mathbf{E}\xi^{(n)}(m_0+m,\mu_1)\xi^{(n)}(m_0,\mu_2) = D^{(n)}(m,\mu_1,\mu_2)$$

exist for all $m_0, \mu, m, \mu_1, \mu_2 \in \mathbb{Z}$ and do not depend on m_0 . The function $c^{(n)}(\mu)$ is called the mean value and the function $D^{(n)}(m, \mu_1, \mu_2)$ is called the structural function of the stationary *n*th increment. Stochastic sequence $\{\xi(m), m \in \mathbb{Z}\}$ which defines a stationary *n*th increment $\xi^{(n)}(m, \mu)$ by formula (1) is called the sequence with stationary *n*th increments.

Theorem 1. The functions $c^{(n)}(\mu)$ and $D^{(n)}(m, \mu_1, \mu_2)$ can be presented in the forms

$$c^{(n)}(\mu) = c\mu^{n},$$

$$D^{(n)}(m,\mu_{1},\mu_{2}) = \int_{-\pi}^{\pi} e^{i\lambda m} (1 - e^{-i\mu_{1}\lambda})^{n} (1 - e^{i\mu_{2}\lambda})^{n} \frac{1}{\lambda^{2n}} dF(\lambda), \qquad (2)$$

where c is a constant, function $F(\lambda)$ is a left-continuous nondecreasing bounded function, $F(-\pi) = 0$. That is more the constant c and the function $F(\lambda)$ are uniquely defined by the stochastic increment $\xi^{(n)}(m,\mu)$.

Using the representation (2) of the structural function of the stationary nth increment $\xi^{(n)}(m,\mu)$ and the Karunen's theorem [4], we get the following representation of the stationary nth increment $\xi^{(n)}(m,\mu)$:

$$\xi^{(n)}(m,\mu) = \int_{-\pi}^{\pi} e^{im\lambda} (1 - e^{-i\mu\lambda})^n \frac{1}{(i\lambda)^n} dZ(\lambda), \tag{3}$$

where $Z(\lambda)$ is an orthogonal random measure on $[-\pi, \pi)$ related with the structural function $F(\lambda)$.



Let the stationary $n {\rm th}$ increment $\xi^{(n)}(m,\mu)$ admit the one sided moving average representation

$$\xi^{(n)}(m,\mu) = \sum_{k=0}^{\infty} \varphi^{(n)}(k,\mu)\varepsilon(m-k)$$
(4)

with some renovating sequence $\{\varepsilon_m : m \in Z\}$ [15] and some complex sequence $\{\varphi^{(n)}(k,\mu) : m \ge 0\}, \sum_{k=0}^{\infty} |\varphi^{(n)}(k,\mu)|^2 < \infty$. Then the spectral function $F(\lambda)$ of the increment $\xi^{(n)}(m,\mu)$ has a spectral density $f(\lambda)$ admitting a canonical factorization

$$f(\lambda) = |\Phi(e^{-i\lambda})|^2, \tag{5}$$

where the function $\Phi(z) = \sum_{k=0}^{\infty} \varphi(k) z^k$ has a convergence radius r > 1 and does not have any zero in $|z| \leq 1$. Let us define $\Phi_{\mu}(z) = \sum_{k=0}^{\infty} \varphi_{\mu}(k) z^k$ where $\varphi_{\mu}(k) = \varphi^{(n)}(k,\mu)$ are the coefficients from the representation (4). Then the following equality takes place

$$\left|\Phi_{\mu}(e^{-i\lambda})\right|^{2} = \frac{|1 - e^{-i\lambda\mu}|^{2n}}{\lambda^{2n}}f(\lambda).$$
(6)

We use the representation (4) to find the optimal mean square estimation of unknown values of the sequence with stationary increment.

3 Extrapolation of the linear functional $A\xi$

Let a stochastic sequence $\{\xi(m), m \in Z\}$ determine a stationary *n*th increment $\xi^{(n)}(m,\mu)$ with absolutely continuous spectral function $F(\lambda)$ which has spectral density $f(\lambda)$. Let the increment $\xi^{(n)}(m,\mu)$ admit the one sided moving average representation (4) and the spectral density $f(\lambda)$ admit the canonical factorization (5). From now we make the assumption that observation of the sequence $\xi(m)$ for m < 0 are known. Consider a problem of finding optimal in mean square sense linear estimation of the functional $A\xi = \sum_{k=0}^{\infty} a(k)\xi(k)$ depending on unknown values $\xi(m), m \geq 0$.

Let $\{d_{\mu}(k) : k \geq 0\}$ be the coefficients from a relation $\sum_{k=0}^{\infty} d_{\mu}(k)x^{k} = \left(\sum_{j=0}^{\infty} x^{\mu j}\right)^{n}$. The functional $A\xi$ can be presented as $A\xi = B\xi - V\xi$, where

$$B\xi = \sum_{k=0}^{\infty} b_{\mu}(k)\xi^{(n)}(k,\mu), \quad V\xi = \sum_{k=-\mu n}^{-1} v(k)\xi(k),$$
$$v_{\mu}(k) = \sum_{l=\left[-\frac{k}{\mu}\right]'}^{n} (-1)^{l} C_{n}^{l} b_{\mu}(l\mu+k), \quad k = -1, -2, \dots, -\mu n,$$
(7)

$$b_{\mu}(k) = \sum_{m=k}^{\infty} a(m)d_{\mu}(m-k) = (D^{\mu}a)_k, \ k \ge 0,$$
(8)



and [x]' denotes the least integer number among numbers greater or equal to x. D^{μ} is a linear operator defined by the elements $D^{\mu}_{k,j} = d_{\mu}(j-k)$ if $0 \le k \le j$, and $D^{\mu}_{k,j} = 0$ if j < k; vector $a = (a(0), a(1), a(2), \ldots)$.

We will suppose that the following conditions hold true

$$\sum_{k=0}^{\infty} |b_{\mu}(k)| = \sum_{k=0}^{\infty} |(D^{\mu}a)_{k}| < \infty, \ \sum_{k=0}^{\infty} (k+1)|b_{\mu}(k)|^{2} = \sum_{k=0}^{\infty} (k+1)|(D^{\mu}a)_{k}|^{2} < \infty.$$
(9)

Under these conditions the functional $B\xi$ has the second moment and the operator B^{μ} defined below is compact.

Let $\widehat{A}\xi$ denote the mean square optimal linear estimation of the functional $A\xi$ from observations of the sequence $\xi(m)$ for m < 0, $\widehat{B}\xi$ denote the mean square optimal linear estimation of the functional $B\xi$ from observations of the stochastic increment $\xi^{(n)}(m,\mu), \mu > 0$, for m < 0. Let $\Delta(f,\widehat{A}) := \mathbf{E}|A\xi - \widehat{A}\xi|^2$ and $\Delta(f,\widehat{B}) := \mathbf{E}|B\xi - \widehat{B}\xi|^2$ denote the mean square errors of the estimations $\widehat{A}\xi$ and $\widehat{B}\xi$. Since the values $\xi(m)$ for $m = -1, -2, \ldots, -\mu n$ are known, we have the equality $\widehat{A}\xi = \widehat{B}\xi - V\xi$. Therefore the following equalities take place:

$$\Delta(f,\widehat{A}) = \mathbf{E}|A\xi - \widehat{A}\xi|^2 = \mathbf{E}|A\xi + V\xi - \widehat{B}\xi|^2 = \mathbf{E}|B\xi - \widehat{B}\xi|^2 = \Delta(f,\widehat{B}).$$

Let $L_2^{0-}(f)$ be a subspace of the space $L_2(f)$ generated by the functions $\{e^{i\lambda k}: k \leq -1\}$. The optimal estimation $\widehat{B}\xi$ can be written as

$$\widehat{B}\xi = \int_{-\pi}^{\pi} h_{\mu}(\lambda)(1 - e^{-i\lambda\mu})^n \frac{1}{(i\lambda)^n} dZ(\lambda),$$
(10)

where $h_{\mu}(\lambda) \in L_2^{0-}(f)$ is a spectral characteristic which provides the minimum value of the mean square error $\Delta(f, \hat{B})$. Using the projection method of the theory of Hilbert spaces, we obtained he following formula for calculation he optimal spectral characteristic

$$h_{\mu}(\lambda) = B^{\mu}(e^{i\lambda}) - r_{\mu}(e^{i\lambda})\Phi_{\mu}^{-1}(e^{-i\lambda}),$$
(11)
$$r_{\mu}(e^{i\lambda}) = \sum_{j=0}^{\infty}\sum_{m=0}^{\infty}b_{\mu}(m+j)\varphi_{\mu}(m)e^{i\lambda j} = \sum_{j=0}^{\infty}(B^{\mu}\varphi_{\mu})_{j}e^{i\lambda j},$$

where B^{μ} is a linear symetric operator defined by the matrix with elements $B_{k,j}^{\mu} = b_{\mu}(k+j), k, j \geq 0. \quad \varphi_{\mu} = (\varphi_{\mu}(0), \varphi_{\mu}(1), \varphi_{\mu}(2), \ldots); \quad \varphi_{\mu}(k), k \geq 0$, are the coefficients from the moving average representation (4).

Further the estimation of the functional $A\xi$ can be presented in the form

$$\widehat{A}\xi = -\sum_{k=-\mu n}^{-1} v_{\mu}(k)\xi(k) + \int_{-\pi}^{\pi} h_{\mu}^{(a)}(\lambda)(1 - e^{-i\lambda\mu})^n \frac{1}{(i\lambda)^n} dZ(\lambda),$$
(12)

where the coefficients $v_{\mu(i)}$ for $i = -1, -2, \ldots, -\mu n$ are defined in (7). Using he relationship (8) between the coefficients a(k) and $b_{\mu}(k)$, we obtain the equality $(B^{\mu}\varphi_{\mu})_{k} = (D^{\mu}A\varphi_{\mu})_{k}$, where the linear operator A is defined by the coefficients


 $a(k), k \ge 0$, in the following way: $(A)_{k,j} = a(k+j), k, j \ge 0$. The spectral characteristic and the value of mean square error can be calculated by formulas

$$h_{\mu}^{(a)}(\lambda) = A(e^{i\lambda}) - r_{\mu}^{(a)}(e^{i\lambda})\Phi_{\mu}^{-1}(e^{-i\lambda}),$$
(13)

$$A(e^{i\lambda}) = \sum_{k=0}^{\infty} (D^{\mu}a)_{k} e^{i\lambda k}, \quad r_{\mu}^{(a)}(e^{i\lambda}) = \sum_{j=0}^{\infty} (D^{\mu}A\varphi_{\mu})_{j} e^{i\lambda j}.$$
 (14)

$$\Delta(f, \widehat{A}) = \mathbf{E} |A\xi - \widehat{A}\xi|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |r_{\mu}^{(a)}(e^{i\lambda})|^2 d\lambda = ||D^{\mu}A\varphi_{\mu}||^2.$$
(15)

Theorem 2. Let the stochastic sequence $\{\xi(m), m \in Z\}$ generate the stationary random increments $\xi^{(n)}(m,\mu), \mu > 0$, with absolutely continuous structural function $F(\lambda)$ and spectral density $f(\lambda)$ admitting the canonical factorization (5). If conditions (9) hold true, the optimal linear estimation $\widehat{A}\xi$ of the functional $A\xi$ of unobserved values $\xi(m), m \ge 0$, from observations of the sequence $\xi(m), m = -1, -2, \ldots$, can be calculated by formula (12). Spectral characteristic $h_{\mu}^{(a)}(\lambda)$ of the optimal linear estimation $\widehat{A}\xi$ can be calculated by formula (13). The value of mean square error $\Delta(f, \widehat{A})$ is calculated by formula (15).

4 Minimax-robust extrapolation

If the spectral density $f(\lambda)$ of the sequence $\xi(m)$ with stationary *n*th increments is known, the value of mean square error $\Delta(h_{\mu}^{(a)}(f); f) := \Delta(f, \widehat{A})$ and the spectral characteristic $h_{\mu}^{(a)}$ of the optimal linear estimation $\widehat{A}\xi$ of the functional $A\xi$ can be calculated by formulas (13) and (15). In the case where the spectral density is not known, but the only set \mathcal{D} of possible spectral densities is given, the minimax (robust) approach to estimation of functionals of the unknown values of random sequence with stationary increments is reasonable. In other words we find an estimation that minimizes the mean square error for all spectral densities from the given class \mathcal{D} simultaneously.

Definition 3. For a given class of spectral densities \mathcal{D} spectral density $f_0(\lambda) \in \mathcal{D}$ is called least favorable in \mathcal{D} for the optimal linear estimate of the functional $A\xi$ if the following relation holds true

$$\Delta(f_0) = \Delta(h_{\mu}^{(a)}(f_0); f_0) = \max_{f \in \mathcal{D}} \Delta(h_{\mu}^{(a)}(f); f).$$

Definition 4. For a given class of spectral densities \mathcal{D} spectral characteristic $h^0(\lambda)$ of the optimal linear estimate of the functional $A\xi$ is called minimaxrobust if there are satisfied conditions

$$h^{0}(\lambda) \in H_{\mathcal{D}} = \bigcap_{f \in \mathcal{D}} L_{2}^{0-}(f), \quad \min_{h \in H_{\mathcal{D}}} \max_{f \in \mathcal{D}} \Delta(h; f) = \sup_{f \in \mathcal{D}} \Delta(h^{0}; f).$$



Lemma 1. Spectral density f^0 which admits factorization (5) is the least favorable in \mathcal{D} for the optimal linear estimate of the functional $A\xi$ if

$$f^{0}(\lambda) = \left| \sum_{k=0}^{\infty} \varphi^{0}(k) e^{-i\lambda k} \right|^{2}, \qquad (16)$$

where $\varphi^0 = \{\varphi^0(k) : k = 0, 1, 2, ...\}$ is a solution to the extremum problem

$$||D^{\mu}A\varphi_{\mu}||^{2} \to \max, \quad f(\lambda) = \left|\sum_{k=0}^{\infty} \varphi(k)e^{-i\lambda k}\right|^{2} \in \mathcal{D}.$$
 (17)

If $h^{(a)}_{\mu}(f_0) \in H_{\mathcal{D}}$, the minimax-robust spectral characteristic can be calculated by the formula $h^0 = h^{(a)}_{\mu}(f_0)$. The functions h^0 and f_0 form a saddle point of the function $\Delta(h; f)$ on the set $H_{\mathcal{D}} \times \mathcal{D}$. The saddle point inequalities

$$\Delta(h; f_0) \ge \Delta(h^0; f_0) \ge \Delta(h^0; f) \quad \forall f \in \mathcal{D}, \forall h \in H_{\mathcal{D}}$$

hold when $h^0 = h^{(a)}_{\mu}(f_0)$ and $h^{(a)}_{\mu}(f_0) \in H_{\mathcal{D}}$ where f_0 is a solution to the conditional extremum problem

$$\widetilde{\Delta}(f) = -\Delta(h_{\mu}^{(a)}(f_0); f) \to \inf, \quad f \in \mathcal{D},$$

$$\Delta(h_{\mu}^{(a)}(f_0); f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|r_{\mu}(e^{i\lambda})|^2}{f_0(\lambda)} f(\lambda) d\lambda,$$
(18)

where r_{μ} is defined by the formula (14) when $f(\lambda) = f_0(\lambda)$. The previous problem is equivalent to the unconditional extremum problem

$$\Delta_{\mathcal{D}}(f) = \widetilde{\Delta}(f) + \delta(f|\mathcal{D}) \to \inf.$$

A solution f_0 to this unconditional extremum problem is characterized by the condition $0 \in \partial \Delta_{\mathcal{D}}(f_0)$ [12].

5 Least favorable spectral densities in the class \mathcal{D}_0

Consider the following set of spectral densities

$$\mathcal{D}_0 = \left\{ f(\lambda) | \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) d\lambda \le P_0 \right\}.$$

Using the condition $0 \in \partial \Delta_{\mathcal{D}}(f_0)$, we get an equation $|r_{\mu}^{(a)}(e^{i\lambda})|^2 (f^0(\lambda))^{-1} = \psi(\lambda) + c^{-2}$, where $\psi(\lambda) \leq 0$ and $\psi(\lambda) = 0$ when $f^0 > 0$. Therefore the least favorable density in the class \mathcal{D}_0 for optimal linear estimate of the functional $A\xi$ can be presented as

$$f^{0}(\lambda) = \left| c \sum_{k=0}^{\infty} (D^{\mu} A \varphi_{\mu})_{k} e^{i\lambda k} \right|^{2}, \qquad (19)$$



where the unknown parameters $c, \varphi_{\mu} = (\varphi_{\mu}(0), \varphi_{\mu}(1), \varphi_{\mu}(2), \ldots)$ are calculated using factorization (5), (17) and $\int_{-\pi}^{\pi} f(\lambda) d\lambda = 2\pi P_0$.

Let us consider the following equation

$$D^{\mu}A\varphi_{\mu} = \alpha\overline{\varphi}, \quad \alpha \in R.$$
(20)

For each solution of it such that $||\varphi||^2 = P_0$ the following equality holds true

$$f^{0}(\lambda) = \left|\sum_{k=0}^{\infty} \varphi(k) e^{i\lambda k}\right|^{2} = \left|c\sum_{k=0}^{\infty} (D^{\mu}A\varphi_{\mu})_{k} e^{i\lambda k}\right|^{2}.$$

Let $\nu_0 P_0$ be the maximum value of $||D^{\mu}A\varphi_{\mu}||^2$ on the set of solutions of the equation (20) that satisfy condition $||\varphi||^2 = P_0$ and define canonical factorization of the spectral density $f(\lambda)$. Let $\nu_0^+ P_0$ be the maximum value of $||D^{\mu}A\varphi_{\mu}||^2$ on the set of such φ that satisfy condition $||\varphi||^2 = P_0$ and define canonical factorization of the spectral density $f^0(\lambda)$. Consequently, the following statements holds true.

Theorem 3. If there exists a sequence $\varphi^0 = \{\varphi^0(m) : m \ge 0\}$ that satisfies conditions $||\varphi^0||^2 = P_0$ and $\nu_0 P_0 = \nu_0^+ P_0 = ||D^\mu A \varphi^0_\mu||^2$, the spectral density (16) is the least favorable in the class \mathcal{D}_0 for optimal extrapolation of the functional $A\xi$ of unknown values $\xi(k)$, k = 0, 1, 2..., of the stochastic sequence with stationary nth increments. The increment $\xi^{(n)}(m,\mu)$ admits the one side moving average representation. If $\nu_0 < \nu_0^+$, the density (19) which admits the canonical factorization (5) is the least favorable in the class \mathcal{D}_0 . Sequence $c\varphi_\mu = \{c\varphi_\mu(k) : k \ge 0\}$ is defined by conditions (17) and $\int_{-\pi}^{\pi} f(\lambda) d\lambda = 2\pi P_0$.

6 Least favorable spectral densities in the class \mathcal{D}_M

Consider the following set of spectral densities

$$\mathcal{D}_M = \left\{ f(\lambda) | \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) \cos(m\lambda) d\lambda = \rho_m, \ m = 0, 1, 2, \dots, M \right\},$$

where $\rho_0 = P_0$ and $\{\rho_m, m = 0, 1, 2, \dots, M\}$ form a strictly positive sequence. From condition $0 \in \partial \Delta_{\mathcal{D}}(f_0)$ we get the equation $|r_{\mu}^{(a)}(e^{i\lambda})|^2 (f^0(\lambda))^{-1} = \psi(\lambda) + c \sum_{m=1}^{M} \psi_m \cos m\lambda$. Thus the list favorable density in the class \mathcal{D}_M for the optimal linear estimate of the functional $A\xi$ can be presented as

$$f^{0}(\lambda) = \frac{\left|c_{0}\sum_{k=0}^{\infty} (D^{\mu}A\varphi_{\mu})_{k}e^{i\lambda k}\right|^{2}}{\left|\sum_{k=1}^{M} c_{m}e^{-i\lambda k}\right|^{2}},$$
(21)

where the parameters c_m , m = 0, 1, 2, ..., M, $\varphi_{\mu} = (\varphi_{\mu}(0), \varphi_{\mu}(1), \varphi_{\mu}(2), ...)$ can be calculated using conditions (17) and $\int_{-\pi}^{\pi} f(\lambda) \cos(m\lambda) d\lambda = 2\pi \rho_m$, m = 0, 1, 2, ..., M, factorization (5).

Let $\nu_M P_0$ be the maximum value of $||D^{\mu}A\varphi_{\mu}||^2$ on the set of solutions of the equation (20) that satisfy condition $||\varphi||^2 = P_0$ and define canonical



factorization of the spectral density $f(\lambda)$. Let $\nu_M^+ P_0$ be the maximum value of $||D^{\mu}A\varphi_{\mu}||^2$ on the set of such φ that satisfy condition $||\varphi||^2 = P_0$ and define canonical factorization of the spectral density $f^0(\lambda) \in \mathcal{D}_M$ defined by (21).

Theorem 4. If there exists a sequence $\varphi^0 = \{\varphi^0(m) : m \ge 0\}$ that satisfies conditions $||\varphi^0||^2 = P_0$ and $\nu_0 P_0 = \nu_M^+ P_0 = ||D^\mu A \varphi^0_\mu||^2$, the spectral density (16) is the least favorable in the class \mathcal{D}_M for optimal extrapolation of the functional $A\xi$ of unknown values $\xi(k)$, k = 0, 1, 2..., of the stochastic sequence with stationary nth increments. If $\nu_M < \nu_M^+$, the density (21), which admits the canonical factorization (5), is the least favorable in the class \mathcal{D}_M . Sequence $\varphi_\mu = \{\varphi_\mu(k) : k \ge 0\}$ and unknown parameters c_m , m = 0, 1, 2, ..., M, are defined by conditions (17) and $\int_{-\pi}^{\pi} f(\lambda) \cos(m\lambda) d\lambda = 2\pi\rho_m$, m = 0, 1, 2, ..., M.

References

- 1.U. GRENANDER, A prediction problem in game theory, Ark. Mat. 3(1957), 371-379.
- 2.J. FRANKE, Minimax robust prediction of discrete time series, Z. Wahrsch. Verw. Gebiete **68**(1985), 337-364.
- 3.J. FRANKE AND H. V. POOR, Minimax-robust filtering and finite-length robust predictors, Robust and Nonlinear Time Series Analysis. Lecture Notes in Statistics, Springer-Verlag, 26(1984), 87-126.
- 4.K. KARHUNEN, Uber lineare Methoden in der Wahrscheinlichkeitsrechnung, Ann. Acad. Sci. Fenn., Ser. A I, No.37, 1947.
- 5.A. N. KOLMOGOROV, Selected works of A. N. Kolmogorov. Vol. II: Probability theory and mathematical statistics., Ed. by A. N. Shiryayev. Mathematics and Its Applications. Soviet Series. 26. Dordrecht etc.: Kluwer Academic Publishers, 1992.
- 6.M. P. MOKLYACHUK, Stochastic autoregressive sequences and minimax interpolation, Theory Probab. Math. Stat., 48(1994), 95-103.
- 7.M. P. MOKLYACHUK, Robust procedures in time series analysis, Theory Stoch. Process. **6(22)**(2000), no. 3-4, 127-147.
- 8.M. P. MOKLYACHUK, Game theory and convex optimization methods in robust estimation problems. Theory Stoch. Process. **7(23)**(2001), no. 1-2, 253-264.
- 9.M. P. MOKLYACHUK, Robast estimations of functionals of random processes, Vydavnycho-Poligrafichnyi Tsentr, Kyivskyi Universytet, yiv, 2008.
- 10.M. S. PINSKER, A. M. YAGLOM, On linear extrapolation of random processes with nth stationary incremens, DAN USSR, 94(1954), no. 3, 385388.
- 11.M. S. PINSKER, The theory of curves with nth stationary incremens in Hilber spaces, IAN USSR, 19(1955), no. 3, 319-344.
- 12.B. N. PSHENICHNYI, *Necessary conditions for an extremum*, ' '2nd ed., Moscow, Nauka, 1982.
- 13.N. WIENER, Extrapolation, Interpolation and Smoothing of stationary Time Series. Whis Engineering Applications, The M. I. T. Press, Massachusetts Institute of Technology, Cambridge, Mass., 1966.
- 14.A. M. YAGLOM, Correlation theory of stationary and related random functions. Vol. 1,2: Basic results, Springer Series in Statistics, Springer-Verlag, New York etc., 1987.
- 15.A. M. YAGLOM Correlation theory of stationary and related random processes with stationary nth increments, Math. collection **37(79)**(1955), no. 1, 141-196.



The use of MDS and HCA enabled pharmacists to reveal their roles which reflect on country of practice and cultural differences when improving patients' adherence to asthma medication

Andrea Manfrin¹, Cecilio Mar Molinero², Luca Cacciolatti²

¹ Medway School of Pharmacy University of Kent, Central Avenue, ME4 4TB Chatham Maritime, Kent UK.

(Email: <u>a.manfrin@kent.ac.uk</u>)

² Kent Business School, University of Kent, CT2 7PE Canterbury, UK

(Email: <u>cm235@kent.ac.uk</u>), (Email: l.cacciolatti@kent.ac.uk)

Abstract: Pharmacists are health care professionals who have a lot more to offer to the public rather than just selling pills. Pharmacists live a dichotomy role: health care consultants versus shop keepers. The aim of this research was to assess how pharmacists intervenes improving patient adherence to asthma medication and to see how pharmacists perceived themselves. The research was conducted by means of a multiple choice questionnaire that was completed by 551 pharmacists. The replies were analysed using Ordinal Multidimensional Scaling and Cluster Analysis, with the routine PROXSCAL. Pharmacists revealed their roles in terms of four dimensions: a community health orientation, theoretical-applied drug orientation, the degree of involvement with the patient and the role of the pharmacist as a consultant. These dimensions have different salience in each one of the countries.

Keywords: pharmacy, country, asthma, adherence multidimensional scaling.

1 Introduction

In Europe, over 30 million people have asthma. The total annual cost of asthma was estimated in 2003 to be $\notin 17.7$ billion (£15 billion) in Europe (Loddenkemper et al 2003)¹. The pharmacist is usually the last health professional in contact with such patients giving advice, and will often be the first point of contact in the detection of adverse drug reactions (ADRs). One of the key tasks set for community pharmacists is to improve patients' knowledge of their disease, and to give advice on the use of medicines in order to increase the effectiveness of therapy. However, although there is a common role to all pharmacists, there appears not to be a cross-country study of cultural differences in pharmacy management. We report here on a study aimed at filling this void. The aim of the study was to investigate the activities of community pharmacist in relation to asthma medication. The research instrument was an on-line survey of community pharmacists in Italy, Switzerland, and the UK. In total, 551 pharmacists replied to the survey. Answers to survey questions were coded in the form of keywords, and analysed by means of multivariate statistical techniques, in particular Multidimensional Scaling (MDS), and Cluster Analysis. MDS was chosen because it allows the visualisation of the results, while keeping a solid mathematical basis. After this section we report on the way in which the questionnaire was designed. This is followed by a technical section in which, after giving summary statistics, we explain how the MDS model was implemented and interpreted. The paper ends with a discussion of the results.

2 Terminologies

There are two technical terms relating to pharmacy practice that need to be first defined: adherence and compliance. Adherence is the extent to which the patient's behaviour matches agreed recommendation from the prescriber. Compliance is the extent to which a patient's behaviour matches the prescriber's recommendations (NCCSDO, 2005)². The difference between the two terms relies on the fact that compliance denotes a relationship where the doctor decide on the appropriate treatment (authoritarian role) and the patient follows the doctor's order (passive role), while adherence emphasizes the concept of an agreement between doctor and patient.

3 Aims

The aim of the research was to explore community pharmacist's role (activity) in relation to encouraging adherence to prescribed medication for patients with asthma, and how these roles are culture and country dependent.

4 The research instrument (questionnaire)



The questionnaire had 11 questions, the first asked about the country in which the practice was situated, as one of the aims of the research was to explore international differences in pharmacy practice. Question 2 was aimed at evaluating what pharmacists understood as adherence, and to exploring whether they understood the difference between compliance and adherence. Questions 5 and 9 tried to establish how pharmacists identified patients that were non-adherent with their asthma medications and the most common reasons why patients did not take their asthma medications properly. Question 4 was set to establish whether pharmacists gave advice on the correct inhaler technique only to new patients, only to established patients, both or neither. Question 6 was aimed at providing information on what tools pharmacists have at their disposal to improve and encourage patients' correct use of their asthma medication. Questions 7 and 10 explored pharmacists' perceived level of involvement in improving patients' use of medicines. Questions 3 and 8 looked at pharmacists' level of confidence in giving advice on inhaler techniques and at their perceptions on the result of such advice. The last question, Question 11, was aimed to understanding, which were for pharmacists, the three most positive outcomes of improving patient's adherence. The questionnaire was distributed in November 2009, and responses were collected until the end of March 2010. It was designed with three language options: English, French and Italian. It was sent through Qualtrics® system to UK Swiss and Italian pharmacists. The MUR (Medicine Use Review) database held by the Medway School of Pharmacy (UK) was used to distribute the questionnaire in the UK. In Italy and Switzerland the survey was e-mailed through the pharmaceutical associations of Pordenone (Friuli Venezia Giulia Region, Italy) and Canton Ticino in Switzerland. One reminder was sent to UK pharmacists but no reminders were sent to Swiss and Italian chemists. Of the 6611 emails that were sent to UK pharmacists, 370 replied giving a response rate of 5.6%. In Switzerland, 38 pharmacists answered the survey giving a response rate of 10% and in Italy 103 replied giving a response rate of 27%. Country of origin could not be established in 40 questionnaires.

5 Data analysis

Responses to questions were coded into the package PASW, version 19 (previously SPSS). Each respondent was treated as a case. A set of keywords was derived from the replies to the questions. All keywords were coded with the values zero or one; one, if the respondent had replied positively to the question, and zero otherwise. When a question had several possible answers, a keyword was allocated to each option. The data set was a matrix of 551 questionnaires by 58 keywords. All entries in this matrix were either zero or one. Missing country values were coded as zero.

5.1 Multidimensional scaling

The data set of pharmacies by keywords was analysed by means of Ordinal Multidimensional Scaling (MDS) with the routine PROXSCAL, although standard menus could not be used, and a routine was coded using the syntax facility. Our interest was not in comparing pharmacists, but in finding out if there were patterns in the way the questions were answered by pharmacists. We were particularly interested in finding out which questions were relevant to particular countries and not to others. MDS is a multivariate statistical technique that starts by calculating a measure of similarity between the entities of interest, the keywords. The measure of similarity used was proposed by Russell and Rao (Yin and Yasuda, 2005)³. The Russell and Rao measure assesses the "proximity" between two keywords by counting the number of times that respondents had replied "yes" to both keywords. This procedure is similar to Mar Molinero and Xie $(2007)^4$. A table of similarities (proximities) between pairs of keywords was obtained in this way. The next step was to plot the keywords in the space in such a way that when the measure of similarity between two keywords is high, the keywords are located next to each other in the space. When the similarity between two keywords is low, they are located far apart in the space. The space in which the keywords are located may have more than two dimensions, and it is necessary to work with projections on sets of two dimensions. When working with more than two dimensions, it is necessary to complement MDS with other statistical techniques in order to interpret the results. The quality of the statistical representation varies with the number of dimensions of the space in which the keywords are located. The higher the number of dimensions the better the quality of fit. There are various ways of assessing goodness of fit. Stress1 was chosen as the measure of quality of fit. Stress1 is based on the proportion of unexplained variance. We used the ordinal approach to build the configuration, untying tied observations, $(Coxon, 1982)^5$. The value of Stress1 was calculated for six dimensions: D6 (Stress 1= 0.109), D5 (Stress 1 = 0.124), D4 (Stress 1 = 0.139), D3 (Stress 1 = 0.173), D2 (Stress 1 = 0.224), and D1 (Stress 1 = 0.331). As the number of dimensions increases, the value of Stress1 decreases. The values of Stress1 in D5 and D6 did not improve massively and for this reason rather than estimate the model again in four dimensions, we used the first four dimensions of the six dimensional representations. The projection of



the six dimensional map in D1 and 2 can be seen in Figure 1, and the projection of the six dimensional map in D3 and 4 is reproduced in Figure 2.



Figure 1. Representation in Dimensions 1 and 2 with Ward cluster



Figure 2. Representation in Dimensions 3 and 4 with Ward cluster membership

It is now possible to attach meaning to the dimensions in the plot. We will start with D1. On the far left hand side of Figure 1 we find that pharmacists find out if patients are not using the medicine properly



reviewing asthma management plan, checking PMR (patient medication record), looking for "Sign and Symptoms"; they intervene assessing patient inspiration capability and minimizing adverse drug reaction. From the centre to the left side of the D1 axis we found pharmacists that strongly agree and agree to the fact that is the responsibility of the pharmacist to ensure that the patient takes the medicine properly; and that the pharmacist thinks that a positive outcome of improving patient adherence is to "minimise adverse drug reaction" as well. At the right hand side of Figure 1 we find a disagreement with the view that the pharmacist has to consider patient's adherence more often; and, for example, that when the pharmacist discovers that the patient is not taking the medicine correctly, takes the action of "signposting to another health care professional"; that pharmacists neither agree nor disagree regarding to the fact that it is their "responsibility to ensure patients use their medication properly" and neither agree nor disagree that they "need to consider patient adherence more often". This dimension appears to be discriminating between pharmacists that see their role as community health professionals, and pharmacists that see themselves as just in charge of dispensing medicines as prescribed by the medical doctor. Using similar considerations we interpreted D2 as pharmacists mainly looking at drug from the theoretical point of view (bottom) which means a general knowledge of drug such as frequency of dosing, to a pharmacists looking at drug under the application point of view (patients' implication) identifying specific issues such as assessing patient inhalation technique, poor inspiration capability. D2 is associated with a higher drug orientation form theoretical knowledge (negative side of the axis) to applied knowledge (positive side of the axis) about drugs. It appears also very clear the link between MUR and UK pharmacists which reinforce the concept of the applied knowledge as well. The presence of the keyword "PMR" (patient medication history) is reinforcing the concept of the knowledge which is applied in a real situation. UK and Swiss pharmacists have PMR in place, but Italian pharmacists are not allowed to keep PMR due data protection regulation. Dimension 3 was found to be related to the degree of pharmacists' involvement in the management of the disease: on the negative side of this axis, the pharmacists are actively involved (Pro-Active), assessing and reviewing which means taking the initiative; on the right hand side, the pharmacists tend to state why poor adherence can occur (Re-Active). Dimension 4 appears to be related to the extent to the role of the pharmacist as a health consultant, the lower end of this dimension being associated with a higher involvement in health advice: health consultant. The opposite end is showing pharmacists who do not tend to get engaged with patients (shopkeeper behaviour). The MDS analysis has revealed that pharmacists see their role in four dimensions: as community health professionals prescription dispenser; as people with theoretical-applied knowledge; in terms of their relationship with the patient; and as health consultants. We can see in Figures 1 and 3 that the keyword "Swiss" is located towards the left hand side of D1, and just below the middle of D2, towards the left of D3, and towards the negative side of D4. This would indicate that Swiss pharmacists can be described as community health oriented (D1), with a slightly heavier component of theoretical knowledge, with pro-active attitude involved in the management of the patient (D3), perceiving themselves health consultant. The keyword "UK" is in the middle of D1, towards the positive side of D2, it is also in the middle of D3 and just at the beginning of the negative value of Dimension. This could be explained by the fact that UK pharmacists perceived themselves across two systems: being involved in the management of the patient and dispensing prescriptions where the weight of the two components seems to be balanced. The presence of UK in the positive area o D2 underpins the focus of UK pharmacists on the applied knowledge. They still are in the negative part of D4, but they perceived themselves less health care consultant compared to their Swiss colleagues. With regards to the Italian pharmacists they appeared to be salient in D1 and very close to the middle of this dimension but slightly skewed to the right. The location of Italian pharmacists in D2 is clear: the theoretical knowledge is predominant versus the applied knowledge. D3 shows the same level of involvement in patient management such as the Swiss; D4 placed them slightly below the UK but far from the Swiss meaning that they did not perceive themselves having a consultant role such as their Swiss colleagues.

5.2 Cluster Analysis

Having attached meaning to the dimensions, and having used these meanings in order to explore the health management culture of pharmacists in the three countries from which information was obtained, it is also relevant to establish if there are questions, or groups of questions, that are answered in the same way by pharmacists. We are particularly interested in knowing which positive answers to questions serve to differentiate pharmacists in the three countries. To assess the proximity of the points in the space we used Cluster Analysis. Distances between keywords in the six dimensional spaces were calculated from the coordinates in the MDS configuration using the Euclidean metric. We used Hierarchical Cluster Analysis with the method suggested by Ward, since it maximises the homogeneity within clusters and the heterogeneity between clusters. The output of the analysis is a dendrogram and it is a matter of



judgement to decide how many clusters are relevant. One is looking for "long branches" in the diagram, since these reveal the heterogeneity of the clusters. As the branches become shorter the clusters increase in homogeneity, but also in number. In the end we settled for six clusters. Cluster 6 and Cluster 5 stand out as discordant from the rest. They both relate to single variables, and summarise responses from very few pharmacists. There will not be any further discussion on these two clusters. Cluster 1 group together very many answers to questions related to the pharmacist as a health care provider. The questions included are: "How do you find out if a patient is not using their asthma medication properly" which reflect the patient's attitude towards poor adherence. Very close to this are the answers to the question "How would you intervene when you know a patient is not using their medication properly" which represent the way in which the pharmacists is acting to improve patient adherence to medication. Conducting Medicine Use Review (MUR) appears in this cluster despite the fact that MUR was a formalised service only in UK at the time of this study. In this cluster there is a strong representation of another question: "What do you think are the reasons patient may not take their medication properly". This question is particularly relevant because it shows that pharmacists believe that non-intentional non adherence is very common among patients and it is represented by forgetfulness, misunderstanding, and lack of knowledge. The presence of the last question "What do you think are the 3 major positive outcomes of improving patient adherence" was answered positively by pharmacists who selected improving quality of life, reduce mortality rate, and minimize adverse drug reaction (ADR). Pharmacists located in the UK are represented in this cluster. Cluster 2 relates to the pharmacists' attitude with respect to the patients' problems. This cluster includes statement which shows what pharmacists think are the reasons patients may not take their medication properly, which are frequency of dosing and poor inspiration capability. In this cluster we find Swiss pharmacists. Cluster 3 group pharmacists who misunderstood the meaning of adherence and confused with compliance. They tend to give advice to continuing patient and they confirmed that they need to improve their confidence when giving advice about inhaler technique. Other variable found in this cluster where SA10 and SA7 which are indicating the level of agreement with the fact that pharmacists needs to get involved in medicine management. This confirms the positive pharmacists' attitude towards patient issues. Here we find Italian pharmacists. Finally, the questions that come together under Custer 4 are related to the way in which pharmacists assess if patients are not using their medication properly, and the way in which the pharmacists' act when they find out. Pharmacists who belong to this cluster tend to review asthma management plan and assess patient inspiration capability.

5.3 Classification of pharmacists

Figures 1 and 2 bring together the MDS representation of keywords and their classification using Cluster Analysis. It is to be noticed that there is little overlap between the clusters, something that facilitates interpretation. We can see in figure 1 that Cluster 1 is mainly located at the centre of the axis and UK pharmacists are represented in it. They seem to have good applied knowledge and rely in between the health professional and the pharmacists involved in the dispensing procedures as well. In figure 2 Cluster 1 is mainly concentrated in the lower part of D4 (health consultant) but it is also between the pro-active and re-active dimension. Probably UK pharmacists are acting according to patients' requirement. In figure 1, cluster 2 is located on the negative axis of D1 but it seems to be present in either part of D2. This revels that Swiss pharmacists see themselves as health professionals where the knowledge seems to be spread across theoretical and applied. Figure 3 does not revel very much about cluster although it is mainly present in the negative part of D4 (health consultant). Cluster 3 appears evident in figure 2 and represents the Italian pharmacists; the theoretical knowledge is the dominant element in this cluster and they perceive themselves between health professional and dispensing activity. The dispensing component seems slightly heavier compared to the UK ones. This cluster is also present in figure 3, it is well represented in the pro-active part of D3 (negative axis) and in the health consultant (negative axis) of D4.

6 Conclusions

The visual interpretation has been able to demonstrate the way in which pharmacists reveal themselves through the questionnaire, and the cultural differences that exist in these three countries. There is much debate on the role of pharmacists as a keystone in the provision of heath. The WHO (World Health Organization) together with FIP (International Pharmaceutical Federation) discussed the role that pharmacists should aim for, and observed that pharmacists have been always linked with manufacturing and supplying of medications, but due to the increasing demand for health services, they are facing new and challenging roles. Such future roles were discussed in a conference on the 23rd November 2006 in Geneva (Switzerland) where it was observed that the complexity of new medications coupled with poor



adherence to prescribed medications is contributing to a shift of the pharmacy model from drug-based to patient-based; Wiedenmayer et al, 2006. The above suggests that there are various ways in which pharmacy can be practiced. It is therefore appropriate to ask up to what point differences in the way in which the practice of pharmacy takes place can be discerned by looking at different cultural settings. This issue has been explored by means of a questionnaire completed by pharmacists in the UK, Italy, and Switzerland. The analysis of questionnaire answers revealed that there are four dimensions to pharmacy: a community health orientation (health professional-dispense medication); theoretical-applied knowledge; the degree of involvement with the patient (pro-active – re-active); and the role of the pharmacist as a consultant-shopkeeper. It was also identified the positive answers to questions that are relevant to each type of pharmacist. It was observed that Swiss pharmacists can be described as health professionals, community oriented, where the knowledge seems to be spread across theoretical (slightly heavier) and applied and they tend to act as consultant. UK pharmacists have shown good applied knowledge and were place between the health professional and pharmacists involved in the dispensing procedures. They did not show a clear shift towards pro-active or re-active role, but they perceive themselves as health consultant too. Italian pharmacists appear to be the more shifted towards the theoretical knowledge but also to dispensing. They perceive themselves involved in patient management and so pro-active acting as health consultant. The health consultant shift is really evident in the Swiss pharmacists and less evident in the Italian and UK pharmacists respectively. This study was limited to a (relatively) small sample of pharmacists in only three countries. Further research needs to be conducted involving a larger number of pharmacists from more countries in order to fully understand and share best practice in order to achieve a patient centred pharmacy model.

References

1.Loddenkemper R, Gibson GJ, Sibille Y. (2003) European Respiratory Society/European Lung Foundation European Lung White Book. Sheffield, ERSJ., pp. 16-25.

2.National Co-ordinating Centre for NHS Service Delivery and Organisation R & D (NCCSDO) 2005. Concordance, adherence and compliance in medicine taking [Online] Available at http://www.medslearning.leeds.ac.uk/pages/documents/useful_docs/76-final-report%5B1%5D.pdf [Accessed 6th January 2010]

3.Yin, Y. and Yasuda, K. (2005) Similarity coefficient methods applied to the cell formation problem: a comparative investigation. Computers and Industrial Engineering 48: 471-489.

4.Mar Molinero C. and Xie A. (2007). What do UK employers want from OR/MS. Journal of the Operational Research Society, 58, 1543-1553.

5.Coxon, A.P.M. (1982) The Users Guide to multidimensional scaling, London: Heinemann.

6. Wiedenmayer, K; Summers, RS; Mackie, CA; Gous, AGS; Everard, M; Tromp, D (2006) Developing pharmacy practice: a focus on patient care. World Health Organisation, Geneva, Switzerland.



Optimal Chattering Regimes in Nonhomogeneous Bar Model

Larisa Manita

National Research University Higher School of Economics, Moscow Institute of Electronics and Mathematics, Bolshoy Trehsviatitelskiy Per. 3, 109028 Moscow, Russia. (E-mail: lmanita@hse.ru)

Abstract. We consider an optimal control problem of longitudinal vibrations of a nonhomogeneous bar with clamped ends. We assume that the density of external forces is a control function. Using the Fourier method we prove that the optimal control is the chattering control, i.e., it has an infinite number of switchings in a finite time interval.

Keywords: Vibrations, Optimal control, Singular solutions, Nonhomogeneous bar.

1 Introduction

Consider small longitudinal vibrations of a nonhomogeneous bar of length l. The longitudinal displacement at a typical point x is denoted y(t, x) where t is the time. Let g(x, t) be a density of external longitudinal force at the instant of time t at the point x. Suppose that g(t, x) = u(t)f(x) where the force profile function f(x) is assumed to be given, u(t) is the control function. Assume that

$$-1 \le u\left(t\right) \le 1 \tag{1}$$

The equation of longitudinal vibrations of the bar can be written as

$$p(x) y_{tt}(t,x) - (k(x) y_x(t,x))_x = u(t)f(x)$$
(2)

Here $p(x) = \rho(x) S(x)$, k(x) = E(x) S(x), $\rho(x)$ is the density of bar, S(x) is the cross-sectional area, E(x) is the Young's modulus at x, see, for example, [1–3]. We assume that the ends of the bar are clamped:

$$y|_{x=0} = y|_{x=l} = 0, \quad t > 0 \tag{3}$$

and the initial position and velocity are fixed:

$$y|_{t=0} = y_0(x), \quad y_t|_{t=0} = y_1(x), \qquad x \in [0, l]$$
 (4)

We suppose that the coefficient functions k, p are smooth enough and

$$k(x) \ge k_0 > 0, \quad p(x) \ge p_0 > 0 \quad \forall x \in [0, l]$$
 (5)

We consider an optimal control problem: to find such a control function u(t) that minimize the following mean square functional

$$\int_{0}^{\infty} \int_{0}^{l} p(x) y^{2}(t, x) dx dt \to \inf$$
(6)



under (1) - (5).

The problems of longitudinal vibrations of a bar were considered in [1-3]. In [1,2] the dynamics of the longitudinal vibrations of a bar subjected to viscous boundary conditions was studied. The optimal boundary control problem for longitudinal vibrations of a bar was considered in [3]. By using a maximum principle the optimal control was expressed in terms of an adjoint variable.

In this paper for the problem of controlling the longitudinal vibrations of a bar (1)-(6) we construct a solution y(t, x) in the form

$$y(t,x) = \sum_{j=1}^{\infty} s_j(t)h_j(x) \tag{7}$$

where $\{h_j(x)\}_{j=1}^{\infty}$ are eigenfunctions of the Sturm-Liouville problem, $\{s_j(t)\}_{j=1}^{\infty}$ are corresponding Fourier coefficients. To find Fourier coefficients we consider an optimal control problem in the space l^2 . For the control problem in l^2 we show that the optimal solutions contain singular trajectories and chattering trajectories. A trajectory is called a chattering trajectory if it has an infinite number of a control switchings on a finite time interval. By similar method we studied the optimal control problem for a rotating uniform Timoshenko beam [4,5]. But for the Timoshenko beam similar results hold only for a dense set of initial conditions in the space l^2 .

2 Optimal control problem in l_2

Define an operator L in $C^2([0,l])$ by $Lh = (kh_x)_x$. Consider the following Sturm-Liouville eigenvalue problem with Dirichlet boundary conditions:

$$Lh + \lambda p(x)h = 0, \quad x \in (0, l), \qquad h(0) = 0, \quad h(l) = 0$$
 (8)

Here the functions k(x) and p(x) satisfy (5). It is known (see [6]) that the problem (8) has an infinite sequence of eigenvalues $\{\lambda_j\}_{j=1}^{\infty}$, which are simple and positive: $0 < \lambda_1 < \lambda_2 < \ldots$, $\lambda_j \to \infty$, $j \to \infty$. To each eigenvalue λ_j corresponds a single eigenfunction h_j , and the sequence of eigenfunctions $\{h_j(x)\}_{j=1}^{\infty}$ forms an orthonormal basis of $L_2((0, l); p)$ with the inner product $(z, w)_p = \int_0^l p(x) z(x) w(x) dx$. If k(x), p(x) are smooth enough (for example, $k', p \in C^1([0, l]))$ and the condition (5) holds, then the eigenvalues λ_j admit the asymptotic form [7–9]:

$$\frac{\lambda_j}{j^2} \sim \pi^2 \left(\int_0^l \sqrt{p(x)/k(x)} dx \right)^{-2}, \quad j \to \infty$$
(9)

Assume that $y(t, \cdot) \in L_2((0, l); p)$. For any t > 0 we expand the solution y(t, x) of (2) in the basis $\{h_j(x)\}_{j=1}^{\infty}$:

$$y(t,x) = \sum_{j=1}^{\infty} s_j(t)h_j(x)$$
(10)
$$s_j(t) = \int_0^l p(x) y(t,x)h_j(x) dx = (y,h_j)_p$$



Using (4) we obtain $s_j(0) = (y_0, h_j)_p$ and $\dot{s}_j(0) = (y_1, h_j)_p$. Multiplying the equation (2) by h_j and integrating it in x we get that the function $s_j(t)$ satisfies the following equation:

$$\ddot{s}_j(t) + \lambda_j s_j(t) = c_j u(t), \quad j = 1, 2, \dots$$

where

$$c_{j} = (f, h_{j}) = \int_{0}^{l} f(x) h_{j}(x) dx$$
(11)

Substituting (10) into (6) and using Parseval's equality we get:

$$\int_{0}^{\infty} \int_{0}^{l} p(x) y^{2}(t,x) dx dt = \int_{0}^{\infty} \sum_{j=1}^{\infty} s_{j}^{2}(t) dt$$
(12)

We reduce the control problem for partial differential equation (2)-(6) to a control problem for a countable system of ordinary differential equations:

$$\int_0^\infty \sum s_j^2(t) dt \to \inf$$
 (13)

$$\ddot{s}_j(t) + \lambda_j s_j(t) = C_j u(t), \quad j = 1, 2, \dots$$
 (14)

$$s_j(0) = a_j, \quad \dot{s}_j(0) = b_j$$
 (15)

$$-1 \le u(t) \le 1 \tag{16}$$

Here $a_j = (y_0, h_j)_p$, $b_j = (y_1, h_j)_p$. We shall assume everywhere below that

$$c_j \neq 0 \quad \text{for all } j = 1, 2, \dots$$
 (17)

Remark. Assumption (17) is very essential for the problem (13)–(16). Indeed, let $c_{j_0} = 0$ for some j_0 . Then j_0 -th equation in (14) takes the form

$$\ddot{s}_{j_0}(t) + \lambda_{j_0} s_{j_0}(t) = 0$$

Hence, if $|a_{j_0}| + |b_{j_0}| \neq 0$ then the corresponding solution $s_{j_0}(t)$ does not vanish as $t \to \infty$. Therefore the integral (13) is equal to $+\infty$ and the optimization problem (13)-(16) has not any sense.

Assume that

$$y_0, y_1 \in L_2((0,l); p), \quad f \in L_2(0,l)$$
 (18)

Put $\omega_j = \sqrt{\lambda_j}$, $\tau_j(t) = \dot{s}_j(t)$. Then the problem (13)–(16) takes the form

$$\int_0^\infty \sum s_j^2(t)dt \to \inf$$
(19)

$$\dot{s}_j = \tau_j, \quad \dot{\tau}_j = -\omega_j^2 s_j + c_j u \tag{20}$$

$$s_j(0) = a_j, \quad \tau_j(0) = b_j, \quad j = 1, 2, \dots$$
 (21)



$$-1 \le u(t) \le 1 \tag{22}$$

Denote

$$s(t) = (s_1(t), s_2(t), \ldots), \quad \tau(t) = (\tau_1(t), \tau_2(t), \ldots)$$
$$a = (a_1, a_2, \ldots), \quad b = (b_1, b_2, \ldots), \quad c = (c_1, c_2, \ldots)$$

Consider the standart Hilbert space l_2 :

$$l_2 = \left\{ w = (w_1, w_2, \ldots) : w_n \in \mathbf{R}, \sum_{n=1}^{\infty} w_n^2 < \infty \right\}$$

with inner product $(v, w) = \sum_{n=1}^{\infty} v_n w_n$. Using assumption (18) we get that $a, b, c \in l_2$.

The existence and uniqueness of a solution $(s(t), \tau(t))$ to problem (19)– (22) in the space $l_2 \times l_2$ were proved in [5] for any initial data from an open neighborhood of the origin $(s = 0, \tau = 0)$. We apply a formal generalization of the Pontryagin maximum principle to the problem (19)–(22). Denote by $\psi_i = (\psi_{i1}, \psi_{i2}, ...)$ (i = 1, 2) adjoint variables. Define the Pontryagin function

$$H(\psi_1, \psi_2, s, \tau, u) = \sum_{j=1}^{\infty} \left(\psi_{1j} \tau_j - \psi_{2j} \omega_j^2 s_j + \psi_{2j} c_j u - s_j^2 / 2 \right) =$$

= $H_0(\psi_1, \psi_2, s, \tau) + u H_1(\psi_1, \psi_2, s, \tau)$

where

$$H_0(\psi_1,\psi_2,s,\tau) = \sum_{j=1}^{\infty} \left(\psi_{1j}\tau_j - \psi_{2j}\omega_j^2 s_j - \frac{1}{2}s_j^2 \right), \quad H_1(\psi_1,\psi_2,s,\tau) = \sum_{j=1}^{\infty} \psi_{2j}c_j$$

For breavity we denote $z = (\psi_1, \psi_2, s, \tau)$. In the space $l_2 \times l_2 \times l_2 \times l_2$ let us consider the Hamiltonian system

$$\psi_{1j} = \psi_{2j}\omega_j^2 + s_j, \quad \dot{s}_j = \tau_j
\dot{\psi}_{2j} = -\psi_{1j}, \quad \dot{\tau}_j = -\omega_j^2 s_j + c_j u^*(t) \quad j = 1, 2, \dots$$
(23)

where $u^*(t)$ satisfies the following maximum condition:

$$H(z(t), u^{*}(t)) = \max_{u \in [-1,1]} H(z(t), u) = \max_{u \in [-1,1]} (uH_{1}(z(t)))$$
(24)

It was proved [5] that the Pontryagin maximum principle is the necessary and sufficient condition of optimality for the problem (19)-(22).

If $H_1(z(t)) \neq 0$ along the trajectory the optimal control is uniquely determined as a function of time from the maximum condition (24):

$$u^{*}(t) = \operatorname{sign}(H_{1}(z(t))) = \operatorname{sign}\left(\sum_{j=1}^{\infty}\psi_{2j}(t)c_{j}\right)$$



Suppose that there exists an interval (t_1, t_2) such that

$$H_1\left(z\left(t\right)\right) \equiv 0, \quad \forall t \in (t_1, t_2)$$

To find an optimal control u(t) in this case we will differentiate the identity $H_1(z(t)) \equiv 0$ by virtue of the Pontryagin maximum principle system (23) until the control u occurs in the resulting expression with a non-zero coefficient.

$$\frac{d}{dt} \Big|_{(23)} H_1(z) = \frac{d}{dt} \Big|_{(23)} \left(\sum_{j=1}^{\infty} \psi_{2j} c_j \right) = \left(-\sum_{j=1}^{\infty} c_j \psi_{1j} \right)$$

$$\frac{d^2}{dt^2} \Big|_{(23)} H_1(z) = \frac{d}{dt} \Big|_{(23)} \left(-\sum_{j=1}^{\infty} c_j \psi_{1j} \right) = -\sum_{j=1}^{\infty} c_j \left(\psi_{2j} \omega_j^2 + s_j \right)$$

$$\frac{d^3}{dt^3} \Big|_{(23)} H_1(z) = -\frac{d}{dt} \Big|_{(23)} \sum_{j=1}^{\infty} c_j \left(\psi_{2j} \omega_j^2 + s_j \right) = -\sum_{j=1}^{\infty} c_j \left(-\omega_j^2 \psi_{1j} + \tau_j \right)$$

$$\frac{d^4}{dt^4} \Big|_{(23)} H_1(z) = \sum_{j=1}^{\infty} c_j \omega_j^2 \left(\psi_{2j} \omega_j^2 + 2s_j \right) - u \sum_{j=1}^{\infty} c_j^2$$
(25)

In our case the control appears in the forth derivative. Assume that all series in (25) are convergent in l^2 . Denote the subsequent derivatives by H_2, H_3, H_4 :

$$H_{2}(z) = -\sum_{j=1}^{\infty} c_{j}\psi_{1j}, \quad H_{3}(z) = -\sum_{j=1}^{\infty} c_{j}\left(\psi_{2j}\omega_{j}^{2} + s_{j}\right)$$
$$H_{4}(z) = -\sum_{j=1}^{\infty} c_{j}\left(-\psi_{1j}\omega_{j}^{2} + \tau_{j}\right)$$

From (25) it follows that

$$H_1(z(t)) = H_2(z(t)) = H_3(z(t)) = H_4(z(t)) = 0, \quad t \in (t_1, t_2)$$

We say a solution of the (23)-(24) is singular if it belongs to the surface

$$\Sigma = \{ z : H_1(z) = H_2(z) = H_3(z) = H_4(z) = 0 \}$$
(26)

A singular control $u^{0}(t)$ is determited from the equation $\frac{d^{4}}{dt^{4}}\Big|_{(23)}H_{1}(z)=0.$

Using (25) we obtain

$$u^{0}(t) = \sum_{j=1}^{\infty} c_{j} \omega_{j}^{2} \left(\psi_{2j} \omega_{j}^{2} + 2s_{j} \right) / \sum_{j=1}^{\infty} c_{j}^{2}$$
(27)

Note that the origin ($\psi_1 = 0$, $\psi_2 = 0$, s = 0, $\tau = 0$) is the singular trajectory and the corresponding singular control $u^0(t)$ equals 0.

It was proved [5] that in a certain neighborhood of the origin the structure of the optimal solutions is the following one: the optimal nonsingular trajectories attain the singular surface Σ in finite time with countable many control switchings, after that the optimal trajectories remain on the singular surface. Namely, the following theorem holds.



Theorem 1. [5]. Let $(c_1\omega_1^3, c_2\omega_2^3, c_3\omega_3^3, \ldots) \in l_2$ and $c_j \neq 0 \quad \forall j$. Assume that there exist positive constants δ and K such that

$$|\omega_{j+1}| - |\omega_j| \ge \delta, \quad |\omega_j| \le K \cdot j, \quad j = 1, 2, \dots$$

Then there exists an open neighborhood of the origin in the space (s, τ) such that the following statements hold for all initial data (a, b) from this neighborhood.

- (i) The problem (19)-(22) has a unique optimal solution.
- (ii) In the space $z = (\psi_1, \psi_2, s, \tau)$ there exists the singular surface Σ of codimension 4 given by the equations

$$\sum_{j=1}^{\infty} \psi_{2j} c_j = 0, \qquad \sum_{j=1}^{\infty} c_j \psi_{1j} = 0$$
$$\sum_{j=1}^{\infty} c_j \left(\psi_{2j} \omega_j^2 + s_j \right) = 0, \qquad \sum_{j=1}^{\infty} c_j \left(-\psi_{1j} \omega_j^2 + \tau_j \right) = 0$$

which is filled in by the singular extremals of the problem (19)-(22). The control on singular extremals are defined by (27).

(iii) For all initial data not belonging to the projection of the singular surface Σ on the space (s, τ) , the optimal trajectories arrive at Σ in finite time with countable many control switchings, i.e., the optimal trajectories are chattering trajectories.



The space $l_2 \times l_2 \times l_2 \times l_2$. The singular surface Σ .

3 Optimal solution for controlling vibrations

Denote $Q_T = (0, l) \times (0, T)$, where T > 0. Consider the Sobolev space $H^k(Q_T) = W_2^k(Q_T)$, $k \ge 0$. The space $H^k(Q_T)$ consists of all functions $v \in L_2(Q_T)$ whose generalized derivatives up to order k exist and belong to $L_2(Q_T)$. The space $H_0^k(Q_T)$ can be defined as a completion of $C_0^\infty(Q_T)$ with respect to the norm of the space $H^k(Q_T)$.

Let $g = uf \in L_2(Q_T)$, $y_1 \in L_2(0, l)$.



Definition 1. [10] We say a function $y \in H^1(Q_T)$ is a *weak solution* of the (2)–(4) if

$$\begin{array}{ll} \text{(i)} & y|_{x=0} = y|_{x=l} = 0, \quad t > 0, \qquad y|_{t=0} = y_0(x), \quad x \in [0, l];\\ \text{(ii)} & \int_{Q_T} \left(k y_x v_x - p y_t v_t \right) dx dt = \int_{Q_T} g v \, dx dt + \int_0^l y_1(x) v(0, x) \, dx\\ \text{for each } v \in H^1\left(Q_T\right): \; v \mid_{x=0} = v \mid_{x=l} = 0, \; v \mid_{t=T} = 0. \end{array}$$

Definition 2. [10] We say $u \in H^2(Q_T)$ is a almost everywhere solut

Definition 2. [10] We say $y \in H^2(Q_T)$ is a *almost everywhere solution* of the problem (2)–(4) provided y satisfies equation (2) in Q_T for almost all $(t, x) \in Q_T$ and y satisfies (3)–(4).

The following theorem is the main result for the problem (1)-(6).

Theorem 2. Let $y_0 \in H_0^2(0, l)$, $y_1 \in H_0^1(0, l)$, k(x), p(x) are smooth enough (for example, $k, p \in C^4([0, l]))$, $p(x) \ge p_0 > 0$, $k(x) \ge k_0 > 0$. Assume that $f \in C^4[0, l]$,

$$f(0) = f(l) = 0, \quad f^{(i)}(0) = f^{(i)}(l) = 0, \ i = 1, 2, 3$$
 (28)

and condition (17) holds. Then there exist positive constants q_1 and q_2 such that if

$$||y_0||_{L_2((0,l);p)} < q_1, \quad ||y_1||_{L_2((0,l);p)} < q_2$$

then

- (i) the problem (2)–(6) has a unique optimal solution $(y^*(t,x), u^*(t));$
- (*ii*) $y^* \in H^2(Q_T)$ for all T > 0;
- (iii) the optimal control $u^*(t)$ has an infinite number of switchings in a finite time interval.

Proof. Here we use notations introduced in Section 2. Since the functions p, k, f satisfy conditions of Theorem 2 it follows [8] that $c_j \sim j^{-4}$ as $j \to \infty$, where

$$c_j = (f, h_j) = \int_0^l f(x) h_j(x) dx$$

Then we have

$$\sum_{j=1}^{\infty} \left(c_j \omega_j^3 \right)^2 = \sum_{j=1}^{\infty} \left(c_j^2 \omega_j^6 \right) = \sum_{j=1}^{\infty} c_j^2 \lambda_j^3 < \infty \implies \left(c_1 \omega_1^3, c_2 \omega_2^3, c_3 \omega_3^3, \ldots \right) \in l_2$$

The property (9) of the eigenvalues $\{\lambda_j\}_{j=1}^{\infty}$ of the problem (8) imply that there exist positive constants δ and B such that

$$|\omega_{j+1}| - |\omega_j| \ge \delta, \quad |\omega_j| \le Bj$$



Now we may apply Theorem 1 to the problem (19)-(22). Then there exists an optimal solution $(s^*(t), u^*(t))$ of the problem (19)-(22) and the optimal control $u^*(t)$ has an infinite number of switchings in the finite time interval.

Consider

$$y^{*}(t,x) = \sum_{j=1}^{\infty} s_{j}^{*}(t) h_{j}(x)$$
(29)

where $\{h_j(x)\}_{j=1}^{\infty}$ are eigenfunctions of the Sturm-Liouville problem (8).

Series (29) formally satisfies equation (2), boundary conditions (3) and initial conditions (4). Since the functions f, y_0 , y_1 satisfy the conditions of Theorem 2 it follows (see [10,11]) that the function $y^*(t,x)$ defined by (29) is a unique weak solution of the problem (2)–(4) and $y^* \in H^2(Q_T)$ for any T > 0. Hence [10] $y^*(t,x)$ is the solution almost everywhere of the problem (2)–(4). Thus the function $y^*(t,x)$ satisfies (2) for almost all $(t,x) \in (0,l) \times (0,+\infty)$, boundary and initial conditions (3)–(4).

Since the function $s^*(t) = (s_1^*(t), s_2^*(t), ...)$ minimizes the functional (13) and the identity (12) holds then the function $y^*(t, x)$ minimizes the functional (6). Thus $(y^*(t, x), u^*(t))$ is a solution of the problem (1)–(6).

4 Conclusion

We considered the optimal control problem of longitudinal vibrations of a nonhomogeneous bar with clamped ends. We proved that the optimal trajectories contain singular part and nonsingular one with accumulation of control swithings.

Acknowledgements

This research was supported by the Russian Foundation of Basic Research (projects No. 11-01-00986 and 13-08-00665).

References

- 1.V. Jovanovic, A Fourier series solution for the longitudinal vibrations of a bar with viscous boundary conditions at each end, Journal of Engineering Mathematics (2012).
- 2.F.E. Udwadia, On the longitudinal vibrations of a bar with viscous boundaries: Super-stability, super-instability, and loss of damping, International Journal of Engineering Science 50 (2012) 79-100.
- 3.I.S. Sadek, J.M. Sloss, S. Adali, J.C. Bruch, Optimal Boundary Control of the Longitudinal Vibrations of a Rod Using a Maximum Principle, Journal of Vibrations and control 3 (1997) 235-254.
- 4.M.I. Zelikin, L.A. Manita, Optimal control for a Timoshenko beam, C.R. Mécanique 334, Issue 5 (2006) 292-297
- 5.V. Borisov, M. Zelikin, L. Manita, Optimal Synthesis in an Infinite-Dimensional Space, Proceedings of the Steklov Institute of Mathematics 271 (2010) 34-52.



- 6.M.A. Al-Gwaiz, Sturm-Liouville Theory and its Applications, Springer Verlag, London, 2008.
- 7.A. Zettl, Sturm-Liouville Problems, in D. Hinton, P.W. Schaefer (Eds), Spectral Theory and Computational Methods of Sturm-Liouville Problems, Lecture Notes in pure and applied mathematics, v.191, Marcel Dekker, Inc., N.Y., 1997.
- 8.I.G. Petrovsky, Lectures on Partial Differential Equations, Dover Publications Inc., N.Y., 1992.
- 9.L. Collatz, Eigenwertaufgaben mit technischen Anwendungen, Akademische Verlagsgesellschaft Geest & Portig, Leipzig, 1963
- 10.V.P. Mikhailov, Partial Differential Equations, Mir Publishers, Moscow, 1978.
- 11.O.A. Ladyzhenskaya, The Boundary Value Problems of Mathematical Physics, Springer — Verlag, N.Y. Inc., 1985





Intrinsic Space Scales for Multidimensional Stochastic Synchronization Models

Anatoly Manita

Department of Probability, Faculty of Mathematics and Mechanics, Lomonosov Moscow State University, 119992, Moscow, Russia (E-mail: manita@mech.math.msu.su)

Abstract. We consider the interacting particle system $x(t) = (x_1(t), \ldots, x_N(t)), t \in \mathbb{R}_+$, consisting of N identical Brownian particles with stochastic synchronizationlike interaction. We provide the answer to the question: what is a typical distance between particles in the synchronized system (when $t = +\infty$) and in the system approaching to "stochastic synchrony" (for large finite t). It appears that in some sense a distance between any pair of particles x_i and x_j is of order N. We obtain some asymptotical results about exponential moments of the distances at time t and show that distributions of $(x_i - x_j)/N$ have exponential tails uniformly in N. For the synchronized N-particle system we prove that all $(x_i - x_j)/N$ have a symmetric bilateral exponential distribution with asymptotically constant parameter: $c_{0,N} \to c_0$ as $N \to \infty$.

Keywords: Stochastic synchronization, Brownian particle system, Markov process, renewal process.

1 Introduction

We consider here a stochastic process $x(t) = (x_1(t), \ldots, x_N(t)) \in \mathbb{R}^N$, $t \in \mathbb{R}_+$, previously studied in [8] and [9]. Components of this process evolve as independent Brownian motions with constant drift coefficient $\sigma > 0$ except random time moments τ_1, τ_2, \ldots of synchronization jumps when states of a randomly chosen pair of components become equal. We give the rigorous definition of the model in Sect. 2.

Mathematical study of stochastic systems with synchronization ([3-7]) is motivated by many applications in computer science (asynchronous parallel and distributed algorithms [1]), wireless network theory (clock synchronization problems [11]) etc.

In the papers [8] and [9] it was assumed that intervals between synchronizations $\tau_n - \tau_{n-1}$ are independent and have common distribution not depending on N. Due to the random nature of the dynamics the so called perfect synchronization $(x_1 = \cdots = x_N)$ is not possible for the present system. But as it was proved in [8] the model reaches synchronization in a stochastic sense: in relative coordinates of a moving observer the system has limit in distribution when $t \to \infty$. Unformunately, the limiting distribution cannot be found explicitly and the study of its properties for large N is a challenging problem. The papers [8] and [9] were devoted mainly to a prestationary evolution of x(t), namely, it was proved that on different time scales ($t = t_N \to \infty$ as $N \to \infty$) the particle system has completely different qualitative behaviour. It was shown that the system passes three different phases before it reaches the final synchronization.

A very important question is: what is a natural space scale for the system of N synchronized particles? In other words: what is a typical distance between particles in the synchronized system? Some information on the distribution of $x_i - x_j$ can be derived from [8] and [9]: assuming that $\mathsf{E}(x_i(0) - x_j(0))^2 < \infty$ for all *i* and *j* we conclude that this second moment remains finite for all t > 0, moreover, it is bounded in *t* and $\mathsf{E}(x_i(\infty) - x_j(\infty))^2$ is of order N^2 . As it will be seen from Theorem 1 of the present paper a similar statement is not true for all exponential moments. We prove (Theorem 1) that $\mathsf{E} \exp(c|x_i(t) - x_j(t)|/N)$ is bounded in *t* only for sufficiently small c > 0. Fortunatelly, this result gives us possibility to get sharp probabilistic estimates for distribution of the rescaled distance $|x_i(t) - x_j(t)|/N$ for large *t*. Moreover, for the synchronized system we show that distribution of $|x_i - x_j|/\sqrt{(N-1)N}$ does not depend on *N* and find this distribution in explicit form (Theorem 2). So we come to the conclusion that the natural space scale for our model is of order *N*.

This work is supported by the Russian Foundation of Basic Research (grant 12-01-00897).

2 Model

We remind here the definition of the model $x(t) = (x_1(t), \ldots, x_N(t)) \in \mathbb{R}^N$ called Brownian particles with synchronization ([8,9]). There is a sequence

$$0 = \tau_0 < \tau_1 < \dots < \tau_n < \dots$$

of random variables. On intervals (τ_{n-1}, τ_n) the particles x_1, \ldots, x_N move according to the free dynamics: $dx_i(t) = \sigma dB_t^i$, $i = 1, \ldots, N$, where $B_t = (B_t^1, \ldots, B_t^N)$, $t \ge 0$, is a N-dimensional Brownian motion independent of $\{\tau_n\}_{n=1}^{\infty}$, the diffusion coefficient $\sigma > 0$ is assumed to be constant. At epoch τ_n a pair of particles (i_n, j_n) is chosen at random¹ and the particle j_n jumps to the particle i_n :

$$x_{j_n}(\tau_n + 0) = x_{i_n}(\tau_n), \tag{1}$$

$$x_k(\tau_n + 0) = x_k(\tau_n), \quad k \in \{1, \dots, N\} \setminus \{j_n\}.$$
(2)

In the framework of particle systems such perturbation of the free dynamics can be regarded as a synchronization-like interaction between particles. It is convenient to consider the jump (1)–(2) as a map $S_{(i,j)} : \mathbb{R}^N \to \mathbb{R}^N$:

$$S_{(i,j)}: \quad (x_1, \dots, x_i, \dots, x_j, \dots, x_N) \mapsto (x_1, \dots, x_i, \dots, x_i, \dots, x_N) \qquad (3)$$

We need some assumptions on the sequence of $\{\tau_n\}_{n=1}^{\infty}$ concerning intervals between synchronization jumps $\Delta_n = \tau_n - \tau_{n-1}, n = 1, \dots$.

¹ I.e., any ordered pair of particles (i_n, j_n) is chosen with probability $\frac{1}{(N-1)N}$.



- A1. $\Delta_1, \ldots, \Delta_n, \ldots$ are independent.
- **A2.** $\Delta_1, \ldots, \Delta_n, \ldots$ are identically distributed with some continuous probability distribution function F.

A3. Δ_n has exponential distribution with mean δ^{-1} : $\mathsf{P} \{\Delta_n > s\} = \exp(-\delta s), s > 0.$

Under Assumption A1–A3 the stochastic process $(x(t), t \ge 0)$ is Markovian. Such model was considered in [8]. In [9] a more general particle system was studied under assumptions A1–A2.

For any random vector $z = (z_1, \ldots, z_N)$ with values in \mathbb{R}^N we denote by \mathcal{P}_z the distribution law of z. Hence \mathcal{P}_z is some probability measure on \mathbb{R}^N .

We need also some assumptions on the initial particle configuration x(0) which is assumed to be random.

AI. The initial particle configuration x(0) is independent of $\{\tau_n\}_{n=1}^{\infty}$ and $(\boldsymbol{B}_t, t \ge 0)$.

Assumption AI is supposed to be held throughout all this paper. Sometimes we will make more restrictive assumptions AIS and AI0.

AIS. Assume that Assumption AI holds and that the initial distribution $\mathcal{P}_{x(0)}$ is invariant with respect to permutations of indices, i.e.,

$$\mathcal{P}_{\pi \star x(0)} = \mathcal{P}_{x(0)}$$

where $\pi: (1, \ldots, N) \to (i_1, \ldots, i_N)$ is an arbitrary permutation and

$$\pi \star (z_1,\ldots,z_N) = (z_{i_1},\ldots,z_{i_N}).$$

AI0. $x_i(0) = 0$ for all i = 1, ..., N.

Evidently, AI0 is a subcase of AIS.

3 Long-time synchronization

It is intuitively clear from definition of the model that collective behavior of a particle system with synchronization is a superposition of two opposite tendencies: with the course of time the free dynamics increases the "spread" of the particle system while the synchronizing interaction tries to decrease it. The *first question* is about a long-time behavior of the *N*-particle system $x(t) = (x_1(t), \ldots, x_N(t))$. We fix N and let $t \to \infty$.

It can be shown [8] that the distribution $\mathcal{P}_{x(t)}$ has no limit in law as $t \to \infty$. Let us consider an improved process $x^{\circ}(t) \in \mathbb{R}^N$

$$x_i^{\circ}(t) = x_i(t) - M(x(t)),$$
 $M(x) := \frac{1}{N} \sum_{m=1}^N x_m,$

One can say that $x^{\circ}(t) = (x_1^{\circ}(t), \dots, x_N^{\circ}(t))$ are coordinates of particles viewed by an moving observer placed in the center of mass M(x). Note that



 $\sum_{i=1}^{N} x_i^{\circ}(t) \equiv 0$ for any $t \geq 0$. It appears that $x^{\circ}(t)$ is an ergodic Markov process (see [8]). In particular,

$$\mathcal{P}_{x^{\circ}(t)} \to \mu^{N} \qquad (t \to \infty)$$

where μ^N is some probability measure on \mathbb{R}^N which is the unique stationary distribution of the Markov process $(x^{\circ}(t), t \geq 0)$. This fact holds under the general assumption AI. Of course, the limiting measure μ^N depends on σ and δ .

Hence the particle system of Section 2 reaches synchronization in a stochastic sense. Properties of the synchronized system can be obtained by studying the probability measure μ^N . This measure is supported on the (N-1)dimensional subspace

$$\left\{x^{\circ} = (x_1^{\circ}, \dots, x_N^{\circ}) : \sum_{i=1}^N x_i^{\circ} = 0\right\} \subset \mathbb{R}^N$$

and seems to be very complicated, especially for large N. So we come to the *next* very natural *question*: what is a typical "size" of the particle configuration (x_1, \ldots, x_N) under the distribution μ^N ?

To formalize a notion of "size" or "spread" of the configuration $x = (x_1, \ldots, x_N)$ we define some functions on the configuration space \mathbb{R}^N . Consider functions

$$d_{ij}(x) = x_i - x_j, \quad d_{ij} : \mathbb{R}^N \to \mathbb{R}.$$

Clearly, $|d_{ij}(x)|$ is a distance between particles i and j in the configuration $x \in \mathbb{R}^N$. Note that $d_{ij}(x) = d_{ij}(x^\circ) = x_i^\circ - x_j^\circ$. Define also functions $D : \mathbb{R}^N \to \mathbb{R}$ and $\mathbf{D}_N : \mathbb{R}_+ \to \mathbb{R}_+$, as follows

$$D(x) := \frac{1}{C_N^2} \sum_{i < j} (x_i - x_j)^2 = \frac{2}{N(N-1)} \sum_{i < j} d_{ij}^2(x)$$
$$\mathbf{D}_N(t) = \mathsf{E} \ D(x(t)), \qquad t \in \mathbb{R}_+ \ .$$

Evidently, $\sqrt{D(x)}$ and $\sqrt{\mathbf{D}_N(t)}$ characterize desynchronization of the total particle system. It was proved in [8] and [9] that for any sequence $\{t_N\}$ such that $t_N/N^2 \to \infty$ as $N \to \infty$ we have

$$\mathbf{D}_N(t_N) \sim \delta^{-1} \sigma^2 N^2 \,. \tag{4}$$

Dynamics of the Markov process x(t) is symmetric with respect to different particles. Therefore the unique stationary distribution μ_N is invariant with respect to permutation of indices. Denoting by E_N the expectation taken with respect to the probability distribution μ^N one can derive from (4) that

$$\mathsf{E}_N \, d_{ij}^2(x^\circ(\infty)) \sim \mathrm{const} \cdot N^2$$
 .

Hence we come to the next suggestion: typical distances between particles in the synchronized system are of order N.

Main goal of the present paper is to give more detailed information on the distribution of "rescaled distances"

$$N^{-1}d_{ij}(x(t)) = \frac{x_i(t) - x_j(t)}{N}$$

for large t (Theorem 1) or even for $t = +\infty$ (Theorem 2).



4 Main Results

4.1 Finite-time analysis

Define the function $G_{N,c}$: $\mathbb{R}^N \to \mathbb{R}$,

$$G_{N,c}(x) := \frac{1}{(N-1)N} \sum_{j_1 < j_2} \exp\left(\frac{c}{N} |x_{j_2} - x_{j_1}|\right), \qquad c > 0.$$

It is useful to keep in mind physical dimensions of all parameters of the model. Assume that x_i is measured in meters (m). Then dimension of σ is $m/sec^{1/2}$, dimension of δ is sec^{-1} , values of the function $G_{N,c}$ are dimensionless, dimension of c is m^{-1} .

Theorem 1 Let assumptions A1–A3 and AI0 hold.

Then there exists $c_0 = c_0(\sigma, \delta) > 0$ such that the following properties hold:

i) for $c < c_0$

$$\limsup_{t \to +\infty} \mathsf{E}\,G_{N,c}(x(t)) \le B(c) < +\infty \tag{5}$$

where B(c) > 0 does not depend on N,

ii) for $c = c_0$

$$N - \frac{1}{2} \le \liminf_{t \to +\infty} \mathsf{E} \, G_{N,c_0}(x(t)) \le \limsup_{t \to +\infty} \mathsf{E} \, G_{N,c_0}(x(t)) < N,$$

iii) for any $c > c_0$ there is $N_0 = N_0(c)$ such that

$$\lim_{t \to +\infty} \mathsf{E} \, G_{N,c}(x(t)) = +\infty$$

for all $N \geq N_0$.

We assume here that all $x_i(0) = 0$ only for brevity of explanations. As it will be seen from the proof (Section 5) Assumption AI0 is not necessary and can be replaced by Assumption AI.

Let us discuss some useful corollaries of Theorem 1. Note that if the model satisfies to Assumption AIS then for any t > 0 the distribution $\mathcal{P}_{x(t)}$ is invariant with respect to the action of the permutation group. Hence all $d_{ij}(x(t)) = x_i(t) - x_j(t)$ are equally distributed and

$$\mathsf{E} G_{N,c}(x(t)) = \frac{1}{2} \mathsf{E} \exp\left(\frac{c}{N} |x_1(t) - x_2(t)|\right).$$

It follows from (5) that distance between particles $|x_i(t) - x_j(t)|$ has exponential moments

$$\mathsf{E} \exp\left(\frac{c}{N} |x_1(t) - x_2(t)|\right) < +\infty \quad \text{for} \quad 0 < c < c_0.$$

By the Chebyshev's inequality this gives an exponential bound for the distance between particles. Namely, under Assumption A10 it follows from Section 5 that for any $0 < c < c_0$

$$P(|x_1(t) - x_2(t)| > rN) \le 2B(c) \exp(-cr), \quad r > 0$$

uniformly in $t \ge 0$. The meaning of this result: a typical distance between particles in the particle system is of order N. Under Assumption AIS a similar results holds for sufficiently large $t \ge t_0(c, N)$.



4.2 Distances between particles of the synchronized system

Let us rescale the functions d_{ij} as follows

$$d_{ij}^{(N)}(x) = (x_i - x_j) / \sqrt{(N-1)N}$$

It seems very surprising that for the synchronised system of Brownian particles distribution of the functions $d_{ij}^{(N)}(x)$ can be given in explicit form.

Theorem 2 Let assumptions A1–A3 and AI hold. Consider the particle system $x = (x_1, \ldots, x_N)$ of Section 2 being in its stationary state μ^N . All random variables $d_{ij}^{(N)}(x)$, $(i, j = 1, \ldots, N)$ have a common distribution not depending on N which appears to be the bilateral exponential distribution with density

$$p(y) = \frac{1}{2}c_0 \exp(-c_0 |y|), \quad y \in \mathbb{R}^1,$$

where $c_0 = \sigma^{-1} \sqrt{2\delta}$.

In particular, in the synchronised system

$$\mathsf{E}_N |x_i - x_j| = \frac{\sigma}{\sqrt{2\delta}} \sqrt{(N-1)N} \sim \frac{\sigma}{\sqrt{2\delta}} N.$$

5 Proof of Theorem 1

Consider the function

$$V_{c_N}(x) = \frac{1}{(N-1)N} \sum_{j_1 < j_2} \cosh\left(c_N |x_{j_2} - x_{j_1}|\right).$$

Note that

$$\frac{e^{|x|}}{2} < \cosh x = \frac{e^x + e^{-x}}{2} \le \frac{e^{|x|} + 1}{2}$$

So it is sufficient to prove main result for the function V_{c_N} with $c_N = c/N$:

$$2V_{c/N}(x) - \frac{1}{2} \le G_{N,c}(x) < 2V_{c/N}(x).$$
(6)

From this point and till the end of the below Lemma 3 we fix some *arbitrary* sequence $\{c_N\}$. To have shorter notation we write V(x) instead of $V_{c_N}(x)$. We introduce also an auxiliary function

$$V_0(x) = \frac{1}{(N-1)N} \sum_{j_1 < j_2} g_N(x_{j_1}, x_{j_2}), \qquad g_N(a, b) = \cosh\left(c_N(a-b)\right) - 1.$$

It is easy to check that $V(x) = V_0(x) + \frac{1}{2}$. Note that

$$g_N(a,b) = g_N(b,a)$$
 and $g_N(a,a) = 0$. (7)



Moreover, $g_N(a, b) > 0$ for $a \neq b$, hence $V_0(x) \ge 0$ and

$$V_0(x) = 0 \quad \Leftrightarrow \quad x_1 = \dots = x_N$$
.

Let f = f(x) be some function on the configuration space \mathbb{R}^N . Denote

$$f^{(n)} = \mathsf{E}\left(f(x(\tau_n)) \mid \{\tau_j\}_{j=1}^{\infty}\right), \qquad n = 1, 2, \dots.$$
(8)

Hence $f^{(n)}$ is a random variable functionally depending on the sequence $\{\tau_j\}_{j=1}^{\infty}$. Consider now $\{V^{(n)}\}$ and $\{V_0^{(n)}\}$. Evidently, $V^{(n)} = V_0^{(n)} + \frac{1}{2}$.

Define a map-valued random variable \mathcal{S} such that

$$\mathsf{P}\left\{\mathcal{S} = S_{(i,j)}\right\} = \frac{1}{(N-1)N}, \quad i \neq j.$$

To continue the proof we need next two lemmas.

Lemma 1 There exists $\varkappa > 0$ such that for any $x \in \mathbb{R}^N$

$$\mathsf{E} V_0(\mathcal{S}x) = k_N V_0(x),$$

where $k_N = 1 - \varkappa / ((N - 1)N)$.

Lemma 2 For $s > 0, x \in \mathbb{R}^N$

$$\mathsf{E} V(x + \sigma \mathbf{B}_s) = V(x)e^{2qs} \tag{9}$$

where $q = \left(c_N \sigma\right)^2 / 2$.

Proof of Lemma 1 can be obtained using properties (7) and ideas of Lemma 2 in [7]. The constant \varkappa is dimensionless. Moreover, it follows from [7] that $\varkappa = 2$ for the case of pair interaction (1)–(2). We omit details. The proof of Lemma 2 is a straigtforward calculation based on the identity

$$\mathsf{E} \exp\left(bB_s^i\right) = \exp\left(b^2s/2\right).$$

It follows from Lemma 1 that

$$\mathsf{E}V(\mathcal{S}x) = k_N V(x) + \frac{1 - k_N}{2}.$$
 (10)

Note that the stochastic process x(t) has left continuous paths. Since on the intervals $[\tau_{n-1}, \tau_n)$ the process has only one synchronization jump at time τ_n , combining (9)–(10) with Assumption A1 we get

$$V^{(n)} = k_N V^{(n-1)} e^{2q\Delta_n} + l_N$$

where $l_N = \frac{1}{2}(1-k_N)$. Hence $V^{(n)} = k_N^2 V^{(n-2)} e^{2q(\Delta_{n-1}+\Delta_n)} + k_N e^{2q\Delta_n} l_N + l_N$. Finally, we come to the following representation

$$V^{(n)} = k_N^n V^{(0)} e^{2q(\Delta_1 + \dots + \Delta_n)} + k_N^{n-1} e^{2q(\Delta_2 + \dots + \Delta_n)} l_N + \dots + k_N e^{2q\Delta_n} l_N + l_N.$$



Fix some t > 0. The number of synchronization jumps on the time interval [0,t] is random. To proceed with the proof we recall some facts and notation from renewal theory [2]. Let $\Pi_t = \max\{m : \tau_m \leq t\}$ and hence $\tau_{\Pi_t} = \max\{\tau_i : \tau_i \leq t\}$. From Assumption A2 it follows that (with probability 1) on the open time interval (τ_{Π_t}, t) there is no synchronization jump. Hence

$$\mathsf{E} \left(V(x(t) \mid \{\tau_j\}_{j=1}^{\infty} \right) = V^{(\Pi_t)} e^{2q(t-\tau_{\Pi_t})} = \\ = k_N^{\Pi_t} V^{(0)} e^{2qt} + k_N^{\Pi_t - 1} e^{2q(t-\tau_1)} l_N + \cdots \\ + k_N e^{2q(t-\tau_{\Pi_t - 1})} l_N + e^{2q(t-\tau_{\Pi_t})} l_N .$$

Now we are going to consider a conditional expectation $\mathsf{E}\left(\cdot \,|\, \Pi_t\right)$ of this expression:

$$\mathsf{E} \left(V(x(t) \mid \Pi_t = n) = k_N^n V^{(0)} e^{2qt} + l_N \sum_{i=1}^n k_N^{n-i} \mathsf{E} \left(e^{2q(t-\tau_i)} \mid \Pi_t = n \right).$$
(11)

Under Assumption A3 ($\Pi_t, t \ge 0$) is a Poisson process of intensity δ . It is easy to check (see, for example, the technical lemma 4.2 on page 62 of [10]) that conditional distributions of $t - \tau_m$ belong to the class of Beta-distributions:

$$\mathsf{P}\left\{t - \tau_m \in (y, y + dy) \,|\, \Pi_t = n\right\} = n \, C_{n-1}^{m-1} \, \frac{y^{n-m} (t-y)^{m-1}}{t^n} \, dy,$$
$$y \in (0, t), \ m = \overline{1, n}.$$

Considering the second summand in (11) and applying the Binon formula

$$\sum_{i=1}^{n} k_N^{n-i} \int_0^t e^{2qy} n C_{n-1}^{i-1} \frac{y^{n-i}(t-y)^{i-1}}{t^n} \, dy = t^{-1} \int_0^t e^{2qy} n \left(\frac{k_N y + (t-y)}{t}\right)^{n-1} \, dx$$

we get

$$\mathsf{E} \left(V(x(t) \mid \Pi_t = n) = k_N^n V^{(0)} e^{2qt} + l_N t^{-1} \int_0^t e^{2qy} n \left(\frac{k_N x + (t-x)}{t} \right)^{n-1} dy.$$

Averaging in n and using notation $\varphi(y) = \mathsf{E} y^{\Pi_t}$ for the generating function of Π_t we come to the following representation.

$$\mathsf{E} V(x(t) = \varphi(k_N) V^{(0)} e^{2qt} + l_N t^{-1} \int_0^t e^{2qy} \varphi'\left(1 - \frac{(1-k_N)y}{t}\right) dy$$

Recall that under Assumption A3 the random variable Π_t is Poissonian with the mean δt . Since $\varphi(y) = \exp(\delta t (y - 1))$ and $\varphi'(y) = \delta t \exp(\delta t (y - 1))$, we



have

$$\begin{split} t^{-1} \int_0^t e^{2qy} \,\varphi' \left(1 - \frac{(1-k_N)\,y}{t} \right) \,dy \ &= t^{-1} \int_0^t e^{2qy} \,\delta t \,\exp\left(-\delta t \,\cdot\, \frac{(1-k_N)\,y}{t} \right) \,dy \ &= \\ &= \delta \int_0^t \,\exp\left(2qy - \delta \left(1 - k_N \right) y \right) \,dy \ &= \\ &= \begin{cases} \frac{\delta}{a_N} \,\left(1 - e^{-a_N t} \right) \,, & \text{if } a_N \neq 0, \\ \delta t, & \text{if } a_N = 0, \end{cases} \end{split}$$

where $a_N = \frac{\varkappa \delta}{N(N-1)} - 2q$. Moreover,

$$\varphi(k_N) \exp(2qt) = \exp(2qt - \delta t (1 - k_N)) = e^{-a_N t}.$$

So we have got explicit formulae for $\mathsf{E} V(x(t))$. Let us summarize this result in the next lemma.

Lemma 3 In the case $a_N \neq 0$

$$\mathsf{E} V(x(t)) = e^{-a_N t} \mathsf{E} V(x(0)) + \frac{l_N \delta}{a_N} \left(1 - e^{-a_N t} \right)$$
(12)

and in the case $a_N = 0$

$$EV(x(t)) = EV(x(0)) + l_N \delta t =$$

$$= EV(x(0)) + (c_N \sigma)^2 t$$
(13)

where

$$l_N = \frac{\varkappa}{2N(N-1)}, \qquad a_N = \frac{\varkappa\delta}{N(N-1)} - (c_N \sigma)^2.$$

In other words, the function $f(t) = \mathsf{E} V(x(t) \text{ solves the following equation})$

$$f' = -a_N f + l_N \delta \,.$$

With this lemma in mind we can finish the proof of Theorem 1. Indeed, we see from (12) and (13) that $\mathsf{E} V(x(t))$ is bounded in t if and only if $a_N > 0$. Putting $c_N = c/N$ we conclude

$$a_N > 0 \quad \Leftrightarrow \quad c^2 < \frac{\varkappa \delta}{N(N-1)} \cdot \frac{N^2}{\sigma^2} \quad \Leftrightarrow \quad c < \frac{\sqrt{\varkappa \delta}}{\sigma} \cdot \sqrt{\frac{N}{N-1}}.$$
 (14)

From this observation the statement of Theorem 1 follows with $c_0 = \sigma^{-1}\sqrt{\varkappa\delta}$. Indeed, the assumption on initial configuration $(x_i(0) = 0 \text{ for all } i = 1, \ldots, N)$ implies that $V(x(0)) = \frac{1}{2}$. If $c \leq c_0$ then $a_N > 0$. Hence $\mathsf{E} V(x(t)) \uparrow w_N(c)$ as $t \to +\infty$ where

$$w_N(c) = \frac{l_N \delta}{a_N} = \frac{1/2}{1 - \frac{c^2 \sigma^2}{\varkappa \delta} \cdot \frac{N-1}{N}}.$$



Taking into account (6) in the case $c < c_0$ we put

$$B(c) = \frac{1}{1 - (c/c_0)^2}$$

and get the item i) of the theorem. In the case $c = c_0$ we have $w_N(c_0) = N/2$. Now the item ii) of Theorem 1 immediately follows from (6).

Assume that $c > c_0$. It is easy to see from (14) that there exists N_0 such that $a_N < 0$ for all $N \ge N_0$. For any such N the r.h.s. of (12) tends to infinity as $t \to +\infty$.

The proof of Theorem 1 is over.

Evidently, this proof is valid also for the general initial assumption AI. The only difference is that for $c \leq c_0$ we still have convergence $\mathsf{E} V(x(t)) \to w_N(c)$ as $t \to +\infty$ but we cannot ensure that this convergence is monotonic.

6 Laplace Transform: Proof of Theorem 2

We proved in Section 5 that if $c \leq c_0$ then

$$\mathsf{E} V_{c/N}(x(t)) \to w_N(c) \qquad (t \to \infty),$$

where $c_0 = \sigma^{-1} \sqrt{\varkappa \delta}$ and

$$w_N(c) = rac{1/2}{1 - rac{c^2}{c_0^2} rac{N-1}{N}}$$

Let ξ be some random variable. Denote by $\mathcal{L}(\xi; s)$ is the Laplace transform of distribution of ξ :

$$\mathcal{L}(\xi; s) := \mathsf{E} \exp\left(-s\xi\right).$$

If the distribution of ξ is symmetric, i.e. $\mathcal{P}_{\xi} = \mathcal{P}_{-\xi}$, then $\mathcal{L}(\xi; s) = \mathsf{E} \cosh(s\xi)$.

It follows from Section 3 the limiting stationary distribution μ^N is invariant with respect to permutations of indices. Hence under the distribution μ^N all random variables

$$d_{ij}(x) = x_i - x_j = x_i^\circ - x_j^\circ$$

are equally distributed and symmetric. Hence

$$\mathsf{E}_N V_{c/N}(x^\circ) = \frac{1}{2} \, \mathsf{E}_N \cosh(c d_{12}(x^\circ)/N) = \frac{1}{2} \, \mathcal{L}(d_{12}(x^\circ)/N; c).$$

From all these arguments and the following observation

$$\mathsf{E}_N V_{c/N}(x^\circ) = w_N(c) = \frac{1/2}{1 - \frac{c^2}{c_0^2} \frac{N-1}{N}}$$

we conclude that the rescaled differences $(x_i - x_j)/N$ have the Laplace transform

$$\mathcal{L}(c) = \frac{1}{1 - \frac{c^2}{c_0^2} \frac{N-1}{N}}$$
(15)



and hence they have the common bilateral exponential distribution with probability density $p(y) = \frac{1}{2}c_{0,N} \exp\left(-c_{0,N} |y|\right), y \in \mathbb{R}$, where

$$c_{0,N} = \frac{c_0}{1 - \frac{1}{N}} \,. \tag{16}$$

The presence of N in the formulae (15) and (16) can be eliminated if we improve the scale factor as follows:

$$d_{ij}^{(N)}(x) = (x_i - x_j)/\sqrt{(N-1)N}$$

For this case in the synchronized system any $d_{ij}^{(N)}$ has the bilateral exponential distribution with parameter $c_0 = \sigma^{-1} \sqrt{\varkappa \delta}$:

$$\mathcal{L}(c) = \frac{1}{1 - \frac{c^2}{c_0^2}}, \qquad p(y) = \frac{1}{2}c_0 \exp\left(-c_0 |y|\right).$$

In particular, $\mathsf{E}_N \left| d_{ij}^{(N)} \right| = c_0^{-1} = \frac{\sigma}{\sqrt{\varkappa\delta}}, \quad \mathsf{Var}_N \left(d_{ij}^{(N)} \right) = 2c_0^{-2} = \frac{2\sigma^2}{\varkappa\delta}$ and $\mathsf{E}_N \left| x_i - x_j \right| = \frac{\sigma}{\sqrt{\varkappa\delta}} \sqrt{(N-1)N} \sim \frac{\sigma}{\sqrt{\varkappa\delta}} N.$

We recall remark made after Lemma 1 that $\varkappa = 2$ in the case of pair interaction (see [7]).

It should be noted that the nonstationary "finite-time" distributions $\mathcal{P}_{d_{12}(x(t))}$, t > 0, do not have such simple and explicit form as $\mathcal{P}_{d_{12}(x)}$ in the synchronized system.

7 Concluding remarks

The main result of this paper remains true in more wide situations than assumptions of Sect. 2 and Sect. 4. Instead of pair interaction (1)–(2) we can consider the class of interactions called k-particle symmetrical synchronizations [7,8]. The only difference is in a value of the constant \varkappa . Evidently, the statement of Theorem 1 holds for any initial distributions of x(0) because Lemma 3 is valid in the general situation.

Similar result is expected also for the model studied in [7] where a free dynamics is governed by independent random walks.

We hope that Assumption 3 is not essential and an analog of Theorem 1 can be proved also for non-Markovian case under Assumptions A1–A2. The method of proof would be essentially the same but now one should pay additional attention to a general renewal process related with the sequence $\{\tau_n\}$. To show some difficulties which are present in the non-Markovian case we refer to [9].

Anouther important axis of the future research is the joint distribution of $(d_{ij} = x_i - x_j, i \neq j)$ which seems to be very complicated due to strong correlations in the synchronized particle system. In the present paper we discussed only one dimensional marginal distributions.



References

- 1.D.P. Bertsekas and J.N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*, Belmont, Athena Scientific, 1997.
- 2.D.R. Cox. Renewal theory, London, Methuen, 1962.
- 3.V. Malyshev, A. Manita. Phase transitions in the time synchronization model. *Theory of Probability and its Applications* 50:134–141, 2006.
- 4.A.G. Malyshkin, Limit dynamics for stochastic models of data exchange in parallel computation networks. Problems of Information Transmission 42:234–250. (2006)
- 5.A. Manita, V. Shcherbakov, Asymptotic analysis of a particle system with mean-field interaction, *Markov Processes Relat. Fields* 11:489–518 (2005)
- 6.A. Manita, Markov processes in the continuous model of stochastic synchronization. Russ. Math. Surv. 61:993–995. (2006)
- 7.A.D. Manita. Stochastic synchronization in a large system of identical particles. Theory of Probability and its Applications 53, 155–161, 2009. (See also http://arxiv.org/abs/math.PR/0606040)
- 8.A. Manita, Brownian particles interacting via synchronizations. Communications in Statistics — Theory and Methods, 2011, v. 40, N. 19-20, p. 3440–3451. (see also http://arxiv.org/abs/1012.3140)
- 9.A. Manita, On Markovian and non-Markovian models of stochastic synchronization. Proceedings of The 14th Conference "Applied Stochastic Models and Data Analysis" (ASMDA), 2011, Rome, Italy, p. 886–893. (See also http://goo.gl/8P4ma)
- 10.A.D.Manita, On phases in evolution of multi-dimensional interacting diffusions with synchronization. In *Contemporary problems of mathematics and mechanics*. Vol. 7. Mathematics. 2011. P. 50-67 (In Russian).
 - $http://mech.math.msu.su/probab/cheb190.pdf\#page{=}50$
- 11.O. Simeone, U. Spagnolini, Y. Bar-Ness, S. Strogatz. Distributed synchronization in wireless networks. *IEEE Signal Processing Magazine*, 2008, V. 25, N. 5, pp. 81– 97.



Optimal Scaling with missing categorical data: An application to the study of labour markets in 19th century Spain

Cecilio Mar Molinero¹, and Fabiola Portillo²

University of Kent, Canterbury, CT2 7PE, UK, and Universitat Autonoma de Barcelona, Barcelona, Spain (E-mail: C.Mar-Molinero@kent.ac.uk)

Department of Economics and Business, University of La Rioja, C/ Ciguena 60, 26005 Logrono, La Rioja, Spain (E-mail: fabiola.portillo@unirioja.es)

Abstract. The information contained in a database from the Spanish railway company MZA for the period 1882-1889 is explored by means of Optimal Scaling methods. The sample includes 992 employees who joined the Madrid-Atocha workshop during that period and contains missing qualitative values. A technique that combines correspondence analysis with the k-means clustering algorithm is implemented to impute these values, maximizing internal consistency as measured by Guttman's squared correlation ratio. The results show two characteristics observed in other studies of labour relations: the existence of «ports of entry» for workers at low levels of qualification and long-term labour relations.

Keywords: Optimal Scaling Analysis; Missing Categorical Data; Labour markets.

1 Introduction

In the last decades, the importance of the contributions made by R. A. Fisher within the field of Statistics has been reassessed (Hald[19]; Welsh and Robinson[53]). The relevance of Fisher's methodological contributions, and his innovative approaches to probability and inference, have been recognised in this literature. However, his Optimal Scaling (OS) methods –or «the appropriate scoring technique» as Fisher[12] described them– have received lesser attention (Gower[15]). OS methods remain relatively under-utilised when compared with other contributions he made. This is particularly true of economic studies and has been attributed to the way in which Fisher introduced OS. His presentation of the subject is hard to follow within a modern multivariate data analysis context. In this sense, Savage[41] (p. 443) points out that «the world of R. A. Fisher is at once very near to and very far from the world of modern statisticians».

The idea of allocating quantitative values to qualitative concepts by means of OS methods can be traced to the seventh edition of Fisher's[12] Statistical Methods for Research Workers. After this contribution, Hayashi[20], Kruskal[26], De Leeuw



et al.[6], Takane et al.[47], De Leeuw[5], Meulman et al.[34], and Neal and Roberts[35], have developed the methodological tools that make it possible its systematic application in many areas of research.

In this paper we apply OS methods to analyse the introduction of internal or non-competitive labour markets (ILMs) in the Spanish railway company MZA at the end of the 19th century.

It has long been argued that changes in the nature of work, often as a consequence of technological innovation, impact on the organization of labour (Barley and Kunda[2]). Much attention has been paid to the relationships between employment contracts and organizational structures (Turner[49]). Of particular relevance are the changes brought about by industrialization in the late 19th century. At that time, the introduction of new production methods was sometimes accompanied by paternalistic welfare systems, resulting from the personal ethics of factory owners (Fitzgerald[14]). But welfare systems also emanated from economic and social forces since the new industrial concerns needed to train their workers in firm-specific skills. This gave rise to a retention problem (Mackinnon[30]). A solution, particularly suited to large mechanistic organizations operating in stable environments, was the introduction of ILMs (Fitzgerald[13]).

ILMs consist of a set of explicit or implicit rules and procedures governing labour relations, particularly in respect to the recruitment, training, job ladders, pay policies and job security (Lazear and Oyer[28]). In an ILM, new workers join the company at the so-called «ports of entry» (Kerr[25]), which tend to be associated with low qualification levels, and are paid the market rate for the job (Doeringer and Piore[7]). Career progression from low skilled to high skilled jobs would be limited to employees who have been in the firm for some time, in order to use senioritybased promotion ladders, something that the transaction costs model would predict (Sundstrom[46]; Siebert and Addison[44]). Pay policies are consistent with the efficiency wage model in that internal workers receive relatively high rewards with respect to the competitive or external labour market (Idson and Feaster[24]); such rewards may take the form of welfare programmes.

The development of ILMs is often associated with large modern organizations that operate in well-defined and relatively stable markets, and are able to hire workers with long term contracts (Fitzgerald[13]). This situation applies particularly to the iron and steel industry, and to large railway companies (Howlett[21]). During their growth period, railway companies had to recruit, control, train, and retain a large number of workers, becoming more bureaucratic and hierarchical (Drummond[8]). There is general agreement in the ILMs literature that these personnel policies evolved in the 20th century, but their presence in the 19th century is still an issue (Doeringer and Piore[7]). Evidence of ILMs in railway companies, in the UK and Australia, at the late 19th century is found by Drummond[9], Mackinnon[30][31], Savage[42], and Howlett[22][23], but no evidence of ILMs appears to exist for other developed countries during that period. The question is



whether the Spanish railway company MZA was operating ILM regulations at the end of the 19th century.

The MZA company was established in Madrid in 1856, and continued to operate until it was nationalised in 1941, when it became part of RENFE, the public Spanish railway monopoly. Until that moment MZA was the largest or the second largest firm in Spain, and the 10th in Europe (Tedde[48]). The study is based on the personal records of MZA employees.

The sample includes 992 employees who joined the MZA Madrid workshop during the period 1882-1889 and contains missing qualitative values. These had to be estimated using a technique that combines correspondence analysis with the k-means clustering algorithm. The approach maximises data internal consistency as measured by Guttman's squared correlation ratio (Van Buuren and Van Rijckevorsel[50]).

The predictions of the standard theories on labour markets were supported by the data. The results reveal two characteristics observed in other studies of labour relations: the existence of «ports of entry» for workers at low levels of qualification, and long-term labour relations. This paper, therefore, attempts to contribute to the understanding of the introduction of advanced management methods at the beginning of the Second Industrial Revolution, by means of a multivariate statistical approach, which is innovative in this area of research.

The rest of the paper is organised as follows. The description of the data and a brief discussion of the fundamentals of the statistical model employed are followed by the results obtained, which contains several parts. The first analysis subsection describes the way in which categorical missing data imputation took place; the second and the third analysis subsections deal with the main body of the research using the techniques of Categorical Principal Components (CPC) and Property Fitting, respectively. The paper ends with conclusions.

2 Data and methods

The data used in this study was taken from the personal information records of 992 employees of the Spanish railway company MZA hired between 1882 and 1889. Some of the information contained in the files, such as initial salary, is quantitative, and some information, such as reasons for leaving the company, is qualitative. To be able to properly analyse the multivariate information contained in the data set, it is necessary to jointly deal with qualitative and quantitative data and, therefore, with variables measured in several scale types, something that determines the statistical methods to be used (Stevens[45]). Optimal Scaling (OS) methods were employed to quantify qualitative data. This made it possible the subsequent implementation of standard multivariate analysis techniques such as CPC and Property Fitting.

There was a problem of missing categorical data that had to be addressed: out of



the 992 files, only 537 contained complete information. Missing data were estimated using an OS based method appropriate for the imputation of categorical data, MISTRESS, suggested by Van Buuren and Van Rijckevorsel[50].

2.1 Variables

The original information was recoded into eight variables: enrolment age, experience in the firm, initial wage, initial job level, leaving reason, marital status, place of birth, and section in which new workers carried out their activity. The first three variables are quantitative, while the last five are qualitative. Leaving reason was often left blank, and there was complete information for only 537 employees.

The description of initial job performed was carefully recorded in the original personal files. Taking into account job descriptions, as defined by the firm, these jobs were coded into an ordinal variable with six categories or levels, ranging from deputy director of the workshop to apprentice. It is relevant to point out that although entries into the company took place at all levels of hierarchy, most workers were hired with low level qualifications. This is consistent with the existence of «ports of entry» for new workers at low levels of qualification.

It is clearly different to be a skilled carpenter or a skilled mechanical engineer, but it was felt that this difference would be captured by the section in which the work took place. Most skilled workers were hired into jobs traditionally performed by crafts. The management and control of the crafts was a problem that had to be faced by railway companies in the 19th century (Drummond[8]). Unlike other employees, craftsmen had well established working practices, based in their tradition, and craft autonomy had to be reduced by managerial control. To address this situation, companies increased the division of labour in the workshops through a bureaucratization of human resources, this being a common characteristic of ILMs.

The personal files did not register changes that took place during the period of employment, other than the date of its termination and the reasons for this. The period studied suffered from high levels of labour unrest. This resulted in legislative attempts to improve working conditions, which had deteriorated as a consequence of the industrialization process. In this context, the reasons for leaving the firm reflect social tensions arising from the organisation of work. These reasons were classified into: death, retirement, illness, resignation, transfer away from the Madrid-Atocha workshop, redundancy, and disciplinary dismissal. It is to be noticed that the 60.5% of the workers left the company either through resignation or disciplinary dismissal.

Place of birth was recoded into four categories: Madrid, elsewhere in Spain, France, and others. The largest group of workers was from Madrid, where the workshop was situated. Separating Madrid from the rest of Spain made it possible to check differences in education between workers with an urban background, and those with an agricultural or rural industry background. MZA was set up with French capital, and French workers were the largest non-Spanish contingent.


2.2 Optimal Scaling with categorical missing data

The data used in this study contains a mixture of variables measured in nominal, ordinal, and ratio scales. OS is appropriate in this situation since it quantifies qualitative data and, by doing so, makes it possible to apply the standard methods of multivariate analysis. Fisher's[12] OS methodology has now been extended, and can be imbedded within a large set of multivariate analysis approaches, such as regression, canonical correlation analysis, discriminant analysis, principal components analysis, and multiple correspondence analysis.

OS methods treat observations as if they were categorical –either because they are thus, or because they are seen as the result of a finite measurement process, and allocates parameters to each category through an optimisation approach. Fisher's[12] approach is based on the estimation of a linear model of the ANOVA type. Following this approach, let *Y* be the $(n \times q)$ normalised indicator matrix of the *n* observations, and *q* categories that need to be scaled. Let β be a *q* length vector of the OS parameters to be estimated. The model is:

$$Y\beta = Xz + \varepsilon, \qquad [1]$$

where *X* is the $(n \times p)$ matrix that contains the values of the remaining variables in the model, *z* a vector of parameters. and ε the random disturbance term.

The parameter vector, β , can be estimated using the Alternating Least Squares (ALS) approach. This requires minimising a loss function, which could be some version of Least Squares; the most popular function being Kruskal's[26] stress. There exist some algorithms that have few local minima problems (Young[54]).

An alternative way of obtaining the values of the OS parameters relies on the Singular Value Decomposition (SVD) of the data matrix:

$$X^{T}Y = \sum_{k=1}^{r} l_{k}a_{k}b_{k}^{T},$$
[2]

where $r = \min(p, q)$, $a_k \in \mathbb{R}^p$ y $b_k \in \mathbb{R}^q$ and $l_1 \ge l_2 \ge ... \ge l_r \ge 0$. The solution would be contained in b_1 , the vector with the highest singular value in the decomposition (Welsh and Robinson[53]). In this study, we follow this approach because it has the advantage that it makes possible to perform hypotheses tests on the OS parameters, as done by Fisher[12] and Barlett[1].

In this study, much effort went into the treatment of missing values, so that these could be estimated without affecting the statistical properties of the data set. The existence of missing data is a common problem when working with data bases. Which method is appropriate to deal with missing values depends on the missingness pattern and mechanism that generates the absences (Little and Rubin[29]). The missing data pattern describes which values are missing in the data, and the mechanism describes the relationship between absent data and variable values. In this particular case, the only variable with missing values is "leaving reason", hence the missing data pattern is univariate. The absence of leaving reason is significantly related to the values of other variables –experience in the firm, initial job level, place of birth, and section, therefore the mechanism is not Missing Completely at Random (MCAR). For example, leaving reason is missing more frequently amongst low experienced workers employed in low level jobs. One can only imagine that, if a worker does not turn up to work on a particular day, there will be an inclination to wait to see if the absence is temporary or permanent, and that permanent absences may just go unrecorded. On the other hand, death during service is a traumatic event and will probably be immediately recorded. Thus, the absence of the value of a variable may depend on the value of another variable, against the assumption of the MCAR process.

When the data is MCAR, ignoring observations with incomplete data results in correct inferences, although there is some information loss. When the presence or absence of an observation in a particular variable depends on the value of another variable, one should take into account, during model estimation, the process by which observations are missing to avoid bias (Rubin[40]).

If all the variables were measured on a ratio scale, we would be taking advantage of correlations between variables in order to estimate the most likely values for the missing data. In this particular data set, because some of the variables, including leaving reason, are qualitative, imputation of missing values is more complex (Von Hippel[51]; Ferrari and Annoni[11]). Missing data were estimated using an OS based method appropriate for the imputation of categorical data, MISTRESS, suggested by Van Buuren and Van Rijckevorsel[50]. This technique combines correspondence analysis with the k-means clustering algorithm is implemented to impute these values, maximizing internal consistency as measured by Guttman's squared correlation ratio.

A several step procedure was followed for the estimation of missing values: we first applied the OS model to the data that excluded employees with missing observations (Portillo et al.[38]); second, correlations between variables were studied; third, the relationships between variables were explored using Path Analysis (Retherford and Choe[39]); and, finally, a homogeneity analysis based procedure due to Van Buuren and Van Rijckevorsel[50] was followed in order to obtain the imputations that were used in the final OS analysis of the data.

3 Results

3.1 Categorical Missing Data

The application of OS to the data set that contains no missing values –i.e., 537 observations– generated a set of correlations between variables. A study of the correlation structure of the data was next performed using Path Analysis. The



AMOS routine of SPSS was run using as endogenous variables experience in the firm, initial salary, and leaving reason. The remaining variables –enrolment age, initial job level, marital status, place of birth, and section of work– were treated as exogenous. The only two relevant path coefficients associated with leaving reason that took values significantly different from zero were the ones that linked leaving reason with enrolment age and with experience in the firm.

The final step was to use the MISTRESS imputation procedure of Van Buuren and Van Rijckevorsel[50]. This procedure maximizes consistency in the data set as measured by Guttman's squared correlation ratio, η^2 (Guttman[18]). The MISTRESS technique combines correspondence analysis methods with the k-means clustering algorithm. The variables found to be associated with leaving reason in the previous analyses –enrolment age, and experience in the firm– were used to obtain estimates of the missing values, whose distribution is presented in table 1.

Table	1.	Mis	sing	data	im	putation	for	empl	ovees	leaving	reason
			~								

Leaving	Obs	erved	MC exp	MCAR [‡] expected			TRESS puted	χ^2 -statistic [†]
reason	Count	%	Count	%		Count	%	(p-value)
Missing	352	39.6	0	0.0		0	0.0	68.720
Resignation	228	25.6	378	42.5		472	53.1	(≤0.01)
Dismissal	80	9.0	132	14.9		98	11.0	
Redundancy	27	3.0	45	5.0		59	6.6	
Transfer	38	4.3	63	7.1		76	8.5	
Illness	38	4.3	63	7.1		39	4.4	
Death	86	9.7	142	16.0		95	10.7	
Retirement	40	4.5	66	7.4		50	5.6	

‡ Missing Completely at Random.

 $\dagger~\chi 2$ test for the difference between the expected distribution under MCAR mechanism and MISTRESS estimates.

The percentage of missing values was 39.6% with respect to the total in leaving reason, and 4.9% with respect to the total data base. These percentages are within the acceptability limit of unknown entries required to avoid consistency bias, obtained by Van Buuren and Van Rijckevorsel[50] using the bootstrap. Guttman's η^2 -statistic took the value 0.68, which is higher than the 0.50 limit required to maximize consistency, also obtained by these two authors using the bootstrap.

3.2 Optimal Scaling and Categorical Principal Components (CPC)

After imputing missing values, the data for the eight variables on 889 workers were analysed using CPC. The first two components accounted for more than 50% of the variation in the data, and the first five components accounted for almost 90%.

			Optim	al Scaling
Variable		Frequency		F-statistic [†]
		n = 889	Scores	(p-value)
Initial job level	Third level	4	2.548	26.358
	Fourth level	59	2.238	(≤0.01)
	Fifth level	353	0.766	· · · ·
	Sixth level	473	-0.873	
Leaving reason	Resignation	472	-0.601	25.698
	Disciplinary dismissal	98	-0.438	(≤ 0.01)
	Redundancy	59	-0.224	(_ ••••-)
	Transfer	76	-0.158	
	Illness	39	2.021	
	Death	95	1.599	
	Retirement	50	2.421	
Marital status	Single	484	-0.903	3.682
	Married	391	1.035	(< 0.01)
	Widower	14	2.323	(_ ••••-)
Place of birth	Madrid	283	0.692	12.335
	Rest of Spain	559	0.015	(≤ 0.01)
	France	30	4.257	(= 0101)
	Other countries	17	3.511	
Section	Forge	70	-1.358	15.249
	Assembly	96	-1.181	(≤0.01)
	Boiler making	107	-0.898	
	Foundry-elect. installations	74	-0.738	
	Fitting	64	-0.172	
	Carriages	219	0.352	
	Upholstery- Paintwork	252	0.959	
	Offices	7	1.715	

Table 2. Optimal Scaling scores for qualitative variables

† Fisher's[12] test for the global significance of OS parameters.

The scores assigned to the categories of each non-numerical variable are given in table 2. Score values are important for interpretation purposes.

Scores in initial job level increase as seniority increases: the lowest score in initial job is associated with apprentices and the highest score is attached to supervisors. Moving on to the scores of the various categories of leaving reason, it is to be noticed that abnormal reasons –resignation, dismissal, and redundancy–show negative scores, while categories that are a normal end to an employment career –illness, death, and retirement– show positive scores. The ordering of the scores for the different categories of place of birth is to be noted. The higher scores are attached to the categories France and other countries. This makes sense in the context of the MZA firm, since it is suspected that the not-Spanish workers were specialists who joined the firm for a short period of time.





Fig. 1. Workers plotted in the first two categorical principal components by leaving reason with indication of experience in the firm.

Table 2 also shows the result of Fisher's[12] test for the global significance of estimated optimal scores for qualitative variables. The F–statistic for all variables was significant at the 1% level, indicating that all variables included in the model are relevant. So, we conclude that there are highly significant differences amongst the workers hired by MZA, and that these differences are captured by their initial job, leaving reason, marital status, place of birth, and section in which they were employed. Fisher[12] highlights the importance of this result, since it is only under this condition that the scores estimated by the OS model are meaningful.

The variables that load highest in the first component were found to be: enrolment age, marital status, initial job level, and initial salary. Taking into account the quantifications produced by OS and factor loadings, we interpreted the first principal component in terms of experience before the worker joined the firm, or "prior experience". The variables that load high in the second component are experience in the firm, and leaving reason. The second component is, therefore, associated with "permanence in the firm".

The first two categorical principal components are represented in figure 1. In

that figure, the points are labelled with experience in the firm, and with leaving reason. At the top right hand side of figure 1 we find the workers who abandoned the company for causes of retirement, illness, or death. In the same region of the figure we find the workers with a long period of attachment to the company. At the bottom left hand side of figure 1 we find two main groups of workers who left through resignation, dismissal, and redundancy. We see that workers with a short career in the company are also to be found in this area of figure 1.

It is clear from the representation that there is a strong negative association between experience in the company and leaving reason. There are other relationships of interest between the variables, and these will be represented using the method of Property Fitting.

3.3 Property Fitting

Property Fitting is a regression based technique that can be located within the context of Biplots (Gower and Hand[16]). Details of the particular form of the algorithm implemented here can be found in Kruskal and Wish[27], Schiffman et al.[43], and Mar Molinero and Mingers[32]. Property Fitting draws normalized vectors of unit length through the space of the principal components in the direction in which a property of the data grows. The coordinates of vectors can be interpreted as directional cosines. The longer the projection of the vector, the more relevant is the two dimensional plot to the interpretation of the results. The angle between any two vectors depends on the correlation between the variables involved. Acute angles indicate positive correlation, the smaller the angle the higher the correlation. Orthogonal vectors are associated with a lack of correlation.

The vectors obtained are shown in figure 2. Take, for example, the variable enrolment age: the associated vector points towards the right hand side of the figure, indicating that the older a worker was when joining, the more to the right of the configuration he will be plotted. Two further features highlighted in figure 2 are whether the observation contains imputed values or not, and initial job level. Most workers for whom the leaving cause was estimated are plotted towards the left and towards the bottom of the figure. Two groups of workers have been circled in this area. Both groups are similar in the sense that the workers did not stay long in the firm and left the firm by resignation or dismissal. But, as the oriented vectors indicate, the two groups are different in the sense they correspond to workers with different initial salaries and job levels. The first group, situated towards the left, is made up of individuals who did not stay long in the firm, were rather young when they joined, and were hired in low level jobs with low initial salaries. The second group of workers with low job tenure, situated at the bottom of the figure 2, were hired with high initial salaries and at high level jobs. We also see that initial salary, initial job level, section, and place of birth are not related to the absence in the worker's file of the leaving reason. This confirms the validity of the missing data procedure used in the analysis.





Fig. 2. Workers plotted in the first two categorical principal components by missing data for leaving reason, and projection of Property Fitting vectors

In conclusion, figure 2 jointly represents the workers and the variables measured on them. It would be reasonable to assume that there is association between joining age, and seniority –i.e., that workers who join when young are less well qualified and are assigned lower seniority than workers who join when they are older. This implies a positive correlation between enrolment age, and initial wage. Such positive correlation is present, as can be seen by the acute angle between the vectors enrolment age and initial salary in figure 2. An examination of figure 2 also reveals a positive correlation between enrolment age and initial job performed, indicating that in the firm MZA experienced workers tended to be hired into more skilled jobs. This is consistent with the influence of external labour markets on internal wages through ports of entry.

Two vectors point towards the top right hand side of figure 2; these are associated with experience in the firm, and leaving reason. The angle between these two vectors is acute and small, indicating a strong positive correlation between these two variables. This means that workers who remained in the firm for a long period of time left it for reasons of retirement, illness or death; and that workers with short periods of employment in the firm left it for reasons of dismissal, redundancy, or resignation. In other words, workers who "survived" the initial period of employment stayed on until retirement or death, or until they had to leave for reasons of illness: they do not resign. These results are consistent with the specific human capital theory. This theory predicts that resignation and dismissal rates decrease as experience in the firm increases (Parsons[37]). However, these findings are at odds with the outcomes to be expected under the deferred compensation model, which states that, once the specific training period is over, salaries exceed worker's marginal productivity, dissuading workers from leaving, but incentivizing dismissals (Medoff and Abraham[33]; and Siebert and Addison[44]).

It is also apparent in figure 3 that leaving reason and initial salary are not associated –the relevant vectors are orthogonal. There is also a lack of association between initial salary and experience in the firm. This is consistent with the view that conditions of employment were determined by the labour market and did not influence workers' permanence in the firm. Permanence in the firm is much better explained by the long-term relationships that ILMs need in order to transform into benefits the costs incurred during the initial employment period. Under the human capital theory, and under the deferred compensation model, ILM relationships act as a mechanism for the selection of the most stable workers (Becker and Stigler[4]; Becker[3]). The explanation is that, since both the firm and the workers share specific training costs, there are incentives not to break the labour relationship through premature dismissals or resignations. Furthermore, the governance structures that arise under ILMs dissuade both workers and employers from adopting opportunist behaviours (Wachter and Wright[52]).

4 Conclusions

In this paper we have addressed methodological and business questions. The methodological questions relate to missing values and to the correct procedures to deal with a mixture of qualitative and quantitative variables. The treatment of missing data is not straightforward. A model based only on observations for which complete information is available may not be representative of the population unless the data is MCAR. In this study we have established that the data was not MCAR, and that the missing data contained valuable information about the process under study. We followed a several step process for missing value imputation that culminates on the maximization of internal consistency. This made it possible to operate with the full data set without dropping any observations.

A further difficulty was the co-existence of qualitative and quantitative variables. OS procedures are appropriate in these cases. OS was introduced by Fisher[12], but it has seldom been applied to the analysis of business problems. In



this paper we have used OS in order to transform qualitative information into quantitative. After this, we have applied standard tools of multivariate analysis. We have also used hypotheses tests originally developed by Fisher[12] in order to establish the relevance of the variables in the model.

Turning now to the business agenda, the study of the way in which ILMs developed and evolved has taken momentum in the last decades (Elvira and Graham[10]; Osterman[36]). Much recent work on ILMs has focused on their evolution and actual implications, arguing that some of the key principles of ILMs have been now dismantled (Grimshaw et al.[17]).

The theory suggests that, when ILMs are present, companies relate to the external market through the so-called «ports of entry». Companies use these in order to hire new workers with no particular training. Workers are expected to undergo an initial period of hardship that will be eventually rewarded by means of wage and social benefits. The theory also predicts that most workers who exit the company in an irregular way do so soon after joining, since opportunity costs can be high for those who have accumulated some experience in the firm.

The main objective of this paper was the study of the introduction of ILMs in the Spanish railway company MZA at the end of the 19th century. The results reveal two characteristics observed in previous studies of labour relations: the existence of «ports of entry» for workers at low levels of qualification, and longterm labour relations. This paper, therefore, attempts to contribute to the understanding of the introduction of advanced management methods at the beginning of the Second Industrial Revolution, by means of a multivariate statistical approach, which is innovative in this area of research.

References

- 1.Barlett, M. S. The goodness of fit of a single hypothetical discriminant function in the case of several groups. Annals of Eugenics, 16, 199-214, 1951.
- 2.Barley, S. R. and Kunda, G. Bringing work back in. Organization Science, 12, 76-95, 2001.
- 3.Becker, G. S. Human capital: A theoretical and empirical analysis. New York: Columbia University Press, 1975.
- 4.Becker, G. S. and Stigler, G. J. Law enforcement, malfeasance, and compensation of enforces. Journal of Legal Studies, 3, 1-18, 1974.
- 5.De Leeuw, J. Least Squares Optimal Scaling of Partially Observed Linear Systems. In Recent developments on structural equation models: Theory and applications (eds. K. van Montfort , J. Oud and A. Satorra). Norwell, Mass.: Kluwer, 2004.
- 6.De Leeuw, J., Young, F. W. and Takane, Y. Additive structure in qualitative data: an alternating least squares method with optimal scaling features. Psychometrika, 41(4), 471-503, 1976.
- 7.Doeringer, P. B. and Piore, M. J. Internal labour markets and manpower analysis. Lexington, Mass.: Lexington Books, Heath, 1971.
- Drummond, D. Specifically designed? Employers' labour strategies and worker responses in British railway workshops, 1838-1914. Business History, 31, 10-11, 1989.

9.Drummond, D. Crewe: Railway town, company and people, 1840-1914. Aldershot: Scholar



Press, 1995.

- 10.Elvira, M. M. and Graham, M. E. Not just a formality: Pay system formalization and sexrelated earnings effect. Organization Science, 13(6) 601-617, 2002.
- 11.Ferrari, P and Annoni, P. Missing data in Optimal Scaling. Working Paper 2005-19, Dipartimento di Scienze Economiche e Statistiche. Universita di Milano, Milano, 2005.
- 12.Fisher, R. A Statistical methods for research workers (7th and 8th editions). Edinburgh: Oliver and Boyd, 1938, 1941.
- 13.Fitzgerald, R. British Labour Management and Industrial Welfare, 1846-1939. London: Croom Helm, 1988.
- 14.Fitzgerald, R. Employment relations and industrial welfare in Britain: Business ethics versus labor markets. Business and Economic History, 28 167-179, 1999.
- Gower, J. C. Fisher's Optimal Scores and Multiple Correspondence Analysis. Biometrics, 46, 947-961, 1990.
- 16.Gower, J. C. and Hand, D. J. Biplots. London: Chapman & Hall, 1996.
- 17.Grimshaw, D., Ward, K., Rubery, J. and Beynon, H. Organisations and the transformation of the Internal Labout Market. Work, Employment and Society, 15(1), 25-54, 2001.
- 18.Guttman, L. The quantification of a class of attributes: A theory and method of scale construction. In The prediction of personal adjustment (P.H. Horst ed.), 319-348. Social Sciences Research Council, New York, 1941.
- 19. Hald, A. A History of Mathematical Statistics from 1750 to 1930. New York: Wiley, 1998.
- 20.Hayashi, C. On the predictions of the phenomena from qualitative data and quantifications of qualitative data from the mathematico-statistical point of view. Annals of the Institute of Statistical Mathematics, 3, 69-92, 1952.
- 21.Howlett, P. Evidence of the existence of an internal labour market in the Great Eastern Railway Company, 1875-1905. Business History, 42(1), 21-40, 2000.
- 22.Howlett, P. Careers for the unskilled in the Great Eastern Railway Company, 1870-1913. London School of Economics Working Paper Series 63/01, London, 2001.
- 23.Howlett, P. The internal labour dynamics of the Great Eastern Railway Company, 1870-1913. Economic History Review, LVII(2), 396-422, 2004.
- 24.Idson, T. and Feaster, D. A selectivity model of employer-size wage differentials. Journal of Labor Economics, 8(1), 99-122, 1990.
- 25.Kerr, C. The Balkinization of Labor Markets. In Labor mobility and economic opportunity (E. W. Bakke ed.), pp. 92-110. Cambridge, MA: MIT Press, 1954.
- 26.Kruskal, J. B. Multidimensional Scaling by optimizing goodness of fit to a non-metric hypothesis. Psychometrika, 29(1), 1-27, 1964.
- 27. Kruskal, J. B. and Wish, M. Multidimensional Scaling. London: Sage, 1978.
- 28.Lazear, E. P. and Oyer, P. Internal and external labor markets: a personnel economics approach. NBER Working Paper Series, No.10192, Cambridge, MA, 2003.
- 29.Little, R. J. A. and Rubin, D.B. Statistical analysis with missing data. Hoboken, New Jersey: Wiley & Sons Inc, 2002.
- 30.Mackinnon, M. The Great War and the Canadian labour market: railway workers 1903-39. In Labour Market Evolution (G. Grantham and M. MacKinnon eds.), pp. 205-224. London: Routledge, 1994.
- 31.Mackinnon, M. Trade Unions and Employment Stability at the Canadian Pacific Railway, 1903-1929. In Origins of the Modern Careers (D. Mitch, J. Brown and M.H.D. van Leeuwen eds.), pp. 126-144. Aldershot: Ashgate, 2004.
- 32.Mar Molinero, C. and Mingers, J. Mapping MBA Programmes: an alternative analysis. Journal of the Operational Research Society, 58, 874-886, 2007.
- 33.Medoff, J. L. and Abraham, K. G. Are those paid more really more productive? The case



of experience. Journal of Human Resources, 16(2), 186-216, 1981.

- 34.Meulman J. J., van der Kooij, A. J. and Heiser, W. J. Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. In Handbook of quantitative methodology for the social sciences (D. Kaplan ed.), pp. 49-70. Thousand Oaks, CA: Sage Publications, 2004.
- 35.Neal, P. J. and Roberts, G. O. Optimal Scaling for partially updating MCMC algorithms. The Annals of Applied Probability, 16(2), 475-515, 2006.
- 36.Osterman, P. The Wage Effects of High Performance Work Organization in Manufacturing. Industrial and Labor Relations Review, 59(2) 187-204, 2006.
- 37.Parsons, D.O. The Employment relationship: Job attachment, work effort and the nature of contracts. In Handbook of Labor Economics (R. Layard and O. Ashenfelter eds.), pp. 789-848. Amsterdam: North-Holland, 1986.
- 38.Portillo, F., Mar Molinero, C. and Martinez, T. Interpreting a data base of railway workers using optimal scaling techniques. Kent Business School Working Paper Series, 127, Canterbury, UK, 2006.
- 39.Retherford, R. D. and Choe, M. K. Statistical Models for Causal Analysis. New York: John Wiley & Sons, 1993.
- 40. Rubin, D. B. Inference and missing data. Biometrika, 63 581-592, 1976.
- 41.Savage, L. J. On reading R.A. Fisher. The Annals of Statistics, 4, 441-500, 1976.
- 42.Savage, M. Discipline, surveillance and the career: Employment on the Great Western Railway 1833-1914. In Foucault, management and organisation theory (A. McKinlay and K. Starkey eds.), pp. 65-92. London: Sage, 1998.
- 43.Schiffman, S. S., Reynolds, M. L. and Young, F. W. Introduction to Multidimensional Scaling: Theory, Methods and Applications. London: Academic Press, 1981.
- 44.Siebert, W. S. and Addison, J. T. Internal labour markets: causes and consequences. Oxford Review of Economic Policy, 7(1), 76-92, 1991.
- 45.Stevens, S. S. Mathematics, measurement, and psychophysics. In Handbook of Experimental Psychology (S. S. Stevens ed.), pp. 1-49. New York: Wiley, 1951.
- 46.Sundstrom, W. A. Half a carrer: discrimination & Railroad Internal Labor Markets. Industrial Relations, 29(3), 423-440, 1990.
- 47.Takane, Y., Young, F. W. and De Leeuw, J. An individual differences additive model: An Alternanting Least Squares method with Optimal Scaling features. Psychometrika, 42, 7-67, 1980.
- 48.Tedde, P. La expansión de las grandes compañías ferroviarias españolas: Norte, MZA y Andaluces. In La empresa en la historia de España F. Comín and P. Martín-Aceña eds.), pp. 265-284. Madrid: Civitas, 1996.
- 49.Turner, T. Internal labour markets and employment Systems. International Journal of Manpower, 15(1), 15-26, 1994.
- 50.Van Buuren, S. and Van Rijckevorsel, J. L. A. Imputation of missing categorical data by maximizing internal consistency. Psychometrika, 57, 567-580, 1992.
- 51.Von Hippel, P. T. Biases in SPSS 12.0 missing value analysis. The American Statistician, 58 160-164, 2004.
- 52.Wachter, M. L. and Wright, R. D. The economics of internal labour markets. Industrial Relations, 29(2), 240-262, 1990.
- 53.Welsh, A. H. and Robinson, J. Fisher and inference for scores. International Statistical Review, 73, 131-150, 2005.
- 54. Young, F. Quantitative analysis of qualitative data. Psychometrika, 46(4), 357-388, 1981.





Couple-wise divorce rates in Japan

Shigeru Mase¹ and Kouji Yoshimoto²

¹ Department of Mathematical and Computer Sciences, Tokyo Institute of Technology, Tokyo, Japan

(E-mail: mase@is.titech.ac.jp)

² Department of Mathematical and Computer Sciences, Tokyo Institute of Technology, Tokyo, Japan

(E-mail: yoshimoto.k.aa@m.titech.ac.jp)

Abstract. In Japan, ratios of numbers of divorce to those of marriage per year exceed one-third from 2000 onward. This fact is frequently referred to as if one-third of married couples will divorce in future. This interpretation is doubtful since numbers of marriage has been rapidly decreasing in Japan because of declining birth rates and, on the other hand, numbers of divorce include those of married couples from older generations with higher birth rates. What are true couple-wise divorce rates, that is, rates of couples married in one year who will divorce afterward, in Japan? Estimates of these rates are important because short marriage durations will frequently result in children with insufficient parental cares and financial supports.

In order to estimate couple-wise divorce rates, we have to forecast numbers of divorce in future for couples married each year. This can be done using methods to construct life table. The result shows that actual couple-wise divorce rates will be over one-third for couples married from 1997 onward, contradictory to our first hypotheses that actual couple-wise divorce rates are much less than one-third. Moreover, it will be almost 40% for those married in 2002.

Keywords: couple-wise divorce rates, vital statistics, Japan.

1 Introduction

Divorce as well as marriage is one of important demographic factors. It lessens not only numbers of babies born but also increases children with insufficient parental and financial cares. In this paper, three kinds of divorce rates are discussed and compared, that is,

- mean annual no. of divorces per 1,000 populations,
- ratios of no. of divorces to that of marriages per year,
- ratios that couples married in a year who will divorce afterward.

The first one is called crude divorce rates (CDR) in literature and is used for cross-country comparison. There seem no established terms for the rest and, for convenience, we refer them to as annual divorce rates (ADR) and couple-wise divorce rates (CWDR) respectively in the sequel.

In Japan, ADRs have been larger than one-third since 1997 and this fact is frequently quoted as if one-third of married couples will divorce in future, that is, ADR is confused with CWDR. This interpretation is doubtful. Although annual ratios of death to birth varied from 80.8% to 111.48% from 2000 to 2010 in Japan, "person-wise death rates" should be always 100%. Numbers of



marriage has been rapidly decreasing in Japan because of declining birth rates and, on the other hand, numbers of divorce include those of married couples from older generations with higher birth rates. What are true CWDRs, that is, ratios of couples married in a year who will divorce afterward, in Japan?

In order to study this question, it is necessary to predict numbers of couples who married in a year and will divorce afterward. This resembles construction of life tables and can be treated in a similar way. Main differences are that one may not marry, may marry and divorce several times, and cannot divorce after his death.

In the following, we study the "life table" of divorce in Japan using divorce data of vital statistics of the Ministry of Health, Labour and Welfare, Japan.

2 Dataset

Basic data used is the vital statistics of the Ministry of Health, Labour and Welfare, Japan[5]. Divorce data is summarized in "Vital Divorce Statistics 2009"[6]. This dataset consists of number of divorced couples classified according to their marriage duration (in year) for 1947–2010. Several missing data was taken from the website[5].

We should remark the followings about this dataset:

- Only years of divorce are given and we cannot recover corresponding wedding years. For example, couples who divorced in 2010 with marriage duration 0 (i.e. less than one year) might marry either in 2009 or 2010.
- Couple's previous marital status are unknown.
- Numbers of couples with marriage durations over 20 years are added up till 1989 except 1970, 1975 and 1980.
- International couples one of which are Japanese are included.
- Years of divorce mean those when couples began to live separately and marriage durations mean periods couples lived together.
- There are relatively few couples marriage durations of which are unknown.

We constructed working data consisting of wedding years and divorce years for all couples in the original data as follows:

• Estimated marriage durations in intervals

 $I: [20, 25), [25, 30), [30, 35), [35, \infty)$

for those with divorce years before 1990 so that their proportions are the same as those of (nearest) 1970,1975 and 1980.

- Applied monotone spline fitting[1] to cumulative frequencies of marriage durations and, then, estimated numbers with marriage duration $d = 0, 1, 2, \ldots, 55$ for each divorce year 1947–2010. Here the largest duration d = 55 was chosen so that it is long enough and estimated marriage durations became almost decreasing for $d \geq 30$.
- Relatively few data for which marriage duration are unknown were distributed to $d = 0, 1, 2, \ldots, 55$ proportionally.



• For each couple with divorce year n and marriage duration n, their divorce day $e = 1, 2, \ldots, 365$ was chosen uniformly in the year n and, then, corresponding wedding day (the day starting to live together) $m = 1, 2, \ldots, 365$ was chosen uniformly in 365 days ending with e. 27 leap days were taking account of.

Thus we could get data of numbers of divorced couples classified by both wedding years and marriage durations.

3 Probability law of marriage durations

The basis of life table is the assumption that life spans obey probability laws. Before starting analysis, it is worth checking if marriage durations also obey probability laws since marriages and divorces seem more subjective matters than births and deaths and may be influenced heavily by current socio-economic conditions. Above all, one may not marry and, if married, one may not divorce or may die before divorce. The left of Fig. 1 shows relative frequencies of marriage durations of couples who divorced in 2001–2010 and the right is the corresponding probabilities estimated by least square method. It is seen that marriage durations shows almost the same patterns at least for these period. Couples with marriage duration d = 3 has the highest probability 0.0797. The probability for $d \leq 5$ and $d \leq 9$ is 0.330 and 0.533 respectively. The median and the mean is 8.26 and 11.0 respectively.



Figure 1. Relative frequencies of durations of marriage of couples who divorced from 2001 to 2010 (left) and estimates of corresponding probabilities by least square method (right)

4 CDRs and ADRs of Japan

The left of Fig. 2 is plots of populations and ADRs of Japan for 1951–2010. The right is plots of CDRs and marriage rates per 1,000 population. Both CDRs and ADRs were calculated by yearly data and show fairly similar variations.

CDRs vary from 1.96 to 2.27 for 2001–2010. According to the international comparison around 2007, Japanese CDR is 26th. Russian and American CDR, which was first and fourth, are about 4 and 5. Korean CDR 2.6 is 9th and the highest in Asia. Chinese CDR 1.6 is 32th. CDRs in Japan took a peak value 1.53 in 1899, became 0.99 in 1920, varied around 1.0 from 1947–1950,



decreased gradually afterward and showed 0.73 in 1963 which was the lowest value after the World War II.

ADRs vary from 0.346 to 0.383 for 2001–2010. This is the reason of the widespread belief that one-third of newly married couples in Japan will divorce afterward.



Figure 2. Left: Populations (per 10 million) and ratios of divorces and marriages(%) for 1950–2010, Right: Crude divorce rates and marriage rates (per 1,000) for 1950–2010.

5 Couple-wise divorce rates in Japan

In life table analysis, one is interested in the decrease of a population with the same birth year. In the same way, we can consider a "life table of marriage", that is, the decrease of a population of couples by divorce with the same wedding year. In particular, it is important to guess CWDR, rates that couples who will divorce eventually. Although both life tables are similar in nature, there are several distinctions. Although rates that person dies eventually is 100%, there are no such values for CWDRs decided beforehand. One may divorce several times and there is possibility that CWDRs for first marriage are different for those of remarried couples. One couple may not divorce or may die before divorce.

In life table analysis, Lexis diagram[2] is well-known. Fig. 3 is the Lexis diagram for marriage duration. It shows numbers of couples with x-axis being wedding years and y-axis being marriage durations. In true Lexis diagrams, x-axis is for years of birth and y-axis for ages at death. Numbers with x = 1 corresponds to those couples who married and divorced in the same year and, therefore, are smaller than others. The shaded area in Fig. 3 is the area where we have data. There are no available data for the upper-left triangular area. The lower-right triangular area belongs to future and we have to predict their values using existing ones. In order to predict the CWDR for couples married in 2010, it is necessary to predict their numbers of divorce after $d = 2, 3, \ldots, 55$ year.





Figure 3. Lexis diagram of marriage and divorce

The procedure to predict CWDRs of couples married in 1947–2010 using spline function fitting is as follows:

1. Fit cubic splines to times series of divorce numbers with length 64

$$X_{1948-d}^{(d)}, X_{1949-d}^{(d)}, \dots, X_{2011-d}^{(d)}$$

for each marriage duration $d = 2, 3, \ldots, 55$.

2. Predict numbers of divorce after 2012 - d

$$\hat{X}_{2012-d}^{(d)}, \hat{X}_{2013-d}^{(d)}, \dots, \hat{X}_{2010}^{(d)}$$

using thus fitted spline functions.

3. Add 55 values of predicted divorce numbers and known divorce numbers before 2010 and get total divorce numbers of couples married in 1947–2010. Dividing them by corresponding marriage numbers, we get CWDR.

Predicted values using cubic splines may vary considerably depending on nodes used. Comparison of predicted values for nodes 25(5)55 and 64 showed almost the same predicted values on average. But individual values may be non-realistic, such as negative values, depending on node numbers. Cubic splines are mainly tools for interpolation but not for prediction. So results by spline prediction seem not reliable as seen in Fig. 4.

A more reliable prediction method for time series X_t is the linear state space (STL) model[4]. This model has the following structure:

$$X_{t} = M_{t} + S_{t} + \epsilon_{t}, \quad \epsilon_{t} \sim N(0, \sigma_{\epsilon}^{2}),$$

$$S_{t+1} = -S_{t} - S_{t-1} - \dots - S_{t-s+2} + w_{t}, \quad w_{t} \sim N(0, \sigma_{w}^{2})$$

where S_t is a periodic component with period s, M_t is a local trend. ϵ_t and w_y are iid Gaussian errors with respective variances σ_{ϵ}^2 and σ_w^2 . σ_w^2 may be equal to 0 and there may not exist a periodic component S_t at all.

Table 1 and Fig. 4 shows results of CWDR prediction. Table 2 shows predicted results of numbers of marriage (per 10,000 population), numbers of divorces (per 10,000 population) and CWDRs (%) for 2011–2020 by STL models. CWDRs increases gradually around 40% The variation of CWDRs by STL model is fairly similar to that of ADRs. On the other hand, the variation of CWDRs by spline method differs from them and almost monotonically increases. Table 3 is a summary of how they increase.

6 Conclusion

From Fig. 4, predictions by STL model seem plausible although there may be other prediction methods. Contrary to our first opinion, CWDRs of Japan are larger than 1/3 from 1997. Moreover, it is about 40% for couples married in 2001–2010.

Changes of CWDRs anticipate those of ADRs by a few years and larger than ADRs by 2.38–5.90% since 1966. Fig. 5 shows differences of CWDRs and ADRs. It has three peaks at 1968,1976 and 1997. First two peaks seems to be results of increases of marriages of females and males of the first baby boomers respectively. The differences became smaller gradually in 1879–1993. This seems a result of subsequent increase of divorces and gradual increase of divorces of middle and older age. The third peak seems to be a result of the second marriage boom (marriages of children of the first baby boomers) accelerated by a decrease of younger populations. The differences became again smaller after 2000.

From Fig. 6, we can see a fairly similar pattern of increase of divorce regardless of marriage durations. In particular, a rapid increase from 1990 and a stagnation and/or a decrease from 2001 are common. As Table 2 suggests, this stagnation seems to continue in future. A long-standing effect of baby boomers after the World War II has been almost diluted. It becomes more difficult to divorce as well as marriage after a long recession. Also it may be a result of new social norms on marriages and families. In fact, it has been now common in Japan for youngs to live together without getting married and get married only after having children. Also it has been becoming common not to marry at all.

	min	median	max	mean
CWDR(LST)	9.22% (1961)	22.0% (1980)	40.1% (2002)	22.2%
CWDR(spline)	11.0% (1958)	21.7% (1980)	38.7% (2003)	24.2%
ADR	7.54% (1963)	18.95% (1980)	37.5% (2003)	19.6%
CDR	0.732(1963)	1.22(1980)	2.27(2002)	1.31

Table 1. Summary of divorce rates (1947–2010)



divorces and corresponding C (CD1) for 2010 2020										
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
marriage	69.1	68.3	67.4	66.5	65.6	64.7	63.8	63.0	62.1	61.2
divorce	26.2	26.1	25.9	25.8	25.7	25.6	25.5	25.4	25.2	25.1
CWDR	37.9	38.2	38.5	38.8	39.2	39.5	39.9	40.3	40.7	41.0

Table 2. Predictions by STL model of number of marriages and
divorces and corresponding CWDR for 2010–2020

Table 3. Summary of how three divorce rates increase for $1951{-}2010$

	CWDR by spline	CWDR by STL	ADR
$\geq 20\%$	$1976 \sim$	$1978 \sim$	$1982 \sim$
$\geq 25\%$	$1985 \sim$	$1983 \sim (\text{except } 1988)$	$1993 \sim$
$\geq 30\%$	1988~	1996~	$1998\sim$
$\geq 1/3$	1994~	$1997 \sim$	$2000\sim$
$\geq 35\%$	$1996 \sim (\text{except } 1999, 2000)$	$1998 \sim$	$2001\sim$
$\geq 40\%$	none	2002	none



Figure 4. Plots of CWDRs by STL and spline models and ADRs for couples married in 1951–2010





and ADRs for 1956-2010



Figure 5. Plots of differences of CWDRs Figure 6. Plots of numbers of divorces with marriage duration d for divorce years 1947-2010

References

- 1. F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation, SIAM Journal on Numerical Analysis, 17, 238-246, 1980.
- 2. H. U. Gerber. Life Insurance Mathematics, 3rd ed., Springer-Verlag, Berlin Heidelberg, 1997.
- 3. P. J. Green and B. W. Silverman. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach, Chapman and Hall, 1994.
- 4. A. C. Harvey. Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge, 1989.
- 5. The Statistics Bureau and the Director-General for Policy Planning of Japan Statistics and Information Department Minister's Secretariat, e-Stat, The Ministry of Internal Affairs and Communications, Japan, 2012. http://www.stat.go.jp/english/index.htm
- 6. Health and Welfare Statistics Association. Vital Divorce Statistics 2009 (in Japanese), the Ministry of Health, Labour and Welfare, Japan, 2010. http://www.hws-kyokai.or.jp/english-index.htm



Data analysis and reliability models of pumping stations of the haoud Elhamra-Bejaia pipeline (Algeria)

^{1*}Medjoudj Rafik, Imine Aghiles², ¹Medjoudj Rabah, ²Aissani Djamil ¹Electrical Engineering Department, University of Bejaia, Algeria ²LAMOS Laboratory, University of Bejaia, Algeria

Abstract, in this paper, a case study for reliability data analysis of pumping stations of a pipeline is introduced. By modeling the times to failure and the mean times between failures of components at these stations, the Weibull distributions were obtained on the basis of Kolmogorov Smirnov tests. When analyzing the shape parameters of the obtained distributions, it was observed that their values are less than unity, indicating that, theoretically, the considered components are in a youthful period or that the components have been operating for about 25 years. These statements are contradictory with practical considerations (e.g., an aging period) and the shape parameters values should be greater than the unity. Regarding also the multiple failure causes of the considered components, the adjustment of data using mixed Weibull distributions is deemed adequate by the tests and the shape parameters values are globally greater than the unity and the trend confirms the nature of aging equipment. The reliability evaluation of a station's components and the reliability change for different variants of the pipeline have been ascertained.

Keyword: Pumping station, Reliability, Weibull distribution.

1 Introduction

Reparable electrical components were usually modeled using two state (working-failure) diagrams. Nowadays, this assumption is relaxed in favor of multi-degraded items. In reference [1], the authors stated in their review of a Markov process that deterioration of the equipment is modeled as occurring in a limited number of steps. It is assumed that the duration of each stage of deterioration as well as times for repairing failed components is exponentially distributed. During the last decade, the Weibull-Markov approach has been applied to electrical components and systems interruption modeling and maintenance applications. Introduced by Van Castaren [2, 3] for state probabilities, transition probabilities and frequencies assessment and cost evaluation, it was reproduced by Pivatolo [4] for system generation reliability evaluation. Recently we have compared both the Markov method and the Weibull-Markov approach in interruption modeling of electrical distribution systems. It has been proved that the results obtained using this latter were more realistic [5, 6]. This paper develops reliability models of the components constituting the different pumping stations of the pipeline Haoud Elhemra-Bejaia and finds reliabilities of three different configurations of the same stations. As for the statistical treatments, two modeling methods are used and compared: the first uses simple distributions, however the second introduces the mixed distributions.



2 Description of the system

2.1 Haoud Elhamra Bejaia pipeline

Sonatrach (National Society of Transportation and Marketing of Hydrocarbons) was born December 31, 1963. It was considered to be the instrument of state intervention in the oil sector, alongside French companies. Until 1966, it was confined to the mission of the transport and marketing. But then it became 'National Society for the Research, Production, Transportation, Processing and Marketing of Hydrocarbons'. Currently, it is the first company in Africa, It is ranked 12th among global oil companies. It is the second explorer of LNG and LPG and the third largest explorer of natural gas in the world. Its global production (all products) was estimated to 230 million tons in 2006. Our study concerns one part of the oil transportation by pipeline in Algeria, the line HEH (Houdh El Hamra in the sahara) - Bejaia (petroleum port). The maximum flow is estimated to 2540 m³/hr and the actual transit between 700m³/hr and 1800m³/hr. Depending on the demand and the use of principal stations, the flow can reach 2000m³/hr. The pipeline " OB1 24 / 22 " is the first structure transport via pipeline installed in Algeria, by the oil company managers of SOPEG in 1959. It connects the storage facility (HEH) and marine terminal (Bejaia). The pipeline diameter OB1 24 `` (609.60 mm) and HEH-Bejaia has a total length of 660,721 km and it is composed of two sections: The diameter of the first section line between HEH and neck Selatna is "24"" (609.60 mm) for the length of 533,217 km. The second one, with diameter 22" (558.8 mm), is connecting the neck Selatna and marine terminal on a length of 127,551 km. HEH-Bejaia line, consists of : A departure terminal HEH-SP1; three (3) main pumping stations SP1 bis, SP2 and SP3; four (4) intermediate pump stations (satellites) SPA, SPB, SPC and SPD; terminal arrival Bejaia; an oil port. The station SPA (Touggourt, ALT = 89.5, PK = 135129, the station SPB (Oumach, ALT = 83, PK = 189.83), the station TPS (Oued El Fida, ALT = 411.75, PK = 443.083) and the station SPD (Beni-Mansour, ALT = 356, PK = 573.62) each have a turbo-pump and auxiliaries consisting of a combustion turbine "THMs" and "Guinard" Pump.

2.2 Operation mode of the pipeline

To perform a reliability analysis, it is necessary to know the functional and technical specifications of the system components. The line can operate in three modes.

Each mode is characterized by the speed of operation and the elements to start. The operation mode is represented as a variant of the system. The following table (1) represents the different operation modes of the system.



Mode	Flow (m ³ /h)	Tons (MTA)*	Number of groups per station
1 st Mode	700 to 800	5.52	1EPS/SP1, 1MPS/SP2, 1 MPS /SP3
2 ^{sd} Mode	1400	9.65	2 EPS/SP1, 2 MPS/SP2, 2MPS/SP3
3 th Mode	1700 to 1800	12.41	3 EPS/SP1,1 TP(TP1 or TP2) / SP1bis, 3 MPS/SP2,3 MPS/SP3

Table 1. Operation modes of the pipeline

Starting with an annual quota set by the marketing department, we determine the hourly flow transported by the pipeline. It will be calculated based on the density of the product to transit. To attain 16 MT / year of crude oil, its density must be 779 Kg/m³. We will have an hourly capacity of:

$$\frac{6.10^9 / 365 \times 24 \text{ Kg/h}}{779 \text{ Kg/m}^3} = 2354 \text{ m}^3 / h$$

To determine the starting elements in the system, in order to attain that flow along the line from the departure terminal SP1, it must take into account the speed provided by the elements of line, and the pressure provided by each station.

3. Reliability of the each variant of the system

The objective of this analysis is to reduce the overall system to a set of functional variants of the system. From a reliability point of view, each variant is composed of a set of elements, where the failure of each of them causes the failure of the variant (i.e., series configurations).

 $R_{TP}(t)$: Reliability of the turbo-pump of type THM (satellite stations and TPC) $R_{MPS}(t)$: Reliability of motor Pump set

 $R_{EPS}(t)$: Reliability of electro-pump set

 R_{vari} (t) : Reliability of variant i

Variant 1:

This variant represents the operation of stations working with an hourly capacity of 800 m3 / h. In this case, each station of the principal stations operates with one equipment only among the five. One Electro Pump Set in SP1 and 1 motorpump set in two other stations (SP2, SP3) will operate. So this variant represents a series configuration of all stations and each station represents the configuration of (1/5). Its reliability is presented as follows:

$$\begin{split} R_{\text{varl}}\left(t\right) &= R_{\text{SP1}}(t) \times R_{\text{SP2}}(t) \times R_{\text{sp3}}(t) \\ R_{\text{SP1}}\left(t\right) &= C_{5}^{1} R_{\text{EPS}}(t) \times \left[1 - R_{\text{EPS}}\left(t\right)\right]^{4} \times C_{5}^{2} R^{2}_{\text{EPS}}(t) \times \left[1 - R_{\text{EPS}}\left(t\right)\right]^{3} + C_{5}^{3} R^{3}_{\text{EPS}}(t) \times \\ \left[1 - R_{\text{EPS}}\left(t\right)\right]^{2} + C_{5}^{4} R^{4}_{\text{EPS}}(t) \times \left[1 - R_{\text{EPS}}\left(t\right)\right] + C_{5}^{5} R^{5}_{\text{EPS}}(t). \end{split}$$



 $\begin{aligned} R_{SP2}(t) &= C_5^{l} R_{MPS2}(t) \times [1 - R_{MPS2}(t)]^{4} \times C_5^{2} R_{MPS2}^{2}(t) \times [1 - R_{MPS2}(t)]^{3} + C_5^{3} \\ R_{MPS2}^{3}(t) \times [1 - R_{MPS2}(t)]^{2} + C_5^{4} R_{MPS2}^{4}(t) \times [1 - R_{MPS2}(t)] + C_5^{5} R_{MPS2}^{5}(t). \\ R_{SP3}(t) &= C_5^{l} R_{MPS3}(t) \times [1 - R_{MPS3}(t)]^{4} \times C_5^{2} R_{MPS3}^{2}(t) \times [1 - R_{MPS3}(t)]^{3} + C_5^{3} \\ R_{MPS3}^{3}(t) \times [1 - R_{MPS3}(t)]^{2} + C_5^{4} R_{MPS3}^{4}(t) \times [1 - R_{MPS3}(t)] + C_5^{5} R_{MPS3}^{5}(t). \end{aligned}$

Variant 2 :

This configuration allow to the system to bring the flow to 1400 m³/h. The SP1 requires 3 EPS and 2 MPS for each of the two other stations. The SP1 station represents the 3/5 configuration. The SP2 and SP3 each represent the 2/5 configuration. The reliability is given by:

$$\begin{split} &R_{var2}\left(t\right) = R_{SP1}(t) \times R_{SP2}(t) \times R_{SP3}(t) \\ &R_{SP1}\left(t\right) = C_5^3 \; R_{EPS}^3(t) \times \left[1 - R_{EPS}\left(t\right)\right]^2 + C_5^4 \; R_{EPS}^4(t) \times \left[1 - R_{EPS}\left(t\right)\right] + \; C_5^5 \; R_{EPS}^5(t). \\ &R_{SP2}\left(t\right) = C_5^2 \; R_{MPS2}^2(t) \times \left[1 - R_{MPS2}\; (t)\right]^3 + C_5^3 \; R_{MPS2}^3(t) \times \; \left[1 - R_{MPS2}\; (t)\right]^2 + C_5^4 \\ &R_{MPS2}^4(t) \times \left[1 - R_{MPS2}\; (t)\right] + \; C_5^5 \; R_{MPS2}^5(t). \\ &R_{SP3}\left(t\right) = C_5^2 \; R_{MPS3}^2(t) \times \left[1 - R_{MPS3}\; (t)\right]^3 + C_5^3 \; R_{MPS3}^3(t) \times \left[1 - R_{MPS3}\; (t)\right]^2 + C_5^4 \\ &R_{MPS3}^4(t) \times \left[1 - R_{MPS3}\; (t)\right] + \; C_5^5 \; R_{MPS3}^5(t). \end{split}$$

Variant 3:

In this variant, the SP1 station operates with three EPS. For the stations SP2 and SP3 two MPS are planned, and one TP for the two other stations, namely SP1bis and SPB.

$$\begin{split} & R_{var2}\left(t\right) = R_{SP1}(t) \times R_{SP2}(t) \times R_{SP1bis}(t) \times R_{SP3}(t) \times R_{SPB}(t) \\ & R_{SP1}\left(t\right) = C_{5}^{3} R_{EPS}^{3}(t) \times \left[1 - R_{EPS}\left(t\right)\right]^{2} + C_{5}^{4} R_{EPS}^{4}(t) \times \left[1 - R_{EPS}\left(t\right)\right] + C_{5}^{5} R_{EPS}^{5}(t). \\ & R_{SP2}\left(t\right) = C_{5}^{3} R_{MPS2}^{3}(t) \times \left[1 - R_{MPS2}\left(t\right)\right]^{2} + C_{5}^{4} R_{MPS2}^{4}(t) \times \left[1 - R_{MPS2}\left(t\right)\right] + C_{5}^{5} \\ & R_{MPS2}^{5}(t). \\ & R_{SP3}\left(t\right) = C_{5}^{3} R_{MPS3}^{3}(t) \times \left[1 - R_{MPS3}\left(t\right)\right]^{2} + C_{5}^{4} R_{MPS3}^{4}(t) \times \left[1 - R_{MPS3}\left(t\right)\right] + C_{5}^{5} \\ & R_{MPS3}^{5}(t). \\ & R_{SP1bis}\left(t\right) = R_{TP1}^{3}(t) \\ & R_{SP1B}\left(t\right) = R_{TP2}^{3}(t) \times \left[1 - R_{MPS3}\left(t\right)\right]^{2} + C_{5}^{4} R_{MPS3}^{4}(t) \times \left[1 - R_{MPS3}\left(t\right)\right] + C_{5}^{5} \\ & R_{MPS3}^{5}(t). \end{split}$$

4. Data analysis

The data collected from the rural industry was a very important step in our work. These data concerns the different mean times of the components of the



different stations, and the period is from 1/01/1997 to 12/31/2007 or a period of 11 years.

4.1 Adjustment tests

The following table shows the results of adjustment tests of the mean lifetimes of the different components by using the Weibull distribution.

Vari	Variables		Adjusted	Parameters		Test of	K-S	Validation
			law		1	D _n	d _{n,0.05}	test
V 1	SP1	2	Weibull	Veibull β=1.1139846 η=9194.127654		0.171	0.172	Not rejected
Л	Booster	3	Exponential	λ=0:0001192653		0.332	0.172	Not rejected
v2	SP1 GEP	3	Weibull	β=5.348812	η=2600.770844	0.076	0.172	Not rejected
112	SFIGE	5	Exponential	λ=0.0004214963		0.202	0.172	Not rejected
X3	SP1bis TP 77		Weibull	β=0.64560084	η=612.0831252	0.103	0.172	Not rejected
			Exponential λ=0.0011918398			0.122	0.172	Not rejected
V4	SP2	6	Weibull	β=0.8119306	η=7762.198197	0.106	0.270	Not rejected
717	Booster	0	Exponential	λ=0.0001252087		0.119	0.270	Not rejected
X5	SP2 MPS	29	₉ Weibull β=0.6284866 η=1266.52		η=1266.52	0.107	0.309	Not rejected
210	51 2 MI 5	27	Exponential	λ=0.0005792701	=0.0005792701		0.309	Not rejected
V6	SP3	0	Weibull	β=0.6313168	η=4401.447566	0.107	0.328	Not rejected
ло	Booster	0	Exponential	λ=0.000149379		0.136	0.328	Not rejected
V 7	CD2 MDC	12	Weibull	β=1.3268163	η=4212.132403	0.219	0.294	Not rejected
Λ/	SP3 MPS	15	Exponential	λ=0.0002756515		0.477	0.294	Not rejected
vo	SDD TD	11	Weibull	β=0.58311194	η=276.9759413	0.219	0.294	Not rejected
X8 SPB TP		4	Exponential	λ=00022518519		0.477	0.294	Rejected

 Table 2. Results of adjusted test of mean times between failures using parametric laws

After analyzing the results obtained by the adjustment tests, it was observed that their values are less than unity indicating that, theoretically, the considered components are in a youthful period, which is contradictory with practical considerations (aging period). For this reasons we have proceeded to an adjustment test using the mixed Weibull function as the distribution function of the mean lifetimes of the components.

Variables		n	Adjusted law	Parameters		justed law Parameters $D_n = d_{n.0.0}$		K-S d _{n.0.05}	Validation test
X1	SP1 Booster	3	-	-	-	-	-	-	
X2	SP1 GEP	3	-	-	-	-	-	-	
X3	SP1bis TP	77	Mixed Weibull	$\begin{array}{l} \beta_1 = 1.007182 \\ \beta_2 = 1.141648 \\ P_2 = 0.4010147 \end{array}$	$\begin{array}{l} \eta_1 = 96.94063 \\ \eta_2 = 1566.4557 \\ P_2 = 0.5989853 \end{array}$	0.134	0.148	Not rejected	
X4	SP2 Booster	6	-	-	-	-	-	-	
X5	SP2 MPS	29	Mixed Weibull	$\begin{array}{l} \beta_1 = 1.1562567 \\ \beta_2 = 0.9798344 \\ P_2 = 0.5006783 \end{array}$	$\eta_1 = 342.6612 \\ \eta_2 = 4113.4872 \\ P_2 = 0.4993217$	0.149	0.24	Not rejected	
X6	SP3 Booster	8		-	-	-	-	-	
X7	SP3 MPS	13	Mixed Weibull	$\beta_1 = 5.36067$ $\beta_2 = 2.553703$ $P_2 = 0.381716$	$\eta_1 = 669.1127$ $\eta_2 = 5150.0732$ $P_2 = 0.618284$	0.231	0.361	Not rejected	
X8	SPB TP	11 4	Mixed Weibull	$\beta_1 = 1.020364 \\ \beta_2 = 1.224585 \\ P_2 = 0.5612362$	$\eta_1 = 71.5822 \\ \eta_2 = 1037.4013 \\ P_2 = 0.4387638$	0.084	0.121	Not rejected	

Table 3. Results of adjusted test of mean times between failures using mixed

 Weibull distribution

5. Reliability study of the system

5.1 Turbo-pumps and motor-pumps reliabilities

The figures (1) represents the changing reliabilities of different components of the pumping stations. Their expressions are given as follows:

$$R_{TP\,1bis} = 0.4010147 \exp\left[-\left(\frac{t}{16.94063}\right)^{1.007182}\right] + 0.5989853 \exp\left[-\left(\frac{t}{1566.45570}\right)^{1.141048}\right]$$

$$R_{MPS\,2} = 0.5006783 \exp\left[-\left(\frac{t}{342.6612}\right)^{1.1562567}\right] + 0.4993217 \exp\left[-\left(\frac{t}{4113.4872}\right)^{0.9798344}\right]$$

$$R_{MPS\,3} = 0.381716 \exp\left[-\left(\frac{t}{669.1127}\right)^{5.360671}\right] + 0.6182845 \exp\left[-\left(\frac{t}{5150.07322}\right)^{2.553703}\right]$$

$$R_{TPSPB} = 0.5612362 \exp\left[-\left(\frac{t}{71.5822}\right)^{1.020364}\right] + 0.4387638 \exp\left[-\left(\frac{t}{1037.4013}\right)^{1.224585}\right]$$





The figure (1) shows respectively the reliabilities of turbo-pumps and the motorpumps installed in the pumping stations. We remark that the and the motorpumps are more reliable than the turbo-pumps, this because the stations contain 5 motor-pump sets installed in parallel, While it contains only one turbo-pump in the stations SP1bis and SPB.

. 5.1 Reliability of different variants of the pipeline

The figure (3) shows the reliabilities of the different variants of the pipeline.



Figure1: Reliabilities of three variants of the system

We remark that the first variant is more reliable, these because it operates by only one component in each station and the four others are in standby.

Conclusion

Pumping station components are an important part of the oil transportation network, and they are often subject to failures due to environment and exploitation conditions. In this work, three variants of system reliability were given taking into account the state of the components of the pumping stations. A real case study of an oil pipeline was considered to explain the strategy used variants of the system, and we observe that the reliability of variant three is the most important, based on the statistical treatments of data collected in a rural industry. The results were presented and discussed, and the reliability of the components were determined under real functioning conditions.

References

1.G.K Chan, Asgarpoor, S. Optimum maintenance policy with Markov process. *Electric Power Systems Research*, 76, 452-456, 2006.

2.J Van Castaren, Reliability assessment in electrical power systems: the Weibull-Markov stochastic model. *IEEE Trans on Industry Applications*, 36(3), 911-915, 2000.

3.J Van Castaren, Assessment of interruption costs in electric power systems using Weibull- Markov model. *PhD Thesis, Department of electric power engineering: Chalmers University of technology, Sweden, 2003.*

4.A Pievatolo,. The downtime distribution after a failure of a system with multistate independent components. *IMATI-CNR Technical report, Presented at the Mathematical Methods on Reliability conference, MMR*, Glasgow, 2007.

5.R Medjoudj,, Aissani, D., Boubakeur, A and Haim, K.D., Interruption modelling in electrical power distribution systems using the Weibull-Markov Model. *Journal of Risk and Reliability*, DOI:10.1243/1748006XJRR215, 2009.

6.S-H Jin, Lu, J-G., Wan, Y-P and Sun, S-G., Application of the Bayes' theory to the failure rate evaluation of electrical apparatus. *Journal of Zhejiang University Science A*8(3): 501-505, 2007,.



MDG and 1000 Day Window: Health and Mortality Disparities in India

Barun Kumar Mukhopadhyay

Retired Scientist, Population Studies Unit, Indian Statistical Institute Kolkata, India Email: barun mukhopadhyay@yahoo.com

Abstract: Achieving universal access to reproductive health by 2015 is one of the two targets of Goal 5 - Improving Maternal Health - of the eight Millennium Development Goals (UN, 2000). The scientists, economists and health experts of the International Food Policy Research Institute (IFPRI), USA (2012) are of the opinion and agree that improving nutrition during the critical 1,000 day window is one of the best investments to achieve lasting progress in global health and development. Considering the Goal 5 of MDG and objective of IFPRI, an attempt is made in the paper to investigate situation prevailing in India in relation to health of mother, infant and children under 2 years of age. At the outset it must be mentioned that India's concern is much voiced in international arena perhaps because of her overwhelming population of more than 1.2 billion mark (1,210,193,422 in Census, 2011). And India is likely to surpass China in population by 2030 (Rosenberg, 2012). IFPRI (2012) made some comment about India that 1998 to 2005, India's gross domestic product (GDP) grew by 40 per cent, yet the number of children who were stunted (generally defined as being significantly below the median height for their age) declined only from 51 to 44.9 per cent, and those underweight from 42.7 to 40.4 per cent. It is very unlikely that India having boast of the largest democracy and second largest food producing country in the world still are struggling to overcome the curse of ill health of mother and children, in particular, more about gender bias. It is well known that there are countries in sub-saharan regions, in particular and other under developed region in the world are still clutched with antecedent of health hazards, hunger etc. Contrary, India having glimpses of her past and present history in many fields but lag behind health, mortality, literacy, poverty and many others simply because of her heterogeneous character of people which is found less in general in other countries in the world. In this juncture, the paper tries to analyze Indian situation from different sources of data. The findings show predicaments from different major states, in general but boast of having good condition in other smaller states as well. Keywords: MDG, Empowered Action Group, Index of Dissimilarity

1 Introduction

Achieving universal access to reproductive health by 2015 is one of the two targets of *Goal 5 - Improving Maternal Health* - of the eight <u>Millennium Development Goals</u> (UN, 2000). And the scientists, economists and health experts of the International Food Policy Research Institute (IFPRI), USA (2012) are of the opinion and agree that improving nutrition during the critical 1,000 day window is one of the best investments to achieve lasting progress in global health and development. IFPRI remarked Indian green revolution at the same time economic development without development being activated in health of child and mother. Considering the *Goal 5* of MDG and objective of IFPRI, a joint analysis of the two schemes is made in the paper to investigate situation prevailing in India in relation to death of mother, infant and children under 2 years of age for the major states in India.

A perpetual attempt is made in India to reduce the maternal mortality. WHO (Network), 199,000 maternal deaths in South Asia, nearly 74 per cent would be accounted for by India. This amounts to about 140,000 maternal deaths. India is a country with diversity. But this diversity is not the "unity in diversity" rather a much problematic situation by which India suffers. There are states with much development whereas majority of big states without development. As a result overall India is defamed in many country around the world. Because of heterogeneity of the people of India overall estimates of MMR remain as high as around 300 in 2003 and reduced to 212 during 2007- 2009. In so far as the data are concerned, experts in the RG office of the GOI in association with many other scientists realised the grim situation prevailing in India which must be curbed. How long the process of state level stratification will continue without people's stratification. As data of later kind is lacking, the present paper likewise attempts to analyse state wise situation in order to know the different situation prevailing in the different major states in India so that action may be done to improve by inputting different ingredients.



The critical 1,000 day window envisages mother health as well as infants, in particular and children under 2 years of age. The situation prevailing about mother has been explained, to some extent above. But the level of infant mortality rates in India is also very high as compared to neighbouring countries. The differentials in the statuses of infancy period is subjected to their morbidity and mortality at the cost of people's lack of awareness, education, mental set up nutritional food, prophylaxis against nutritional anemia so on and so forth. These are all responsible for in some way or other to have raised the infant deaths in India. Moreover gender disparity right from the embrayo stage at the mother's womb to infancy to adulthood is widely accepted fact as far as India is concerned. It has long been observed that India's population shows an unusually masculine composition. This was generally attributed to higher mortality among females than males in contrast to what has been observed in the western world. However, since the 1990s, the young ages show greater masculinity than in the past and a steep rise in the sex ratio at birth has been observed. There is evidence that sex-selective abortions are widely practiced, facilitated by easy availability of the technology of pre- natal sex-detection and access to medical termination of pregnancy, at least in some parts of the country (Kulkarni, 2007).

2 Sources of Data

i) Special Bulletin on Maternal Mortality in India, 2007-2009, Office of Registrar General, New Delhi-110 066.

ii) Statistical Outline of India, 2007-2008, TATA Services Limited, Department of Economics and Statistics, Bombay House, Mumbai-400 001.

iii) DLHS 3, District level household and facility survey, 2007-2008, International Institute for Population Sciences (IIPS), Mumbai-400 004.

iv) Statistical Abstract of India 2007, Central Statistical Organisation (CSO), Ministry of Statistics and Programme Implementation, New Delhi-110 066.

- v) Census of India for different years
- vi) Sample Registration System for different years and
- vii) NFHS at different time point of Government of India.

3 Methodology

The 1000 Day window is: 273 days of gestation period, 365 days of infancy and again 365 days of children under two years of age above infancy. The first period of gestation signifies the health statuses of mother as well as baby in the womb, the second one the infancy which is quite related to very common vocabulary "infant mortality rate" and finally children under two years of age above infancy. The study is done as par the above criteria. The gestation period has been taken up in terms of differentials of maternal mortality ratios among the states. MMR may give some idea of lack of intake of nutrient food, breast feeding and medicine and so on during ante natal and post natal periods by the population of the different states. Moreover, second category may be studied through infant mortality rates, gender bias among the infants which is very common among people of some of the states. The gender differential is measured through boys/girls ratio of each subject under study. An Index of Dissimilarity (ID) is developed to examine the differentials among population. The formula for the gender differences was originally established by Mukhopadhyay and Majumdar (2012) is as follows:



Index of Dissimilarity =
$$\frac{1}{n} = \sum |\mathbf{R}_{x}|$$

where R_x is the ratio of male by female, x is for different aspect, n the number of observation. As the Institute is of the opinion that if the 1000 days are safe, much of the problem of malnutrition, ill health could be curbed if importance is put on these days of population. What they actually said is important here to quote, "The 1,000 days between a woman's pregnancy and her child's 2nd birthday offer a unique window of opportunity to shape healthier and more prosperous futures. The right nutrition during this 1,000 day window can have a profound impact on a child's ability to grow, learn, and rise out of poverty. It can also shape a society's long-term health, stability and prosperity".

4 Analysis of the data

At the initial stage of the analysis, a table may be given for overall India about the level and trend in MMR in order to cope up of the MDG program of the UN. Later, the analysis may be done for the level of MMR in respect of the major states in India. There are other states (small) and some Union Territories (UTs) which are not considered here. From the major states much of the finding are usually obtained.

5 Overall level and trend of Maternal Mortality Rate in India

The latest figure of MMR is 212 of India during 2007-2009 (RG, 2011). To explain a few words (Table 1 given below) from the declining values of maternal mortality ratios from 398 in 1997-1998 to 301 in 2001-2003, i.e., during the period of seven years is of the order of around 25 per cent. An earlier study

Year	Maternal mortality ratios	Per cent changes in MMR
1997-'98	398	-
1999 - '01	327	-17.84
2001 – °03	301	-8.00

Table 1 Maternal mortality ratios andper cent change in India , 1997-2003

estimated a declining trend from around 1355 maternal deaths per every 100 thousand live births during 1957-'60 to about 330 maternal deaths, in the year 2002 (Ranjan and Gulati, 2008). In NFHS studies maternal mortality rates had not decline during nineties, rather increased from 424 in 1992-'93 (NFHS-1) (95 per cent CI: 324-524) to 540 in 1998-'99 (NFHS 2) (95 per cent CI: 428- 653). The large confidence intervals were due to smaller number of maternal deaths in these surveys. In this respect a downward linear trend was found during 1966 to 1992 in rural areas of India (Mukhopadhyay, 1996) from estimated figures of 676 in 1966 to 352 in 1992. The linear trend was tested through best fit of the data (p<0.00).





Series 1: Absolute values; Series 2: Percentage change

Fig.1. MMR values with trend

6 State MMRs and Trends

8-

Table 2 Maternal mortality ratios of the major states in India and rate of change, 1997-'03 03	

Major States	Maternal Mor	rtality Ratio		Rate of Change (percent)			
	1997-98	1999-2001	2001-2003	(1997-98)-(1999-2001)	(1999-2001)-(2001-2003)		
Andhra Pradesh	197	220	195	-11.68	-11.36		
Assam	568	398	490	-29.93	+23.12		
Bihar/Jharkhand	531	400	371	-24.67	-7.25		
Gujarat	46	202	172	-	-14.85		
Haryana	136	176	162	+29.41	-7.95		
Karnataka	245	266	228	+8.57	-14.29		
Kerala	150	149	110	-0.67	-26.17		
Madhya Pd/Chhattisgarh	441	407	379	-7.71	-6.87		
Maharashtra	166	169	149	+1.81	-11.83		
Orissa	346	424	358	+24.54	-15.57		
Punjab	280	177	178	-36.79	+0.56		
Rajasthan	508	501	445	-1.38	-11.18		
Tamil Nadu	131	167	134	+27.48	-19.76		
UttarPd./Uttaranchal	606	539	517	-11.06	-4.08		
West Bengal	303	218	194	-28.05	-11.01		
India	398	327	301	-17.84	-7.95		



A perusal of the above table reveals a large extent of differences in the level of MMR among the states in 1997-1998 at the initial point of time, i.e., 1997-1998. States like Tamil Nadu ,Haryana and Kerala depicted lower level of MMR being respectively 131, 136 and 150 in 1997-'98. On the other hand, states like Uttar Pradesh/Uttaranchal, Assam, Bihar/Jharkhand and Rajasthan possessed the level at much higher values at the rate higher than 500 mark. Madhya Pradesh/Chhattisgarh depicts slight lower level of 441 which is itself still much high. The levels of the remaining states lie in between these two categories of states. The figure for Haryana being 46 is a doubtful one. Here the contribution of the states in their number of population in overall India seems very high for Uttar Pradesh/Uttaranchal, Bihar/Jharkhand, Rajasthan (Fig.1 as shown in the **Appendix**).

Now the trend of the values during the seven year periods into two periods of 1997-2001 and 2001-2003 clearly show some trend of declining values except in some states where upward trend was observed. As the period of study is very short, hence no definite conclusion can be made so far as trend analysis is concerned. Due to paucity of data, that cannot be done in a proper manner. However, apart from all predicaments, it may be said that a trend of decline at one hand is observed among all the states and on the other the variations are found among the states. As such any developmental programme, if it is done this must go through the in depth analysis before actual application. Otherwise only on the basis of overall Indian values, it would be a waste either of energy or of money etc., as far as national policy is concerned.

After analyzing the data on MMR with respect to MDG, it is now necessary to analyse the data of infant mortality rates at an initial stage giving parity with the 1000 Day Window.

7 Infant Mortality Rates in India with Trend and Gender Bias

The male-female infant mortality rates from 1985 to 2000 are given in Table 3. The values of IMF are declining in both the cases of sexes more or less over the entire period from 1985 to 2000. The CIA World Fact book (Network system, 2011) shows, male: 46.18 deaths/1,000 live births and female: 49.14 deaths/1,000 live births in India, 2011. From these figures the value of DC in 2011 is found to be 0.940 which is still not favouring females. From the table the ratios are found higher than unity in some stray cases where female mortality is less, otherwise in most years the values indicate higher infant mortality rates among females than males

8 Infant Mortality Rates in the States

The table given below shows the level of infant mortality rates of the 15 major states of India. The differentials in the statuses of infancy period is subjected to their morbidity and mortality at the cost of people's lack of awareness, education, mental set up nutritional food, prophylaxis against nutritional anemia so on and so forth. These are all responsible for in some way or other to have raised the infant deaths for which the values of which are put in the table given below for those states in India in 2001. Moreover, these differentials are expected to be similar still after one decade, that is in 2012. Again, the state, Kerala possessed the minimum value of only 11 infant deaths meaning thereby good health of infants, to a great extent, with less malnutrition, good ante natal and post natal cares to have been taken and many other measures rightly and in time seriously have been adopted by the people with their advanced mental set up and awareness, in particular and state as well. The children in that state cannot be said to have been stunted as being significantly below the median height for their age and underweight. Odisha's figure is of the order of 90 may be taken as an exceptional level. Whereas big states like Madhya Pradesh (86), Uttar Pradesh (82), Rajasthan (79) and Assam (73) have very high infant mortality rates, the other states like Andhra Pradesh (66), Haryana (65), Bihar (62) possess moderately high values than the states, Maharashtra (45), Tamil Nadu (49), West Bengal and Punjab (51 each) and Karnataka (58). A picture is given below about these values of these states and India as well.



Major states	IMR	Major states	IMR	Major states	IMR	Major states	IMR
Andhra Pd.	66	Haryana	65	Maharashtra	45	Tamil Nadu	49
Assam	73	Karnataka	58	Odisha	90	Uttar Pradesh	82
Bihar	62	Kerala	11	Punjab	51	West Bengal	51
Gujarat	60	Madhya Pd.	86	Rajasthan	79		

 Table 3 Statuses of infant mortality rates in the major states, 2001

From the differential scenario of data and figure, there must have been some diversities in the states in India which might have played a great role on the overall level of IMF. However, this picture gives the impression of variations in the values of infant mortality rates among the major states in India.



9 Different 'over sex-ratio' at birth for overall India during 1982-2005

An attempt is made in the present paragraph to know overall values of sex-ratio at birth (Male births per 100 female births) in India. This paragraph is given here to understand the level of sex-ratio at birth for India never achieved the international standard of either 105 or 106 or 107. Scientists very commonly sometimes use sex-ratio at birth to be 105 while estimating demographic parameters without verifying the exact figure at the time of their research.



Year		1982-'84	1983-'85	1984-'86	1985-'87	1986-'88	1988-'90	1990-92
Sex-ratio at birth		109.8	110.4	109.5	109.6	109.8	109.8	111.1
1991-93	1992-94	1993-95	1994-96	1995-97	1996-98	1998-'00	1999-'01	2000-'02
111.9	113.0	113.8	113.3	112.2	111.0	111.4	111.9	112.1
1991-93	2001-03	2002-04	2003-05				- -	- -
111.9	113.3	113.4	113.6					

Table 4 Sex-ratio at birth at different time point during 1982-2005, India

The table 5 above shows how the values are too high at different time point. From these high figures one may find some distinct difference between the past and the present values. As for example, during 1982 to 1990 the figures are slightly less values around 110 on an average but after that they started increasing from 111 to 114 during 1991 to 2005. It simply may show more and more female feticide and infanticide have occurred over time.

10 The third category study

Now coming to the third category of population age one, that is those below the age of 2 years as prescribed by IFPRI, the analysis is done on the basis of data on deaths below the age of 5 among the states in rural and urban areas in India in 2002. From these figures the condition of under 2 age population has been attempted here.

Table 5 Per cent under five death rate by sex and residence, 2002

Male 25.7	Female	Male	Female
25.7	20.2		
	30.3	16.7	13.5
23.8	28.0	18.2	15.7
19.4	25.0	9.9	9.8
22.7	24.9	9.0	10.9
22.9	23.9	13.3	18.4
27.8	23.2	12.1	18.0
16.0	22.2	10.2	14.4
11.9	21.5	8.0	13.4
	23.8 19.4 22.7 22.9 27.8 16.0 11.9	23.8 28.0 19.4 25.0 22.7 24.9 22.9 23.9 27.8 23.2 16.0 22.2 11.9 21.5	23.8 23.0 18.2 19.4 25.0 9.9 22.7 24.9 9.0 22.9 23.9 13.3 27.8 23.2 12.1 16.0 22.2 10.2 11.9 21.5 8.0



Contd. from earlier page		I	I	Contd. in next page
Bihar	15.6	19.7	14.1	12.9
Andhra Pradesh	17.1	18.5	10.0	5.9
Karnataka	18.3	17.4	6.9	6.6
West Bengal	13.7	12.9	10.2	6.0
Maharashtra	11.9	11.3	8.5	7.9
Tamil Nadu	12.8	11.3	7.9	6.8
Kerala	1.8	2.8	2.5	1.4

The table given above depicts the deaths of children under age five of the major states in India in 2002. The bottom line figures are for Kerala are the lowest in rural and urban areas of the order of 1.8 for male in rural areas and 2.8 for female in the same areas. The similar low figures are in urban areas as well. There are other states like Tamil Nadu, Maharashtra and West Bengal possessing low average figures around 10 in all the cases. The position of other major states are comparatively higher values as shown in the table. As far as gender bias of deaths below the age of 5 among the states in rural and urban areas in India in 2002 there are gaps. The figures overall show less gender bias among mostly in Kerala and to some extent in other developed states like Tamil Nadu, Maharashtra, West Bengal both in rural and urban areas. In Punjab although rural figure gives lower value (11.9) but urban figure shows almost double. In undeveloped states the figures are much higher namely Madhya Pradesh, Uttar Pradesh, Rajasthan, Assam and Odisha. The figures for Andhra Pradesh, Karnataka, Bihar, Haryana are lying within these two categories.

Finally a comparison of India with South Asian countries in respect of MMR and some health indicators of children under 5 years of age has been done using the following table.

11 Comparison of health status among the South Asian Countries

Three indicators namely maternal mortality rate, per cent children under 5 severely underweight and per cent children under 5 moderately underweight have been considered for the analysis among the countries including **India**, **Sri Lanka**, **Pakistan**, **Maldives**, **B' Desh**, **Nepal**, **Bhutan**, **Myanmar and Afghanistan**. The table given below gives the diversified scenario among the nine countries except Afghanistan about which incomplete data are available.
South Asean	Health Indicators						
Countries	MMR	%Children<5 severely underweight	% Children<5 moderately or severely underweight				
India	200*	15.8**	43.5**				
Sri Lanka	35.0*	3.7 ^d	21.6 ^e				
Pakistan	260*	12.7 ^a	31.3 ^b				
Maldives	60*	3.3 [°]	17.8 [°]				
Nepal	170*	10.6**	38.8**				
Bangladesh	240*	11.8 ^d	41.3 ^d				
Bhutan	180*	3.2*	12.7*				
Myanmar	200*	5.6*	29.6 ^e				
Afghanistan	460*	na	na				
*2010,a>2002,b>20 **2006	001, c>2009, d>2007,	e>2003	1				

Table 6 Health indicators of South Asian Countries

Three indicators namely maternal mortality rate, per cent children under 5 severely underweight and per cent children under 5 moderately underweight have been shown in the table above. The Indian position is worth noticeable. While MMR is concerned majority of the states have reduced their rates except Afghanistan, Pakistan the rates are high and to some extent Bangladesh too. India's position is mediocre. Sri Lanka and Maldives deserve high ranking position. In terms of the figures for under 5 children's health condition in both cases are overwhelmingly high in India. From above figures it is clear that India is to traverse more paths in order to achieve the goal of MDG and Window of 1000 Day.

12 The Root Cause of Overall Situation

There are several reasons for the downfall of India's position in respect of health of mother and child. The most vulnerable point to be noted here that with many sensitive people of India being perpetually defamed in international arena in its many aspects including the present one at the cost of a big chunk of people with their many social taboos, prejudices and backwardness in their mind and thoughts. The root cause of these deficiencies in Indian people cannot be opened before the general masses who are still in darkness that why India's position is so worse when at the same time state like Kerala and some other states in south as well as in west, north or even in east boast of having very good position as compared to major big states contributing overwhelmingly the overall Indian status for which a section of good states suffer. These are all culminating from the very nature of India with much heterogeneities in its health consciousness of their mothers and children (Figures II, III, IV and V as shown in the **Appendix I**) apart from their culture, language, social systems, attitudes towards social responsibilities and so on. In

this diversified nature of the people of India an example of use of different languages as their mother tongue in different states is quoted below in table 8. There are other heterogeneities as well. Ironically the age at marriage in

India, especially in rural areas is still very low. It is, to some extent, also low even among educated youths enjoying urbanities with high fertilities due to non practice of FP methods in their early marital life if investigated.

Language	Speakers	State(s)
0 0	(in millions, 2001)	
Assamese	13	Assam, Arunachal Pradesh
Bengali	83	West Bengal, Tripura, Andaman & Nicobar Islands
Gujarati	46	Dadra and Nagar Haveli, Daman and Diu, Gujarat
Hindi		Andaman and Nicobar Islands, Arunachal Pradesh, Bihar, Chandigarh,
	258 122	Chhattisgarh, the national capital territory of Delhi, Haryana, Himachal
	230-422	Pradesh, Jharkhand, Madhya Pradesh, Rajasthan, Uttar Pradesh and
		Uttarakhand
Kannada	38	Karnataka
Kashmiri	5.5	Jammu and Kashmir
Konkani	2.5-7.6	Goa, Karnataka, Maharashtra, Kerala
Malayalam	33	Kerala, Andaman and Nicobar Islands, Lakshadweep, Puducherry
Marathi	72	Maharashtra, Goa, Dadra & Nagar Haveli, Daman and Diu, Madhya
		Pradesh
Oriya	32	Odisha
Punjabi	29	Chandigarh, Delhi, Haryana, Punjab
Tamil	61	Tamil Nadu, Andaman & Nicobar Islands, Puducherry
Telugu	74	Andaman & Nicobar Islands, Andhra Pradesh, Puducherry
Urdu	52	Jammu and Kashmir, Andhra Pradesh, Delhi, Bihar, Uttar Pradesh and
Ordu	52	Uttarakhand

Table 7 People in India with their languages spoken in millions in different states, 2008

From the above table it may be pointed out that Hindi is the language spoken by most of the major states in India namely Bihar, Chandigarh, Chhattisgarh, the national capital territory of Delhi, Haryana, Himachal Pradesh, Jharkhand, Madhya Pradesh, Rajasthan, Uttar Pradesh and Uttarakhand. No doubt this language is the national language, but unfortunately these states except some smaller ones also contribute maximum in the Indian population with highest MMR, IMF and many other social, demographic characteristics. As a result Indian position could not be remedied unless these people improve their status. No doubt there are other states as well where improvement is also is to be done, not to great extent. Moreover it is also to be mentioned here that India as a democratic country. People can have the right to live in their own choice of states sometimes for their livelihood and other purposes. If detailed analysis is done taking separately each state, it may be found cross language system prevails, which means People from other states migrated in West Bengal. Similarly in other states also. Heterogeneities within the state are not because of entire aborigines but a large amount of migrated people in the state as well. The culture, language, lifestyle are different among the people which means that geographic diversities are not the only ones. But it is the people's diversity above all. In this connection government usually adopt different strategies and targets in order to curb the problem, - Population Policy (NPP), 2000, the National Health Policy (NHP), 2002, the Reproductive and Child Health (RCH) Programme (initiated since 1996) and the National Rural Health Mission (NRHM) 2005-2012. Between 4 and 5 million women in India suffer from ill-health associated with childbearing India's commitment to the eight Millennium Development Goals (MDG) set by the United Nations, among which is the goal of reduction in maternal mortality ratio (MMR) by three-fourths by 2015, will therefore require the necessary political will in the implementation of policy and programmes to safeguard maternal health (Das et al, 2007).

Time has come now to undertake chemotherapy-like treatment to kill the dead cell from the body in each step so that all the bad prejudices, dogmas, unawareness about the situation prevailed among people irrespective of education, social economic status residing in India spreading over different regions, states, districts, blocks and villages are eradicated. But that is a huge task, alternatively GOI have classified the states into three categories. The first category comprises the "Empowered Action Group" (EAG) states of Bihar and Jharkhand, Madhya Pradesh and



Chhattisgarh, Odisha, Rajasthan, Uttar Pradesh and Uttaranchal. Assam is also included in this list. Nearly two-third of the maternal deaths in the country are reported to occur in the EAG states and Assam. Among the three categories, this category accounts for nearly 47 per cent of the births, no doubt with stunted births as well. Numbers of children who were stunted (generally defined as being significantly below the median height for their age) declined only from 51 to 44.9 percent, and those underweight from 42.7 to 40.4 percent during 1998 to 2005 (IFPRA, 2012). Moreover, these states have had, historically higher child mortality, higher poverty levels and lower life-expectancy and other indicators than other states. The second category (Region) covers the "Southern" states of Andhra Pradesh, Karnataka, Kerala and Tamil Nadu. These states traditionally have had better child mortality and other health indicators. The remaining major states formed the third category (Region) and have been classified as 'Others' (Office of the RG in India, 2006) including Maharashtra, Punjab, West Bengal and some other with their mediocre status.

In this respect it is envisaged that although MMR declined from about 400 maternal mortality rate in 1997-'98 to about 300 in 2001-'03 and 212 during 2007-2009 gives satisfaction but at the same time tells that reducing MMR to 195 by 2012 of National Commission of Population (NCP) and National Rural Health Mission (NRHM) and to 109 by 2015 of the Millennium Development Goals (MDG) is going to be a real challenge. Particularly when most of the deaths occurred in the states included in the "Empowered Action Group" (EAG) of states an Assam. For further decline, rapid progress in health sector schemes would be needed in these states, in particular. And, these states are thus focus of the National Rural Health Mission (NRHM) as are quoted elsewhere (Government of India,2006). But Government of India with joint meeting of the different international bodies like UN, UNICEF in its report mentioned in 2010 the target is to achieve at least 135 not 109 keeping in view of the present position.

13 A Proposed Model on Demographic aspect

If a model of demographic identity system is thought to be initiated in India so that every individual will be directly connected with the government machinery so that social responsibility of the people in India would be more proactive for the benefit of the nation building. Moreover any new item occurs in respect of any individual that must either be reported to the concerned department so that the change will have to be included in the corresponding records maintained in the computer network system. It will be easy to follow up each individual on regular basis after certain specified interval of time. Moreover each individual will be liable to the prescribed schedule introduced time to time by the GOI. The National Population Register (NPR) though initiated by the Registrar General and Census Commissioner of India before another new system in the form of "AADHAAR" (Unique Identification Authority of India, 2011) to reach each individual through e-governance as and when necessary for the real benefit of mankind but the demographic data in the NPR are not complete as it seems from the website (see Appendix II). Before this GOI have already introduced Epic card also with the photo of the respective voter in order to get rid of malpractices in the election process. Then it can also be done for detailed demographic identity of the population may it be included in the NPR or a separate card may be introduced as par present model .

14 Conclusion

MDG (UN, 2000),- when developed countries registered substantial reduction in maternal morbidity and mortality, developing countries in particular constituting ninety nine per cent of the total maternal deaths which are occurring throughout the world. India is still lagging behind the target of 109 by the MDG later changed to 135 in 2015 by the GOI with the consultation of UN. International Food Policy Research Institute (IFPRI), USA (2012) recently prescribed 1000 Day Window improving nutrition during the 1000 day of mother and child. IFPRI makes some comment about India, in particular that in spite of India's green revolution and economic development is lagging behind health especially of child and mother and other. India with 2nd largest food producing country with highest democracy in the world, her overall estimates of MMR remain as high as around 212 during 2007-2009 only because of India having much heterogeneities in her culture, language, social values, etc apart from general consensus of geographical diversities. As such, GOI classified the states into different categories as mentioned in earlier paragraph in order to curb the situation by giving emphasis on respective group of states. And the present study also pointed out diversities in different aspect of health of mother and child vis-à-vis diversities among states. From the heterogeneity point of view of Indian people, keeping aside the usual attempt on states, districts, region



basis a new model on individual basis study has been proposed in the present attempt in the form of introducing identity card system of Indians in order to capture their socio economic, demographic statuses for effective planning.

15 References

- 1. Bread for the World Institute, 2012. International Food Policy Research Institute, IFPRI, USA.
- Das N.P., Shah, U and Patel, R. 2007. Emerging Causes and Determinants of Matemal Mortality in India: Based on Large Scale Surveys since the 1990s.
- Dasgupta, Monika, 1987. Selective Discrimination against Female Children in Rural Punjab, India, Population and Development Reviews, 13, The Population Council, USA.
- Govt. of India, 2006. Maternal mortality in India, 1997-2003, Trends, causes and risk factors, Office of the Registrar General, New Delhi, pp 1-18.
- 5. Govt. of India, 2006. AADHAAR, Unique Identification Authority of India, Bengaluru.
- Mukhopadhyay, B.K. and Majumdar P.K. (2012). Status of Gender-Differentials and Trends in India : Population, Health, Education and Employment in Pal, M ; Pathak, P ; Bharati, P ; Ghosh, M; Gender Issues and Empowerment of Women, Nova Science Publishers, INC, New York.
- Mukhopadhyay Barun Kumar, 1996. Trends and developments in maternal mortality vis-a-vis a differential health socio-economic infrastructural transformations among the major states in India, Socio-Economic Transformation in India, Ray, B (ed.), Kanishka Publishers, Distributors, New Delhi, pp 79-110.
- 8. Network 8 ystem, 2011. The CIA World Fact book.
- P.M. Kulkarni, 2007. Estimation of Missing Girls at Birth and Juvenile Ages in India, A paper prepared for the United Nations Population Fund (UNFPA), India.
- Ranjan, A. Gulati, S.C., 2008. INDIA, the state of population, 2007, Oxford University Press, New Delhi, pp1-215.
- 11. RG, 2011. Special Bulletin on Maternal Mortality in India 2007-09, SRS, New Delhi.



Appendix I

Pictures shown below are differentials among state proportions

Figure I: Proportions of populations in states and UTIs, Census 2001





Contd. Appendix I

Figure II: Ante natal check up of expectant mothers Figure III: Institutional delivery, DLHS, 2007-09



Figure IV: Full vaccination of children (12-23 months) Figure V: Current use of FP methods, both, DLHS, 2007-09





APPENDIX II

NATIONAL POPULATION REGISTER (NPR)

The National Population Register (NPR) is a Register of usual residents of the country. It is being prepared at the local (Village level), sub District (Tehsil/Taluk level), District, State and National level under provisions of the Citizenship Act 1955 and the Citizenship (Registration of Citizens and issue of National Identity Cards) Rules, 2003.

The NPR would have the following details of every usual resident in the country (i) Name of Person, (ii) Relationship to head, (iii) Father's name, (iv) Mother's name (v) Spouse's name (if married), (vi) Sex, (vii) Date of Birth, (viii) Marital Status, (ix) Place of Birth, (x) Nationality as declared, (xi) Present address of usual residence, (xii) Duration of stay at present address, (xiii) Permanent residential address, (xiv) Occupation/ Activity, (xv) Educational qualification Also, three biometrics namely photograph, 10 finger prints and 2 iris would be collected for persons of age 5 years and above.

A proposal to issue Resident Identity Cards to all usual residents in the NPR of 18 years of age and above is under consideration of the Government. This Card would be a smart Card and would bear the Aadhaar number.

The NPR is a register of usual residents. The data collected in NPR will be sent to UIDAI for de-duplication and issue of Aadhaar Number. Thus the register will contain three elements of data - (i) demographic data, (ii) biometric data and (iii) the Aadhaar (UID Number).

A proposal to issue Resident Identity Cards to all usual residents in the NPR of 18 years of age and above is under consideration of the Government. This Card would be a smart Card and would bear the Aadhaar number.





Reach and Utilization of Maternal and Child Health Services in India

P. Murugesan Research Officer and Biostatistician Department of Fertility Studies International Institute for Population Sciences (IIPS), Deonar, Mumbai-India 400 088 E-mail: <u>aasai muru@yahoo.com</u>

Summery

Maternal and Child Health has remained an integral part of the Family Welfare Programmes in India. The National Family Health Survey-3 (2005-06) provides enormous data related to Maternal and Child Health care- antenatal, delivery and post natal care for the women while for child care variables like child immunization-use of BCG, DPT 1 to 3 and Polio 1-3 etc. are available. The present paper therefore aims to provide the prevailing situation regarding reach and utilization of the MCH services in India as well as among the different states. It was observed that in India around 77 percent women received antenatal care for their most resent birth doing the five years residing the survey. Interestingly the rate of increase was higher in rural area as compared with urban areas. State-wise variations were noticed with regard reach and utilization of antenatal care services. Further, highest cases of safe delivery assisted by health personnel were found for the state of Tamil Nadu (91 percent) and lowest for Nagaland (12 percent). With regard to child care services percentage values obtained for fully vaccinated children range from 81 percent from Tamil Nadu and 21 percent for Nagaland. Besides providing state wise comparative picture with regard to MCH services the paper intends do discus factors affecting utilization of MCH services among the regions of the country by using logistic regression analysis. In addition attempt would be made to discus some of the plausible reasons for the variation in the utilization of MCH services.

Keywords: Reach and Utilization of Maternal and Child Health Services

Introduction

In India, the striking interregional diversity is an important confounding factor. As such, the current research analyzes data for India as a whole and for the states of Bihar, Madhya Pradesh, Rajasthan, and Uttar Pradesh in the north and Andhra Pradesh, Karnataka, Kerala, and Tamil Nadu in the south. These two groups of states are distinctly different socio-economically and culturally and are fairly representative of the north-south dichotomy observed by Dyson and Moore (1983). Southern women typically enjoy greater freedom, an outcome of the Dravidian culture, and higher levels of literacy, education, and employment. Northern women are strongly subjected to the traditional conservatism of the Mogul legacy and are predominantly illiterate, less educated, and less likely to work outside the home.



This research sets out to investigate the broad nature of the association between utilization of maternal and child health (MCH) services in India. Using data from the 2005-06 National Family Health Survey (NFHS-3), we examine the relationship between maternal schooling and factors known to reduce the risks of maternal and child mortality—utilization of antenatal and delivery care services, utilization of child immunization services, and treatment of childhood diseases. We hypothesize that the practices of educated mothers, with regard to pregnancy, childbirth, immunization, and management of childhood diseases such as diarrhea and acute respiratory infection (ARI), are quite different from those of their uneducated counterparts. In addition, the extent to which the impact of maternal schooling on the utilization of MCH services is confounded by other socioeconomic and demographic influences is examined.

Cross-country comparisons using large data sets, such as the World Fertility Survey and the Demographic and Health Surveys, have shown that education in general and female education in particular exert a very strong influence in reducing child morbidity and mortality (Boerma et al. 1990; Bicego and Boerma 1993; Caldwell and Caldwell 1990; Hobcraft, McDonald, and Rutstein 1985; Murthi, Guio, and Dreze 1995). At the micro level, more in-depth quantitative and qualitative research that examines women's health-enhancing behavior has arrived at similar conclusions (Bhuiya and Streatfield 1991; Bourne and Walker 1991).

In investigating the pathways of influence, research confirms that the causal linkages between these two factors are far from clear and that this relationship is simply not a reflection of a co-occurrence of education with other socioeconomic variables (Desai 1994; Hobcraft 1993). In spite of methodological problems associated with the measurement of maternal mortality and morbidity, several studies have shown a strong relationship between maternal mortality and morbidity and the absence of prenatal care. They have also shown that and that utilization of prenatal care is dependent on, among other factors, maternal education (Monteith et al. 1987; Okafor 1991; Wong et al. 1987). Studies of factors that influence the utilization of modern delivery-care services are, however, scarce. Several other studies have been carried out to explain how maternal education may influence child health, mostly within the conceptual framework put forward by Mosley and Chen (1984), who argue that mortality is the outcome of a combination of social, economic, biological, and environmental factors and that these factors operate through a set of proximate determinants.

Three broad pathways of influence, linking maternal schooling to child mortality, that result in greater utilization of modern health services have been suggested: educated women are better able to break away from tradition to utilize modern means of safeguarding their own health and that of their children (Caldwell and Caldwell 1988; Cleland 1990); educated women are better able to utilize what is available in the community to their advantage (Barrera 1990;



Caldwell 1990; Goodburn, Ebrahim, and Senapati 1990); and educated women may be able to make independent decisions regarding their own and their children's health leading to greater utilization of modern health facilities (Caldwell 1979; Caldwell 1986).

Objectives

The objective t of the present study is aims to provide the prevailing situation regarding reach and utilization of the maternal and child health services in India. And the specific objectives are as follows:

- 1. To study the reach and utilization of maternal health services in India and its major states.
- 2. To study the reach and utilization of child health services in India and its major states and
- 3. To examine the factors affecting the maternal and child health services in India and its states.

Data and Methodology

The NFHS-3 gathered during 2005-06 information on various aspects of maternal-care utilization has been used in India as well as the major states. Specifically, for each live birth in the five years preceding the survey, a woman was asked if she had received antenatal care. If she did, she was asked who administered the care, how many months pregnant she was when she first received antenatal care, how many antenatal care visits she had in all, whether she had received an injection to prevent tetanus during her pregnancy and, if so, how many injections she had received, whether she had received iron/folic-acid tablets while she was pregnant, where she gave birth, and who assisted with the delivery.

The base sample for our study of maternal-care utilization constitutes 49,369 children born in the period 1–42 months prior to the survey. These were children born to ever-married women aged 15–49 for whom information on health was obtained. For child health-care utilization, only the 45,363 children who were alive at the time of the interview are considered. In this analysis, we include five main indicators of maternal-care utilization—antenatal care, tetanus toxoid immunization, receipt of iron/folic-acid tablets, place of delivery, and delivery care. The child health-care indicators of interest are the percentage of children born in the four years preceding the survey who suffered



from ARI in the two weeks preceding the date of interview and were taken to a health facility for treatment, the percentage of children who suffered from diarrhoea in the two weeks prior to the interview and were taken to a health facility for treatment or were treated with ORS and/or RHF, and the percentage of children age 12–23 months who were fully vaccinated.

Information on maternal and child health care is birth based, that is, it is collected for each child born to respondents in the four years preceding the interview. The denominator differs for each dependent variable. For maternal health-care indicators, the denominator is all live births in the four years preceding the survey. For ARI, diarrhoea, and treatment with ORS and/or RHF, it is the number of children who were actually reported sick with the illness in the two weeks before the interview. There are five dichotomous dependent variables for maternal health-care indicators: whether or not mothers received at least two doses of tetanus toxoid vaccine; whether or not mothers received iron/folic-acid tablets; whether or not mothers received antenatal care from a health professional in an institutional setting, from a health worker at home, or from any other person whether or not mothers received delivery care from a trained health professional; and whether or not the birth was delivered in a public or private medical institution. Trained health professionals include doctors, nurses, trained nurse/midwives, and other formally trained personnel. The dichotomous dependent variables for child health-care indicators are: whether or not sick children suffering from symptoms of ARI and/or diarrhoea were taken to a health facility or health provider for treatment; whether children sick with diarrhoea were treated with ORS and/or RHF; and whether children age 12–23 months at the time of interview are fully vaccinated.

A health facility or provider includes government or municipal hospitals, primary health centers, subcenters, mobile clinics, village health guides, and government paramedics in the public sector and private hospitals or clinics, private doctors, mobile clinics, and community health workers in the private sector. A child is considered as having been taken to a health facility if he/she received care from one or more of the above sources. Children are considered fully vaccinated if they received the tuberculosis (BCG) and measles vaccinations and all three doses of the DPT (diphtheria, pertussis, tetanus) and polio vaccines. The dichotomous dependent variable in this case is children fully vaccinated or not. The age group 12–23 months is recommended by the World Health Organisation (WHO) as the age group to examine because it covers children older than 11 months (by which age they should have been fully vaccinated) and because it contains the most recent information (older children may not have been covered by more recent health programmes). Even though the sample is not restricted to one child per woman, it is unlikely to over represent women with higher fertility (who are more likely to be rural and less educated) because there are very few women who had more than one birth in the four years before the survey.



The dependant variables are maternal health care services and child health care services. Several independent variables, both socioeconomic and demographic, which could potentially influence the outcome were included in the analysis: mother's education, place of residence, employment status, caste, and religion and child's sex, birth order, and age (applicable to child health-care indicators only) at the time of interview. Mother's age at the time of birth of her child was not included since it is highly correlated with birth order. Mother's education was categorized into three groups: illiterate, literate less than middle school complete and middle school complete and above. Place of residence was urban or rural. Mother's employment status was categorized into working and not working outside the home. Even though women working for cash may be in a particularly strong position, socio-economically and otherwise, because of the small numbers working outside the home (and even smaller numbers working for cash), it made sense statistically to group women into just these two categories. Caste was divided into two groups: those belonging to scheduled castes and scheduled tribes and those not belonging to such groups. Religion was classified into Hindu, Muslim, and others. Birth order was broken down into three categories: 1, 2–3, and 4+; and child's age in months was grouped into five categories: less than 6 months, 6–11, 12–23, 24–35, and 36–47 months.

A logistic regression is run for each of the two dependent dichotomous variables. This technique examines the potential strength of education when the influence of other socioeconomic and demographic variables is controlled. The model is based on forward stepwise selection using the Wald statistic for deleting variables that exceed the 0.1 cutoff values. The same models are applied at the national level only.

Result and Discussions

The present paper made an attempt to discus two components i.e. Maternal and Child health care services. These two components are analyzed one by one using NFHS-3 all India data and the same is provided below.

Maternal Health Care Services

The Table.1 shows India as well as states differentials in five maternal care services for birth during the five years preceding the survey. These indicators together summarize the extent to which different states have progressed toward achieving safe motherhood goals at all three stages of the birth process: antenatal, delivery and postnatal. The first indicator is a summary antenatal care for the last live birth in the five years preceding the survey, mother received three or more antenatal check-ups(with the first check-up within the first trimester of pregnancy), received two or more tetanus toxoid injections, and took iron and folic acid tablets or syrup for three or more moths. The next two indictors pertain to care during delivery and show the percentage of births



delivered in medical institutions and deliveries assisted by health personnel. The last two indicators pertain to postnatal car for mothers and show the percentage of deliveries with a postpartum check-up within 42 days of the birth and within two days of birth.

For India as a whole, mothers of only 15 percent of births received all of the required components of antenatal care services. This indicator ranges from a high of 64 percent in Kerala and 56 percent in Goa to a low only 2 percent in Nagaland and 4 percent in Uttar Pradesh. Other states that performs almost all as poorly as Uttar Pradesh and Nagaland on this indicator include Bihar, Arunachal Pradesh, Madhya Pradesh, Jharkhand, Meghalaya, Rajasthan and Mizoram, where only 6-9 percent of women received the recommended components of antenatal care. Kerala , followed closely by Goa, also outperform all other states in terms of delivery care, with nearly all deliveries taking place in medical institutions and a similarly high percentage if deliveries assisted by a health professional. Tamil Nadu, with 88 percent of births delivered in medical institutions in Nagaland, Chhattisgarh, and Bihar. Only 25-29 percent of deliveries are assisted by health are assisted by health professionals in Nagaland, Uttar Pradesh, Jharkhand and Bihar.

Tamil Nadu, where 91 percent of deliveries have a postnatal check-up within 42 days of birth and 87 percent within two days, tops the list of states with regard to both of the postnatal care services. Kerala and Goa are the other two states which perform quite well on the postnatal care services.

An examination of the performance of each state on the different safe motherhood indicators shows that several states consistently perform well below the national average one each of the five services. This list includes Rajasthan in the North Region, all states in the Central Region, Bihar and Jharkhand in the East Region, and Arunachal Pradesh, Assam, Meghalaya, and Nagaland in the North East Region. Uttarachal also performs poorly on all the indicators except antenatal care, which is slightly higher than the national average. In contrast, Mizoram performs above the national average on the delivery care service and the postnatal care services, but poorly on the antenatal care service.

In case of the multivariate analysis of logistic regression models the dependent variable is considered as maternal health care services. The maternal health care services are all three types of services are included (antenatal, delivery and postnatal cares). And the independent variables were considered is mothers socio-economic and demographic background variables. The logistic models were carried out only for India as a whole and are presented in the Table.2 providing Adds Ratio or Expected B values of the dependent variables with the mother's socio-economic predictors. It observes that compared with all the background variables the mother's education has high significance in all the three factors of maternal health care services. Another very important finding was that mother's religions have significance (99 percent level of significance) with assistance during delivery and postnatal care services.



preceding the survey by State.									
India /Major States	Percentage who received all recommended types of ANC care	Percentage of births delivered in a health facility	Percentage of deliveries assisted by health personnel	Percentage of deliveries with a postnatal check-up	Percentage of deliveries with a postnatal check- up within two days of birth	Total Number of Women			
India	15.0	38.7	46.6	41.2	37.3	29995			
North									
Delhi	29.0	58.9	64 1	60.9	58.4	797			
Haryana	14 7	35.7	48.9	57.6	55.9	783			
Himachal Pradesh	17.4	43.0	47.8	50.6	43.2	668			
Jammu & Kashmir	17.5	50.2	56.5	51.6	19.2	7/8			
Punjab	10.8	51.3	50.5 68 2	63 7	62.0	837			
Raiasthan	17.0	20.6	08.2	21.9	02.0	0 <i>3</i> 7 1042			
Uttaranchal	0.0	29.0	41.0	51.0 25.0	20.9	1045 644			
Cartal	10.1	32.0	38.3	33.8	32.4	044			
Chattiagorh	11.2	14.2	41 6	262	20.4	1040			
Madhya Pradash	11.3	14.3	41.6	36.3	28.4	1040			
Uttar Pradesh	7.2	26.2	32.7	33.8	28.5	1753			
	4.1	20.6	27.2	14.9	13.3	3196			
East									
Bihar	5.8	19.9	29.3	17.8	15.9	588			
Jharkhand	7.5	18.3	27.8	19.6	17.0	716			
Orissa	18.4	35.6	44.0	40.9	33.3	1166			
West Bengal	12.3	42.0	47.6	44.3	40.7	1699			
Northeast									
Arunachal Pradesh	6.5	28.5	30.2	23.7	22.7	344			
Assam	9.6	22.4	31.0	15.9	13.9	887			
Manipur	10.5	45.9	59.0	50.1	46.4	1267			
Maghalaya	8.1	29.0	31.1	33.2	28.8	540			
Mizoram	87	59.8	65.4	53.5	20.0 50.6	453			
Nagaland	1.0	11.6	05.4 24 7	11.8	10.6	866			
Sikkim	27.2	11.0	24.7 53.7	52.4	10.0	800 474			
Tripura	10.6	47.2	JJ.7 10 0	32.4	44.9	4/4			
XX74	10.0	40.9	48.8	33.7	30.3	410			
west	<i></i>	02.2	04.0	02.0		704			
Guiarat	55.7	92.3	94.0	82.8	/5.5	/84			
Oujaiai Maharashtra	25.6	52.7	63.0	61.4	56.5	949			
Ivialiarasilura	21.6	64.6	68.7	64.0	58.7	2171			
South									
Andhra Pradesh	28.2	64.4	74.9	73.3	64.1	1604			
Karnataka	29.6	64.7	69.7	66.9	58.5	1421			
Kerala	63.6	99.3	99.4	87.4	84.9	822			
I amil Nadu	34.0	87.8	90.6	913	87.2	1325			

Table.1Maternal Health Care Services for births during the five years
preceding the survey by State.

Source: NFHS-3



Table2.Logistic Models of dichotomous variables of Received Antenatal Care,
Assistance received during Pregnancy and Received Postnatal Cares
(Dependent Variables) with Mothers socio-economic background
variables

	Odds Ratio or Expected Beta Values with Significance					
Background Variables	Received Ante Natal	Assistance for	Received Postnatal	Reference Categories		
	Cares	Delivery	Cares			
Place of Residence				Urban		
Rural	1.11*	1.13	1.13*	orban		
Current Age						
21 - 30 Years	2.36***	0.50*	1.61*	Dalary 20		
31 - 40 Years	1.75**	.063*	1.71*	Delow 20		
41 and Above	1.48*	0.89	1.37	1 Cais		
Religion						
Muslim	1.24*	0.48***	0.78*			
Christian	1.05	0.29***	0.68**	Hindu		
Others	0.84	1.58*	0.40***			
Education						
Primary Complete	0.46***	4.21***	0.36***			
SSC complete	0.68*	2.69**	0.59**	Tilitanata		
HSC Complete	0.66*	3.21**	0.48***	Innerate		
Degree and Above	0.82	1.79*	0.60**			
Partner Education						
Primary Complete	0.91	1.13	0.76			
SSC complete	1.28	0.82	0.88	T11		
HSC Complete	1.34	0.54	0.88	Illiterate		
Degree and Above	1.38	0.28*	0.75			
Respondent Occupation						
Working	1.02	0.00	1.23***	Not working		
Partner Occupation						
Working	0.23**	1.86*	0.73	Not working		
Constant	0.11***	.01***	0.21***			

Note: *** (1 % level of significance), ** (5 % level of significance), * (10 % level of significance).

1. Dependent Variable 0 = Not received, 1 = Received any Antenatal Care

2. Dependent Variable 0 = Not Assisted, 1 = Assisted during Delivery

3. Dependent Variable 0 = Not received, 1 = Receive any Postnatal Care



Child Health Care Services

Table.3 shows vaccination coverage rates for each recommended vaccination and the percentage of mothers showing a for children 12-23 months in India and each state. There are considerable interstate differentials in the coverage rates for different vaccinations and for children receiving all vaccinations. The percentage of children who are fully vaccinated ranges from 21 percent in Nagaland to 81 percent in Tamil Nadu. Tamil Nadu, Goa, Kerala and Himachal Pradesh stand out in full immunized. Among the more populous states, Uttar Pradesh (23 percent), Rajasthan (27 percent), Assam (31 percent), Bihar (33 percent), Jharkhand (34 percent), and Madhya Pradesh (40 percent) stand out as having a much lower percentage of children fully vaccinated than the national average of 44 percent. As these states account for nearly one-third of the total population of the country, their low vaccination coverage pulls down the coverage rate for the country as a whole. In addition to Nagaland and Assam, some of the other northeastern states (Arunachal Pradesh and Meghalaya) also have a relatively poor record on vaccination coverage.

A similar picture emerges with respect to individual vaccinations. In Tamil Nadu, Himachal Pradesh, Goa, Kerala, Sikkim, and Maharashtra, the coverage for BCG and at least the first doses of DPT and polio is generally in excess of 90 percent and in some cases, nearly universal. In Tamil Nadu and Goa, measles coverage is also above 90 percent. However, in most of the states, there is a considerable drop from the second to the third dose for both DPT and polio, and in almost every state fewer children have received measles vaccine than any of the other vaccinations except polio 0. The polio 0 is the polio vaccination given at the time of birth.

The analysis of logistic regression models were also carried out to examine the factor affecting the child health care services. In this models there are four dependent variables or indicators of child health information have been considered with the mothers socio-economic background predictor variable of place of the residence, current age, religion, education of the other, partner education, mothers occupation and partners 'occupation were considered in the analysis. The mothers education has found very high significant (99 percent level) with all the four indicators. Similar pattern also observed in the case of current age of the mother except the measles vaccination of the child. In other words the literate mothers all most double as compared with the illiterate mothers in all the receiving child heath care services in India.

Table.3	Percentage of children ag	ge 12-23 months	who received	specific v	accinations at
	any time before the surve	y (according to a	vaccination c	ard or mot	ner's report)

			DPT		Polio				Fully	
India /Major States	BCG	1	2	3	0	1	2	3	Measle s	Vaccinated
India	78.1	76.0	66.7	55.3	48.4	93.1	88.8	78.2	58.8	43.5
North										
Delhi	87.0	83.4	80.5	71.7	70.4	88.5	86.5	79.1	78.2	63.2
Haryana	84.9	83.8	81.0	74.2	52.7	92.2	91.3	82.8	75.5	65.3
Himachal Pradesh	97.2	96.6	91.9	85.1	67.1	96.8	94.6	88.6	86.3	74.2
Jammu & Kashmir	90.9	90.5	88.8	84.5	48.3	95.1	93.8	82.2	78.3	66.7
Punjab	88.0	85.9	80.4	70.5	65.6	90.1	86.7	75.9	78.0	60.1
Raiasthan	68.5	65.0	53.2	38.7	30.0	93.0	84.0	65.2	42.7	26.5
Uttaranchal	83.5	81.04	76.4	67.1	51.8	89.1	84.5	80.3	71.6	60.0
Central										
Chattisgarh	84.6	87.2	77.4	62.8	37.0	96.7	93.8	85.1	62.5	48.7
Madhya Pradesh	80.5	76.0	63.7	49.8	41.3	94.0	88.4	75.6	61.4	40.3
Uttar Pradesh	61.0	55.7	43.6	30.0	34.4	94.6	92.3	87.6	37.7	23.0
East										
Bihar	64.7	65.2	55.5	46.1	30.5	90.6	87.5	82.4	40.4	32.8
Jharkhand	72.7	66.0	53.2	40.3	25.2	93.4	87.2	79.3	47.6	34.2
Orissa	83.6	83.6	77.6	67.9	38.5	85.7	80.3	65.1	66.5	51.8
West Bengal	90.1	89.7	83.2	71.5	53.4	93.2	88.6	80.7	74.7	64.3
Northeast										
Arunachal Pradesh	57.7	57.0	48.4	39.3	34.3	72.6	65.5	55.8	38.3	28.4
Assam	62.4	66.7	56.2	44.9	27.5	81.6	72.7	59.0	37.4	31.4
Manipur	80.0	77.4	72.3	61.2	23.1	93.5	90.2	77.5	52.8	46.8
Maghalaya	65.9	62.0	56.0	47.3	31.0	81.5	74.2	56.6	43.8	32.9
Mizoram	86.4	89.1	84.5	66.8	46.4	89.0	83.7	63.5	69.5	46.5
Nagaland	46.3	47.5	36.3	28.7	13.2	79.8	68.4	46.2	27.3	21.0
Sikkim	95.9	94.9	91.2	84.3	63.4	94.0	91.2	85.6	83.1	69.6
Tripura	81.1	80.2	76.0	60.2	56.0	84.7	77.8	65.3	59.9	49.7
West										
Goa	96.8	95.7	92.6	87.5	85.6	98.6	94.0	87.2	91.2	78.6
Gujarat	86.4	82.2	73.4	61.4	59.9	92.6	83.5	65.3	65.7	45.2
Maharashtra	95.3	94.3	86.8	76.1	71.7	95.9	91.7	73.4	84.7	58.8
South										
Andhra Pradesh	92.9	92.6	76.4	61.4	28.3	96.2	94.5	79.2	69.4	46.0
Karnataka	87.8	86.7	81.5	74.0	75.1	91.8	87.9	73.8	72.0	55.0
Kerala	96.3	94.0	90.8	84.0	86.7	94.5	88.6	83.1	82.1	75.3
Tamil Nadu	99.5	98.9	97.7	95.7	94.5	99.6	96.3	87.8	92.5	80.9

in India and by state, 2005-06.



Table 4.Logistic Models of dichotomous variables of Received BCG, Received
All DPTs, Received All Polios and Received Measles (Dependent
Variables) with Mothers socio-economic background variables

	Odds Ratio or Expected Beta Values with Significance Levels								
Background Variables	Received BCG	Received All DPTs	Received All Polios	Received Measles	Reference Categories				
Place of Residence Rural	1.31***	0.98	0.91***	1.25***	Urban				
Current Age									
21 - 30 Years	1.16***	16.72***	4.56***	0.52***	Dalary 20				
31 - 40 Years	1.42***	1.21***	6.01***	1.02	Below 20				
41 and Above	1.22*	12.28***	9.04***	1.15*	10015				
Religion									
Muslim	1.038	0.92*	1.021	0.94					
Christian	0.58***	0.87***	1.12***	0.61***	Hindu				
Others	0.35***	0.87**	1.059	042***					
Education									
Primary Complete	0.10***	0.18***	0.31***	030***	Illitarata				
SSC complete	0.19***	0.24***	0.31***	049***					
HSC Complete	0.21***	0.29***	0.37***	053***	millerate				
Degree and Above	0.38***	0.44***	0.48***	0.79***					
Partner Education									
Primary Complete	0.85	0.79**	0.81**	085*					
SSC complete	1.15	0.94	0.82**	1.11	Illitarata				
HSC Complete	1.34**	01.11	1.00	1.22*	millerate				
Degree and Above	1.70***	1.20*	1.07	1.28*					
Respondent Occupation Working	3.72	8.70	12.62	6.15	Not working				
Partner Occupation Working	0.982	0.810**	0.82	.0.88	Not working				
Constant	12.18***	0.015***	0.02***	2.56***					

Note: *** (1 % level of significance), ** (5 % level of significance),

* (10 % level of significance).

1. Dependent Variable 0 = Not received, 1= Received BCG

2. Dependent Variable 0 = Not Assisted, 1 = Received All DPTs

3. Dependent Variable 0 = Not received, 1 = Received ALL Polios

4. Dependent Variable 0 = Not received, 1 = Received Measles



Conclusions

To improve the availability of and access to quality health care, especially for those residing in rural areas, the poor, women, and children, the government recently launched the National Rural Health Mission for the 2005-2012 period. One of the important goals of the National Rural Health Mission is to provide access to improved health care at the household level through female Accredited Social Activities (ASHA), who act as an interface between the community and the public the health system. The ASHA acts as a bridge between the ANM and the village, and she is accountable to the Panchayat. She helps promote referrals for universal immunization, escort services for RCH, construction of household toilets, and other health care delivery programmes.

An examination of the performance of each state on the different safe motherhood indicators shows that several states consistently perform well below the national average one each of the five services. The logistic regression analysis has been observes that compared with all the background variables the mother's education has high significance in all the three factors of maternal health care services. Another very important finding was that mother's religions have significance (99 percent level of significance) with assistance during delivery and postnatal care services.

The percentage of children who are fully vaccinated ranges from 21 percent in Nagaland to 81 percent in Tamil Nadu. Tamil Nadu, Goa, Kerala and Himachal Pradesh stand out in full immunized. Among the more populous states, Uttar Pradesh (23 percent), Rajasthan (27 percent), Assam (31 percent), Bihar (33 percent), Jharkhand (34 percent), and Madhya Pradesh (40 percent) stand out as having a much lower percentage of children fully vaccinated than the national average of 44 percent. The mother's education has found very high significant (99 percent level) with all the four indicators of child immunizations was found by using logistic regression analysis.



REFERENCES

- Barrera, A. 1990. The role of maternal schooling and its interaction with public health programs in child health production. *Journal of Development Economics* 32:69–91.
- Bhuiya, Abbas, and Kim Streatfield. 1991. Mothers' education and survival of female children in a rural area of Bangladesh. *Population Studies* 45:253–64.
- Bicego, George, and J. Ties Boerma. 1993. Maternal education and child survival: A comparative study

of survey data from 17 countries. Social Science and Medicine 36:1207-27.

- Boerma, J. Ties, Elisabeth A. Sommerfelt, Shea O. Rutstein, and Guillermo Rojas. 1990. *Immunization: Levels, trends, and differentials*. DHS Comparative Studies, No. 1. Columbia, Maryland: Institute for Resource Development.
- Bourne, Katherine L., and George M. Walker, Jr. 1991. The differential effect of mothers' education on mortality of boys and girls in India. *Population Studies* 45:203–19.
- Caldwell, John. 1979. Education as a factor in mortality decline: An examination of Nigerian data. *Population Studies* 33:395–413.

_. 1986. Routes to low mortality in poor countries. Population and Development Review

12:171-220.

_____. 1990. Cultural and social factors influencing mortality in developing countries. *The Annals of the American Academy of Political and Social Science* 510:44–59.

- Caldwell, John, and Pat Caldwell. 1988. *Women's position and child mortality and morbidity in LDC's*. Research Paper. Canberra: Department of Demography, Research School of Social Sciences, Australian National University.
- Caldwell, Pat, and John C. Caldwell. 1990. *Gender implications for survival in South Asia*. Health Transition Working Paper No. 7. Canberra: National Center for Epidemiology and Population Health, Australian National University.
- Cleland, John. 1990. Maternal education and child survival: Further evidence and explanations. In *What we know about health transition: The cultural, social, and behavioral determinants of health,* pp. 400–419. John Caldwell, Sally Findley, Pat Caldwell, Gigi Santow, Wendy Cosford, Jennifer Braid, and Daphne Broers-Freeman, eds. Vol. 1. Canberra: Australian National University.
- Cleland, John, and Jerome K. van Ginneken. 1988. Maternal education and child survival in developing countries: The search for pathways of influence. *Social Science and Medicine* 27:1357–68.
- Desai, Sonalde. 1994. *Maternal education and child health: Evidence and ideology*. Paper presented at the IUSSP Seminar on Women, Poverty, and Demographic Change, Oaxaca, Mexico, October 2–28.
- Dyson, Tim, and M. Moore. 1983. On kinship structure, female autonomy, and demographic behavior in India. *Population and Development Review* 9:35–60.
- Fauveau, V., M. Koenig, J. Chakraborty, and A. Chowdhury. 1988. Causes of maternal mortality in rural Bangladesh: 1976–1985. *Bulletin of the World Health Organization* 66:643–51.



- Goodburn E., G. J. Ebrahim, and Sishir Senapati. 1990. Strategies educated mothers use to ensure the health of their children. *Journal of Tropical Pediatrics* 36:235–39.
- Govindasamy, Pavalavalli, and B. M. Ramesh. 1996. *Maternal education and gender bias in child care practices in India.* Paper presented at the Annual Meeting of the Population Association of America, New Orleans, May 9–11.
- Govindasamy, Pavalavalli, M. Kathryn Stewart, Shea O. Rutstein, J. Ties Boerma, and A. Elisabeth Sommerfelt. 1993. *High-risk births and maternity care*. DHS Comparative Studies, No. 8. Columbia, Maryland: Macro International Inc.
- Hobcraft, John. 1993. Women's education, child welfare, and child survival: A review of the evidence. *Health Transition Review* 3:159–75.
- Hobcraft, John, J. McDonald, and S. O. Rutstein. 1985. Socioeconomic factors in infant and child mortality: A cross-national comparison. *Population Studies* 38: 193–223.
- International Institute for Population Sciences (IIPS). 1995. National Family Health Survey (MCH and Family Planning), India 1992–93. Bombay: IIPS.
- Monteith, Richard S., Charles W. Warren, Egberto Stanziola, Ricardo Lopez Urzua, and Mark W. Oberle. 1987. Use of maternal and child health services and immunization coverage in Panama and Guatemala. *Bulletin of the Pan American Health Organization* 21:1–15.
- Mosley, W. H., and L. C. Chen. 1984. An analytical framework for the study of child survival in developing countries. *Population and Development Review* 10 (Supplement): 25–45.
- Murthi, Mamta, Anne-Catherine Guio, and Jean Dreze. 1995. Mortality, fertility, and gender bias in India: A district-level analysis. *Population and Development Review* 21:745–82.
- Okafor, C. B. 1991. Availability and use of services for maternal and child health care in rural Nigeria. *International Journal of Gynecology and Obstetrics* 34:331–46.
- Rosenzweig, M., and T. Paul Schultz. 1982. Child mortality and fertility in Colombia: Individual and community effects. *Health Policy and Education* 2:305–348.
- Schultz, T. Paul. 1984. Studying the impact of household economic and community variables on child mortality. *Population and Development Review* (Supplement) 10:215–35.
- UNICEF. 1990. Development goals and strategies for children in the 1990s. New York: UNICEF.
- Ware, Helen. 1984. Effects of maternal education, women's roles, and child care on child mortality. *Population and Development Review* (Supplement) 10:191–214.
- Westoff, Charles. 1986. *Technical Notes No. 1*. Demographic and Health Surveys. Columbia, Maryland: Institute for Resource Development/Westinghouse.
- Wong, Emelita L., Barry M. Popkin, David K. Guilkey, and John S. Akin. 1987. Accessibility, quality of care, and prenatal care use in the Philippines. *Social Science and Medicine* 24:927–44.