ANUL L **2005**

S T U D I A

UNIVERSITATIS BABEŞ–BOLYAI

OECONOMICA

2

Editorial Office: 400015 – Cluj–Napoca Republicii no. 24, Phone: 0264-405352

SUMAR - SOMMAIRE - CONTENTS - INHALT

STUDIA UNIVERSITATIS BABEŞ-BOLYAI, OECONOMICA, L, 2, 2005

LIFE AMORTIZATION IN DETERMINISTIC AND STOCHASTIC ENVIRONMENT

E. CARDONA[*], P. De ANGELIS, E. VOLPE di PRIGNANO, A. FORTUNATI[**]

[*] *Federico II University of Napoli, Italy*
[**] *La Sapienza University of Rome, Italy*

ABSTRACT. In this paper we deal with the life annuity amortization, classical argument of financial mathematics with applications in banking. We are going to analyze the main characteristics and we give the formulations referred to different choices of fixed, variable and stochastic interest rates.

JEL Classification: C15, G22

Keywords: insurance, life amortisation, interest rates, simulation

1. Recalls on the life amortization with constant rate

It is known that the refund of a financial loan, or an *amortization*, can be carried out by one final or gradual payment, in accordance with different methods. One of these sometimes applied which has an actuarial origin, is the *life annuity amortization*, brought into effect gradually by a life annuity as it has the clause characteristic where the outstanding debt is extinguished in the case of the borrower's death prior to the contractual maturity of repayments. In the event of premature death, such a contractual feature will leave any heirs to the borrower free from any outstanding payments.

This clause satisfies a necessity in not leaving the heirs with the responsibility for settling any outstanding debt left by the borrower. The clause is also useful and sometimes essential in the event of limited financial resources to settle the outstanding debt. For example, a single-income family whose only investment is the family house bought through mortgage repayments. In case of the breadwinners' sudden death with or without a low survival pension, not being able to rent the house due its occupation, the surviving relatives could have great difficulty in continuing the residual mortgage payments. The formal development of such model of amortization, mainly with front loaded payments, bears in respect to non risk case an increase in such payments, each consisting of:

1. *Capital paid* in advance, intended to reduce the outstanding loan balance;
2. *Interest paid* in advance, for the interest payments on the remaining debt at the beginning of each period;

3. A *natural periodical premium* paid-due to a term insurance in the event of the borrower's premature death on the capital referred to the remaining debt.

The sum of the 2nd and 3rd components gives the periodical total cost for the borrower. We call it: *actuarial interest amount.*

We recall here the known formulas representing for each period the constraints between the principal share, the actuarial interest, the payments and the outstanding loan balance. For simplicity we will assume annual periods.

Let $n \in \mathbb{N}$ be the amortization length (in years) in absence of death and *i* be the annual interest rate, both agreed upon time 0. Moreover, let $S=D_0$ be the initial debt, referred to a borrower with *x* (integer) age, that he refunds if survivor by annual-due ${}_z\ddot{\alpha}_{x,n,i,S}$ payments, generally variable, in *z*=0,...,*n*-1. It is to be valid the following actuarial equivalence relation

(1) $$\sum_{z=0}^{n-1} {}_z\ddot{\alpha}_{x,n,i,S} \ {}_zE_x = S$$

being ${}_zE_x = l_{x+z}(1+i)^{-z} / l_x$ the *actuarial discount factor* for *z* years and *x* age, which substitutes the *financial discount factor* $(1+i)^{-z}$ applied in non risk amortization. The payment ${}_z\ddot{\alpha}_{x,n,i,S}$, abbreviated in $\ddot{\alpha}_z$, for the period (*z* ,*z*+1) it includes

- the capital *share* ${}_z\ddot{c}_{x,n,i,S}$, abbreviated in $\ddot{c}_z$;
- the *actuarial interest share* ${}_z\ddot{j}_{x,n,i,S}$, abbreviated in $\ddot{j}_z$, which is the sum of the financial interest and the insurance premium for the death risk.

Let $d = 1-(1+i)^{-1}$ be the annual constant *discount rate*; D_z be the *outstanding debts* at the time *z*, q_{x+z} be the death probability in (*z* ,*z*+1), (*z*=0,...,*n*-1). If the capital shares $\ddot{c}_z$ are assigned, subject to the constraint

(2) $$\sum_{z=0}^{n-1} \ddot{c}_z = S$$

it results

(3) $$\ddot{c}_z = D_z - D_{z+1}$$

(4) $$\ddot{j}_z = dD_{z+1} + (1-d)q_{x+z}D_{z+1} = (1-{}_1E_{x+z})D_{z+1}$$

(5) $$\ddot{\alpha}_z = \ddot{c}_z + \ddot{j}_z = D_z - {}_1E_{x+z}D_{z+1}$$

If alternatively we fix the payments $\ddot{\alpha}_z$ subject to the constraint (1), then we deduce for *z*=1,...,*n*-1 the *outstanding debts*

(6) $$D_z = \sum_{k=z}^{n-1} \ddot{\alpha}_k \ {}_{k-z}E_{x+z}$$

and from them the values $\ddot{c}_z$ by (3) and $\ddot{j}_z$ by (5).

The *retrospective reserves* M_z are written

$$M_z = \frac{S - \sum_{k=0}^{z-1} \ddot{\alpha}_k \; {}_k E_x}{{}_z E_x} \quad , \qquad z = 0,\ldots,n\text{-}1 \tag{7}$$

whereas the *prospective reserves* W_z are formally expressed by the right hand of (6) and therefore are equal to the outstanding debts D_z if the technical basis fixed in 0, by which ${}_z E_x$ follows, is unchanged in z. The annual insurance premiums $(1-d)q_{x+z}D_{z+1}$, which are a share of $\ddot{\alpha}_z$, being proportional to the outstanding debts, usually decrease if z increases, in spite of the rise of death probabilities, modestly for young age groups. Consequently they can result initially high, as to discourage the borrower. From a theoretical point of view the constant insurance premiums could be taken into consideration. But in such a manner to not reduce the premium annuity duration to have a constant premium not lower than each natural premium (which does not solve that problem for the borrower), the mathematical reserve that would be constituted would result negative, therefore at the cost of the borrower, such is practically unacceptable. A numerical illustration follows demonstrating a life amortization corresponding to a debt of € 95000 made by a borrower aged 42, refunded by 10 payments-due, rate 3,5%. The demographic tables M/SI99 are used.

Table 1:

Life amortization under fixed rate

	debt =	95000,00	rate =	0,035		duration =	10
z	*l(42+z)*	*E(42+z)*	*c(z)*	*B(z)*	*D(z)*	*j(z)*	*payment(z)*
0	96400,00	0,964460	8700,00	0,00	95000,00	3067,13	11767,13
1	96228,00	0,964356	8650,00	8700,00	86300,00	2767,74	11417,74
2	96046,00	0,964202	9800,00	17350,00	77650,00	2428,91	12228,91
3	95849,00	0,963986	9600,00	27150,00	67850,00	2097,81	11697,81
4	95631,00	0,963708	9300,00	36750,00	58250,00	1776,48	11076,48
5	95386,00	0,963408	10100,00	46050,00	48950,00	1421,59	11521,59
6	95112,00	0,963095	9700,00	56150,00	38850,00	1075,77	10775,77
7	94808,00	0,962861	9750,00	65850,00	29150,00	720,49	10470,49
8	94482,00	0,962512	9200,00	75600,00	19400,00	382,37	9582,37
9	94123,00	0,962314	10200,00	84800,00	10200,00	0,00	10200,00
10	93746,00			95000,00	0,00		

2. Life amortization with two rates in deterministic environment

Analysing the development of § 1 we can observe that the components of the actuarial interest amount are antithetical in reference to the rate involved. In the interest payment, the rate relates to the borrower's debt position which for him it is a *debit rate,* whereas the paid-due premium is accumulated to

construct the tribute at the year's end in case of death, so that the rate is a *credit rate.* Therefore the life amortization model with one constant rate, seen in § 1, must be considered a simple way to model rates behaviour: by market law a credit rate is indeed different and lower than a contemporary debit rate. It follows as better practice, to apply a rigorous approach to the problem, to consider the two components of the actuarial interest amount depending on two different rates. A related formulation is here developed. We denote by *i'* the annual constant debit rate of interest that intervenes on the financial interest share; and with *i"* the analogous credit rate working on the insurance premium. The actuarial interest share is a sum of two components so modified.
The new value

$$\ddot{j}_z = \left\{ \left[1-(1+i')^{-1}\right] + (1+i'')^{-1} q_{x+z} \right\} D_{z+1} \ , \qquad (z = 0,...,n\text{-}1) \tag{8}$$

which helps in making the payment $\ddot{\alpha}_z = \ddot{c}_z + \ddot{j}_z$, results.
It is easy to verify that *i* cost rate for the borrower can be implicitly deduced by (1), where $\ddot{\alpha}_z$ takes in account (8) instead of (4). It usually results: *i* > *i'* > *i".*

The model above developed with two different rates is somehow similar to *two rates American amortization*, where the debit interest rate is different by (and greater than) the credit rate which regulates the sinking fund accumulation.

Another numerical illustration follows in Table II. In this one, with the same data as Table I, we introduce also the 2,5% credit rate.

Table 2:

Life Amortization under two rates

	debt =	95000,00		d. rate =	0,035	duration =	10
				c. rate =	0,025		
z	*l(42+z)*	*E(42+z)*	*c(z)*	*B(z)*	*D(z)*	*j(z)*	*payment(z)*
0	96400,00	0,96446	8700,00	0,00	95000,00	3068,58	11768,58
1	96228,00	0,964356	8650,00	8700,00	86300,00	2769,13	11419,13
2	96046,00	0,964202	9800,00	17350,00	77650,00	2430,22	12230,22
3	95849,00	0,963986	9600,00	27150,00	67850,00	2099,06	11699,06
4	95631,00	0,963708	9300,00	36750,00	58250,00	1777,66	11077,66
5	95386,00	0,963408	10100,00	46050,00	48950,00	1422,64	11522,64
6	95112,00	0,963095	9700,00	56150,00	38850,00	1076,65	10776,65
7	94808,00	0,962861	9750,00	65850,00	29150,00	721,12	10471,12
8	94482,00	0,962512	9200,00	75600,00	19400,00	382,74	9582,74
9	94123,00	0,962314	10200,00	84800,00	10200,00	0,00	10200,00
10	93746,00			95000,00	0,00		

3. Life amortization in term structure environment

In § 1 and 2 we have taken into consideration the classic constant rate approach, distinguishing or not between credit and debit rate. However, in both cases it can be considered the more general approach with variable rates, for instance in conformity with a term structure fixed at the beginning of the redemption and valid for the total period.

The formal development of the amortization under the constraints among principal and actuarial interest amounts, total payments and outstanding debts, can be obtained making the calculus $\ddot{j}_z$ on the basis of a structure of *debit forward uniperiodal rates* $\{i'_{r-1,r}\}$ and *credit* similar ones $\{i''_{r-1,r}\}$, both fixed in 0. If it's not possible, we can put *spot rates*. We can assume that between the structures operating on $\ddot{j}_z$ holds an additive connection, such as

(9) $$i''_{r-1,r} = i'_{r-1,r} - h \quad , \qquad \forall r; \;\; h > 0$$

or multiplicative, such as

(9') $$i''_{r-1,r} = \lambda \, i'_{r-1,r} \quad , \qquad \forall r; \;\; 0 < \lambda < 1$$

In any case with different debit and credit structures, a variable cost-rates structure which generalises IRR can be deduced in infinite ways by establishing a financial equivalence which generalizes (1).

We do not insist upon the general relations regarding rates, but on a particular case. We should note the effect of the gap between credit and debit rates is small in respect to the outstanding debt. With constant rates, however, it is possible to observe that the spread between (8) and (4) values of the actuarial interest share, calculated on the basis of different and equal rates, is $q_{x+z} D_{z+1} \Delta$, where Δ is the gap between the annual discount factors evaluated by using the credit and the debit rate. This allows us to simplify the treatment, $\lambda = 1$ in (9'), so unifying the term structures to apply to the two components of $\ddot{j}_z$. We will mark by "*" such a single structure and quantities connected by the same, which then gives rise to a law generalizing IRR.

Then in such rates scenario a fair life amortization is achieved if the actuarial equivalence relation (1) is replaced by

(10) $$\sum_{z=0}^{n-1} \ddot{\alpha}^*_z \; {}_z E^*_x = S$$

with the following symbols' meaning: the *actuarial discount factor* for the period (z,z+1) referring to a people x aged at the time 0 is clearly given by

(11) $${}_1 E^*_{x+z} = \frac{l_{x+z+1}}{l_{x+z}} (1 + i_{z,z+1})^{-1} \quad ;$$

that one for (0,z) is given by

(12) $$ {}_zE_x^* = \frac{l_{x+z}}{l_x}(1+i_z)^{-z} = \prod_{r=0}^{z-1}\frac{l_{x+r+1}}{l_{x+r}}(1+i_{r,r+1})^{-1} $$

being i_z the *spot rate* on (0,*z*). Clearly it results:

(12') $$ {}_zE_x^* = \prod_{r=0}^{z-1} {}_1E_{x+r}^* $$

the quantity to put in (10).

Let's now remember that the uniperiodal forward discount rates $d_{r-1,r}$ are linked to the similar interest rates $i_{r-1,r}$ by the constraint

(13) $$ 1 - d_{r-1,r} = (1+i_{r-1,r})^{-1} = s_{r-1,r} $$

where $s_{r-1,r}$ is the present value (or forward price) in *r*-1 of the unitary bond payable in *r*. So, fixed the *principal shares* $\ddot{c}_z$ under the constraint (2), the payments value is given, as in (5), by

(14) $$ \ddot{\alpha}_z^* = \ddot{c}_z + \ddot{j}_z^* \quad , \qquad (z = 0,...,n-1) $$

where

(15) $$ \ddot{j}_z^* = \left[d_{z,z+1} + (1-d_{z,z+1})q_{x+z} \right] D_{z+1} = (1 - {}_1E_{x+z}^*)D_{z+1} \quad , \qquad (z = 0,...,n\text{-}1) $$

However, taking also in account (3), it results

(16) $$ \ddot{\alpha}_z^* = D_z - {}_1E_{x+z}^* D_{z+1} \quad , \qquad (z = 0,...,n\text{-}1) $$

and this value is to put in (10). It is to be noticed the analogy of (15) and (16) with (4) and (5).

The following Table 3 refers to an illustration made on the basis of forward rates deduced by a given term structure.

Table 3:

Life Amortization under variables forward rates

	debt =	95000,00			duration =	10	
z	*l(42+z)*	*i(z-1,z)*	*E*(42+z)*	*c(z)*	*D(z)*	*j(z)*	*z)*
0	96400,00		0,950682	8700,00	95000,00	4256,17	12956,17
1	96228,00	0,050	0,952394	8650,00	86300,00	3696,62	12346,62
2	96046,00	0,048	0,954062	9800,00	77650,00	3116,89	12916,89
3	95849,00	0,046	0,955676	9600,00	67850,00	2581,88	12181,88
4	95631,00	0,044	0,957234	9300,00	58250,00	2093,38	11393,38
5	95386,00	0,042	0,958776	10100,00	48950,00	1601,54	11701,54
6	95112,00	0,040	0,955708	9700,00	38850,00	1291,10	10991,10
7	94808,00	0,043	0,952736	9750,00	29150,00	916,93	10666,93
8	94482,00	0,046	0,949667	9200,00	19400,00	513,40	9713,40
9	94123,00	0,049	0,946763	10200,00	10200,00	0,00	10200,00
10	93746,00	0,052			0,00		

4. A stochastic approach for a life annuity amortization

In order to follow a stochastic approach for calculating a life annuity amortization, let's introduce two risk sources referring respectively to a financial risk (i.e. interest rates related to the financial amortization and the technical reserves investments) and a demographic risk. In particular assumptions made in our calculations are described below. We assume that financial and insurance markets are perfectly competitive, frictionless, and free of arbitrage opportunities. Moreover, all the agents are supposed to be rational and non-satiated, and to share the same information.

We consider $\{r_t\}; t = 1, 2, ...$, and $\{\mu_{x+t}\}; t = 1, 2, ...,$ as two diffusion processes driven the instantaneous interest rate and the intensity of mortality (referred to an insured aged *x* at issue), by the filtrations F^r and F^μ respectively; with reference to a generic insurance contract pay-out, the two stochastic processes are defined on a probability space $\left(\Omega, F^{r,\mu}, \mathrm{P}\right)$ such that $F^{r,\mu} = F^r \cup F^\mu$.

The demographic stochastic process $\{\mu_{x+t}\}; t = 1, 2, ...,$ is described by a Mean-Reverting Brownian Gompertz (MRBG) model; in particular, we take into account a traditional actuarial approach, where T_x is a random variable representing the remaining lifetime of a policyholder aged x; as a consequence the probability of survival to time s, conditional on being alive at age x, is equal to

$$ {}_s p_x = \mathrm{Prob}\left(T_x > \mathrm{s} \middle| F_t^\mu\right) \tag{17} $$

We define $\mu_{x+t:t}$ as the hazard rate for an individual aged $x+t$, at calendar year t, it follows that (17) can be arranged as:

$$ {}_s p_x = \hat{E}\left[e^{-\int_0^s \mu_{x+t:t} dt} \middle| F_t^\mu \right] \tag{18} $$

The time evolution of the form as follows $\mu_{x+t:t} = \mu_{x:0} e^{g_{x,s}\, t + \sigma_x Y_t}$, with $g_{x,s}, \sigma_x, \mu_{x:0} > 0$, where $g_{x,s}$ resume on time s the deterministic correction due to age x and the effect of *longevity risk;* $\{Y_t\}$ is a stochastic process introduced to model random variations in the forecast trends; σ_x represents the standard deviation of the process $\{\mu_{x+t}\}; t = 1, 2, ...,$ in particular the stochastic process $\{Y_t\}$ is described by a *mean reverting* diffusion process:

$$ dY_t = -bY_t dt + dW_t, \quad Y_0 = 0, \quad b \geq 0 \tag{19} $$

where b is the *mean reversion* coefficient and $\{W_t\}$ is a standard Brownian motion.

Both interest rate paid over the financial amortization procedure and interest rate earned over technical reserves investment depend on the time dynamic of the instantaneous interest (spot) rate; in particular the spot rate $\{r_t\}; t = 1, 2, ...$, is modelled by a mean reverting square root diffusion equation as in Cox, Ingersoll and Ross model (CIR); therefore, we assume the following stochastic equation:

(20) $$dr_t = k(\theta - r_t)dt + \sigma_r \sqrt{r_t}\, dZ_t^r$$

where k is the *mean reversion* coefficient, θ is the long term rate, σ_r is the volatility parameter and $\{Z_t^r\}$is a standard Brownian motion.

In order to model the dynamic of interest rate earned, we assume to work in a Black-Scholes economy where the reference portfolio in which insurance reserves are invested is compounded mainly by a bond index and a minority by a stock index. The two components are described by the following equation

(21) $$dS_t^{(i)} = r_t\, S_t^{(i)} dt + \sigma^{(i)}\, S_t^{(i)} dZ_t^{(i)}$$

$$i = \begin{cases} 1 : \text{stock index} \\ 2 : \text{bond index} \end{cases}$$

where $S_t^{(i)}, \sigma^{(i)}$ e $\{Z_t^{(i)}\}$, are, for each reference portfolio's component, market price, volatility parameter and a *Wiener* process. At last, the three sources of financial uncertainty are correlated:

(22) $$dZ_t^{(k)} dZ_t^{(j)} = \rho_{k,j} dt \qquad k, j = 1,2,r$$

hence, reference portfolio could be expressed as a combination of the random variables introduced above

(23) $$S_t = (1-\alpha) S_t^{(1)} + \alpha S_t^{(2)}$$

As a result, the annual rate of return of the reference fund at time t is defined as:

(24) $$i''_{t-1,t} = \frac{S_t}{S_{t-1}} - 1$$

Substituting (24) in the equation (9'), it becames:

(25) $$i'_{t-1,t} = \frac{\left(\frac{S_t}{S_{t-1}} - 1 \right)}{\lambda}$$

where we have put $\lambda = 0{,}9$, to yield our calculations.

Via Monte Carlo procedures we have obtained results shown in § 3 in a stochastic environment, running 10000 trajectories respectively for μ_{x+t} , $i''_{t-1,t}$ and $i'_{t-1,t}$; afterwards, in reference to every trajectory of the three processes, we have worked out a life annuity amortization scheme.

In particular in Table 4, first moments of main values of the life annuity amortization stochastic scheme are presented, each one calculated as an arithmetic mean of values produced by the simulation procedure.

Table 4:

Life Amortization in stochastic environment: Mean Values

	debt =	95000,00				duration =	10
z	*deb. rate*	*cr. Rate*	*E*(42+z)*	*c(z)*	*D(z)*	*j(z)*	*□(z)*
0	0,024	0,022	0,977	8.700,00	95.000,00	2.084,55	10.784,55
1	0,031	0,028	0,971	8.650,00	86.300,00	2.415,77	11.065,77
2	0,036	0,033	0,967	9.800,00	77.650,00	2.433,54	12.233,54
3	0,041	0,037	0,963	9.600,00	67.850,00	2.348,85	11.948,85
4	0,045	0,040	0,960	9.300,00	58.250,00	2.144,57	11.444,57
5	0,047	0,043	0,958	10.100,00	48.950,00	1.810,84	11.910,84
6	0,049	0,045	0,956	9.700,00	38.850,00	1.415,81	11.115,81
7	0,051	0,046	0,954	9.750,00	29.150,00	980,94	10.730,94
8	0,053	0,047	0,953	9.200,00	19.400,00	529,17	9.729,17
9	0,054	0,049	0,951	10.200,00	10.200,00	0,00	10.200,00
10					0,00		

Final remarks about computational results of the stochastic approach:

- Table 5 shows values of parameters used in mortality and financial risk diffusion processes. In particular, with reference to CIR model, risk-adjusted parameters are calibrated by means of Brown and Dybvig framework (1986) on market value at the 1th October 2003 of euro swap interest rates. Parameters for MRGB model are calibrated on mortality hazard rates derived from an Italian projected life table called “RG48”. At last, reference fund parameters are estimated on daily market values of Emu-Bond Index and MSCI World Index observed between 2001 and 2003.

Table 5:

Estimated Parameters set

r_t		μ_{x+t}		S_t	
r_0	0.015268	b	0.50	$S_0^{(1)}$	123.57
k	0.245439	σ	0.17	$S_0^{(2)}$	1240.22
θ	0.058359	g	0.10	$\sigma^{(1)}$	0.01
σ_r	0.053524			$\sigma^{(2)}$	0.11

- Figures 1 and 2 show respectively paths of the spot rate (under CIR model) and the hazard rate (under MRBG model).

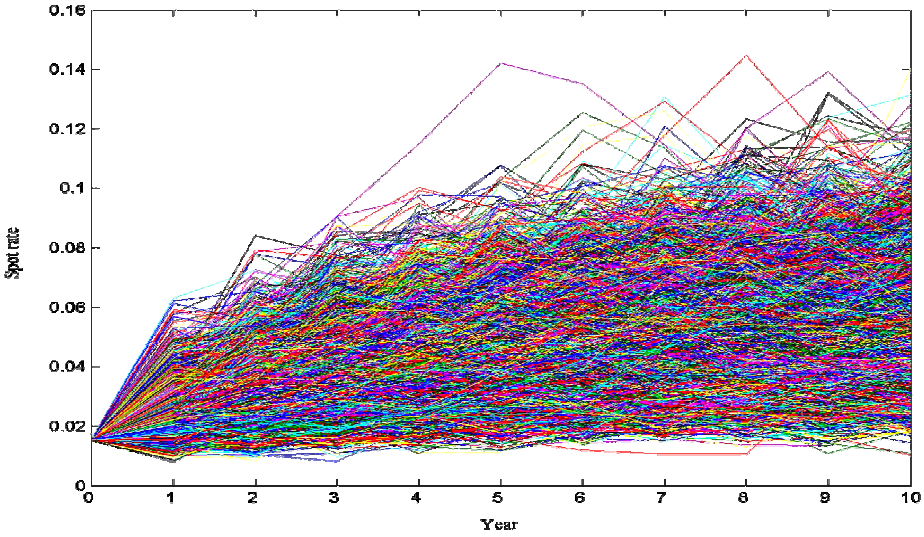


Figure 1. Spot rate paths (10.000 trajectories)

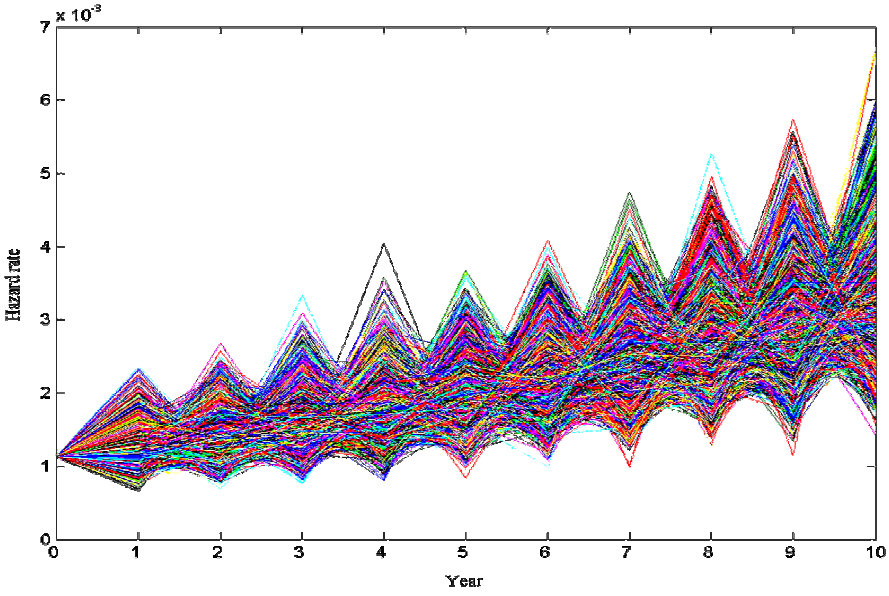


Figure 2. Hazard rate paths (10.000 trajectories)

- Table 6 shows the Internal Cost Rate distribution characteristics derived from simulations results.

Table 6:

Internal cost rate main characteristics

Mean	St. Dev.	1%-quantile	5%-quantile	95%-quantile	99%-quantile	Skewness	Kurtosis
0,036	0,011	0,015	0,020	0,055	0,064	0,355	3,236

REFERENCES

Baione F., De Angelis P., Fortunati A. (2004), "On A Fair Value Model For Traditional Italian Life Insurance Policies", *8th International Congress on Insurance Mathematics and Economics*, Rome, June 14-16

Boggio T., Giaccardi F., *Compendio di Matematica Attuariale,* Ed. Giappichelli.(I)

Brennan, M.J., Schwarz, E.S. (1976), "The Pricing of Equity-Linked Life Insurance Policies with an Asset Value Guarantee", *Journal of Financial Economics* n. 3, 3.

Brown J. S., Dybvig P. H. (1986), "The empirical implications of the Cox, Ingersoll, Ross theory of the term structure of interest rates", *Journal of Finance*, 4, 1-32.

Cox J. C., Ingersoll J. E., Ross S. A. (1985), "A theory of the term structure of interest rates", *Econometrica*, 53:-408.

Harrison J.M., Kreps D.M. (1989), "Martingales and Arbitrage in Multiperiod Securities Markets", *Journal of Economic Theory*, n. 20.

Milevsky M. A., Promislow S.D. (2001), "Mortality derivatives and the option to annuities", *Insurance: Mathematics and Economics*, (29): 231-245.

Volpe di Prignano E. (1985), "*Manuale di Matematica Finanziaria",* Ed. E.S.I. (I).

STUDIA UNIVERSITATIS BABEŞ-BOLYAI, OECONOMICA, L, 2, 2005

ESTIMATION OF THE MARGINAL CHANGES IN WILLINGNESS TO PAY FOR TRIPS AFFECTED BY CONGESTION

DIANA E. DUMITRAS*, KIMBERLY ROLLINS**

* *USAMV Cluj Napoca, Romania*
** *University of Nevada, Reno, U.S.A.*

ABSTRACT. The current study contributes with useful information for managing undeveloped public lands. The research focuses on three wilderness areas from Ontario, Canada. Data were collected through a survey at each site. The contingent valuation method is used with the single bounded dichotomous choice format and the random paired dichotomous choice format. A random effects probit model is utilized to estimate the marginal changes in willingness to pay for changes in congestion levels for recreation trips, for different points within a trip, and for different recreational activities. Grateful acknowledgments we address to Ontario Parks, Ontario Ministry of Natural Resources, for the resources provided for this study.

JEL Classification: C25, Q26

Keywords: recreation trip, probit, marginal effects

Introduction

The determination of the optimal use of public lands has become an important problem for public lands managers in many countries. The increasing demand for recreational use on public lands world-wide is a result of many factors, such as the demographic changes in population, income and preferences for leisure activities.

Public land management agencies are responsible for the creation of policies and management strategies that result in the most socially desirable combination of land uses. Conflicts can emerge when the marginal benefits of one user are lost or reduced due to actions of other users in the same area. Congestion occurs when a resource is used in common by a number of people who affect each others' utility. The level of congestion may have a significant impact on individuals' benefits, as measured by willingness to pay (WTP). The aim of this research is to estimate the marginal changes in willingness to pay for changes in congestion levels for recreation trips, for different points within a trip, and for different recreational activities.

For analyzing goods without market-derived values it is desirable to use the contingent valuation method to measure the benefits and marginal impacts of competing uses. Contingent valuation requires data collection using survey methods that directly elicit individual's valuation of public

goods by finding out what they would be willing to pay for certain qualitative or quantitative changes in the good (Mitchell and Carson, 1993). The valuation formats addressed to respondents are the single-bounded dichotomous choice (SBDC) and the random paired dichotomous choice (RPDC) (Dumitras, Rollins and Balazs, 2005).

The research focuses on three wilderness areas from Ontario, Canada. The areas are Kawartha Highlands, Killarney Provincial Park and Spanish River Valley. The sites have different site characteristics, and can support several types of recreational activities. Data were collected through a survey at each site, being first piloted to improve the efficiency of the estimates (Rollins, 1997). The overall response rate of the 2004 study was 63,46%, resulting in 1818 usable surveys.

The empirical model

The valuation of public goods is based on random utility maximization. The random utility model represents utility as being the sum of systematic and random components (Hanemann, 1984):

(1) $$U(j, y; s) = V(j, y; s) + \varepsilon_j, \quad j = 0,1$$

where U_j is the unobservable indirect utility, V the systematic component of utility, and ε the random component. Income (y) and other individual characteristics (s) can influence individual preferences. The offer will be accepted at an additional cost, P, only if:

(2) $$V_1(j, y - P, s) + \varepsilon_1 \geq V_0(j, y, s) + \varepsilon_0, \quad j = 0,1$$

and refuse it otherwise. The process is deterministic for the respondent. He knows which choice maximizes his utility. In contrast, the researcher does not observe everything about the individual, the process being stochastic from his viewpoint. For him the individual's response remains a random variable with a probability that is given by:

P_1 = Prob(Yes) = Prob$[V_1(1, y-P, s)+\varepsilon_1 \geq V_0(0, y, s)+\varepsilon_0] = F_\eta(\Delta V)$

and P_0 = Prob(No) = 1 - Prob(Yes)

where $\eta = \varepsilon_0 - \varepsilon_1$ and $F_\eta(\cdot)$ is the cumulative density function of η.

The maximum willingness to pay is established by setting ΔV equal to zero and solving for P. The cumulative density function is assumed to be standard normal, the probability that an individual accepts the offer corresponds to a standard probit model.

The present study extends the model presented by Boxall et al. (2003) by analyzing the marginal changes at five parts of the trip and simultaneously for several user types. Having eight responses per individual and willing to account for correlation between and within individuals'

responses it is desirable to arrange the data in a panel format. The first two observations correspond to the RPDC questions; the change in congestion is zero and the dependent variable the “Yes/No” response. The following five observations correspond to the SBDC questions; the change in congestion is either twice or half of actual stated congestion level and the dependent variable the “Yes/No” response. The last observation represents the actual trip taken; the change in congestion and the change in WTP are zero, the dependent variable is “Yes”. The panel has the following format:

(4)

$$V_1(c_1^*, c_2^*, c_3^*, c_4^*, c_5^*, y - P_1, s) + \varepsilon_1 \geq V_0(c_1^*, c_2^*, c_3^*, c_4^*, c_5^*, y, s) + \varepsilon_0,$$
$$V_2(c_1^*, c_2^*, c_3^*, c_4^*, c_5^*, y - P_2, s) + \varepsilon_2 \geq V_0(c_1^*, c_2^*, c_3^*, c_4^*, c_5^*, y, s) + \varepsilon_0,$$
$$V_3(c_1^1, c_2^*, c_3^*, c_4^*, c_5^*, y - P_3, s) + \varepsilon_3 \geq V_0(c_1^0, c_2^*, c_3^*, c_4^*, c_5^*, y, s) + \varepsilon_0,$$
$$V_4(c_1^*, c_2^1, c_3^*, c_4^*, c_5^*, y - P_4, s) + \varepsilon_4 \geq V_0(c_1^*, c_2^0, c_3^*, c_4^*, c_5^*, y, s) + \varepsilon_0,$$
$$V_5(c_1^*, c_2^*, c_3^1, c_4^*, c_5^*, y - P_5, s) + \varepsilon_5 \geq V_0(c_1^*, c_2^*, c_3^0, c_4^*, c_5^*, y, s) + \varepsilon_0,$$
$$V_6(c_1^*, c_2^*, c_3^*, c_4^1, c_5^*, y - P_6, s) + \varepsilon_6 \geq V_0(c_1^*, c_2^*, c_3^*, c_4^0, c_5^*, y, s) + \varepsilon_0,$$
$$V_7(c_1^*, c_2^*, c_3^*, c_4^*, c_5^1, y - P_7, s) + \varepsilon_7 \geq V_0(c_1^*, c_2^*, c_3^*, c_4^*, c_5^0, y, s) + \varepsilon_0,$$
$$V_8(c_1^*, c_2^*, c_3^*, c_4^*, c_5^*, y - P_8, s) + \varepsilon_8 \geq V_0(c_1^*, c_2^*, c_3^*, c_4^*, c_5^*, y, s) + \varepsilon_0.$$

where c_i is the congestion level, i = 1,2,3,4,5 resenting the five parts of the trip. The actual level of congestion is represented by the 0 superscript, the change (twice or half) by the 1 superscript and no change by the * superscript, P_i is the change in cost and *s* a vector of independent variables. The probability is given by (3).

A random effects probit model is used to estimate the welfare measures. The general expression is given by Greene (2003):

(5)
$$Y_{it}^* = \beta X_{it} + u_i + \varepsilon_{it}$$
$$Y_{it} = 1 \text{ if } Y_{it}^* > 0$$

$Y_{it} = 0$ otherwise,

where *i* subscript indicates the individual respondent, *t* indicates the number of responses per individual, Y_{it}^* is an unobserved latent variable, Y_{it} the observed random variable, X_{it} a vector of explanatory variables, and β is a vector of coefficients. The unobservable characteristic, u_i, is specific to individual *i*, does not vary among the *t* observations and is $\text{IN}(0, \sigma_u^2)$. These random disturbances are assumed to be uncorrelated with the

explanatory variables. The error terms, ε_{it}, vary among individuals and across the *t* observations and is IN $(0, \sigma_{\varepsilon}^2)$.

Results and discussions

One model is estimated to determine the marginal changes in WTP for changes in congestion levels, for different points within a trip and for different recreational activities. The dependent variable is the "Yes/No" response. The changes in congestion are included as explanatory variables. Other independent variables separated by site are cross terms representing the increase in trip costs, the trip length, the income per household member, groups with children under the age of 14, a dummy variable for survey versions, and a constant. The estimation results are presented in Table 1.

Table 1:

Estimation results

Variables	Kawartha Highlands		Killarney Park		Spanish River	
Increase in trip cost	-0.0399[a]	(0.0020)	-0.0235[a]	(0.0009)	-0.0163[a]	(0.0015)
Duration	0.0770[a]	(0.0103)	0.2084[a]	(0.0331)	0.2040[a]	(0.0627)
Income	0.0223[a]	(0.0059)	0.0105[a]	(0.0027)		
Children under 14			0.4647[a]	(0.1865)		
Congestion near access and departure points:						
- all activities					0.0499	(0.0539)
- canoeing	-0.1693[a]	(0.0480)	-0.0211[b]	(0.0121)		
- other activities	-0.0464[b]	(0.0257)				
- hiking			-0.0735[a]	(0.0257)		
Congestion while on water:						
- canoeing	-0.0367	(0.0244)	-0.0581[a]	(0.0169)		
Congestion while hiking:						
- hiking			-0.0598[b]	(0.0317)		
Congestion while camping:						
-0.0805[a]		(0.0342)	-0.0515[a]	(0.0135)	-0.4489[a]	(0.1685)
Congestion while portaging:					0.8024	(0.4943)
(Congestion while portaging)^2:					-0.3085[a]	(0.1405)
Constant	2.1466[a]	(0.1967)				
Version dummy	-0.4104[a]	(0.1357)				
Log likelihood	-2130.8745					
ρ	0.7515					
No. of observations	5344					
No. of groups	717					

Standard errors shown in parentheses;
[a] Significant at 1% level or above; [b]Significant at 10% level or above.

The coefficient for the increase in trip cost variable and duration of the trip variable are significantly different across the three sites. Income per household member variable is significant at Kawartha Highland and Killarney. 'Congestion near access and departure points' variable is significant at Kawartha and Killarney, for two different user types. The null hypothesis that 'congestion near access and departure points' is zero at Spanish River was rejected. 'Congestion while on water' is statistically significant at Killarney for canoeing trips. There is no statistical evidence to accept the hypothesis that 'congestion while on water' for people who went canoeing at Kawartha is zero. 'Congestion while hiking' is statistically significant only at Killarney because hiking is the main activity. 'Congestion while camping' is statistically different across sites for all user types. 'Congestion while portaging' for Spanish River is not significant at 5% level, but it is significant when is tested jointly with its corresponded squared term. The coefficient on 'version dummy' is negative and significant, as expected, indicating that increases in congestion do affect the trip value.

The paper presents some of the various analyses we have done based on the estimation results. Marginal WTPs are calculated and graphically represented below. The graphs show how individual's WTP for trips may be influenced by the increase in the level of congestion. In all cases, the marginal WTP curves decrease as congestion level increases. Too many groups at a time affect respondent's trip value.

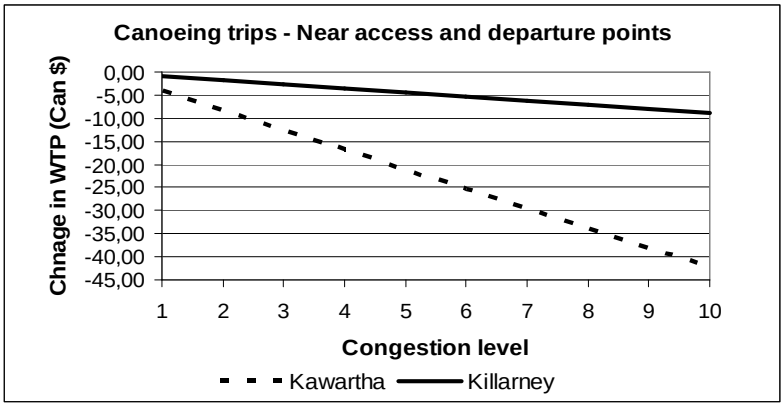


Figure 1. Marginal changes for canoeing trips affected by congestion 'near access and departure points'

Figure 1 shows that canoeing trips are much more affected by changes in congestion level near access and departure points at Kawartha than at Killarney, where people look forward to share their experience with up to 5-6 groups. After they start their activities, they feel annoyed and their trip value decreases. Data indicate that people who took canoeing trips at Kawartha are less affected while they are on water than at Killarney (Figure 2).

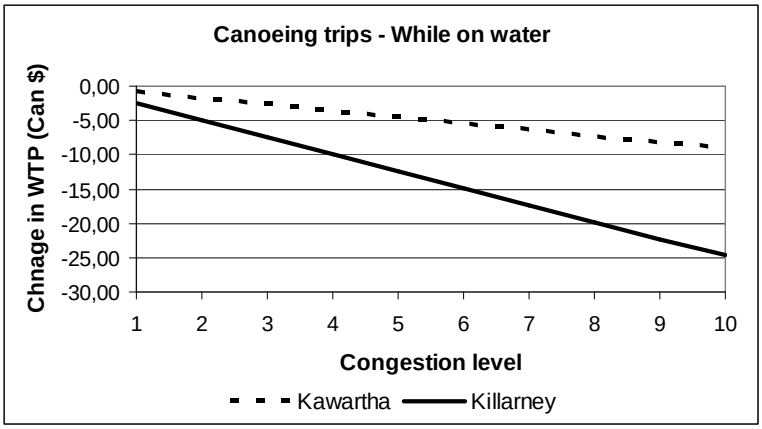


Figure 2. Marginal changes for canoeing trips affected by congestion 'while on water'

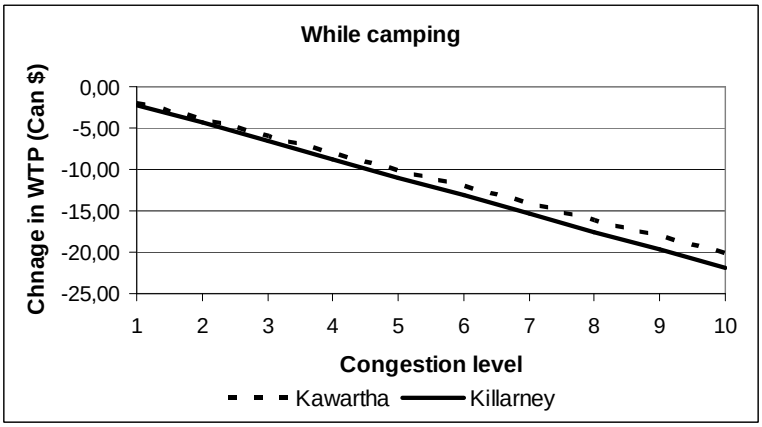


Figure 3. Marginal changes for canoeing trips affected by congestion 'while camping'

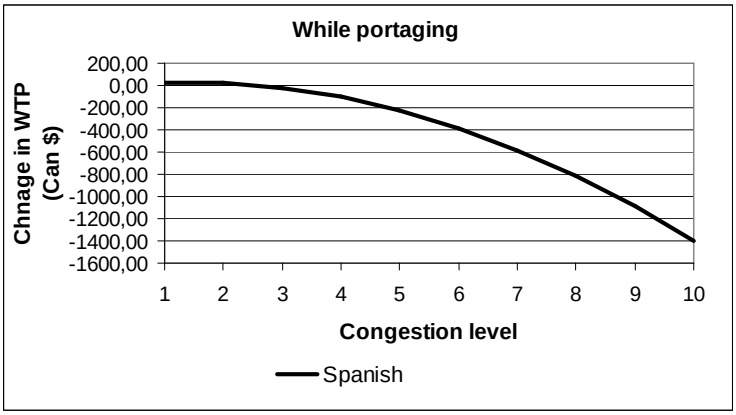


Figure 4. Marginal changes for canoeing trips when affected by congestion 'while portaging'

While people are camping they seem to be affected almost in the same manner at both sites (Figure 3). The curves start decreasing almost in the same point, having a slightly different slope as the level of congestion increases. Spanish River is a unique site, the main activity being white water canoeing. Figure 4 shows how the marginal WTP curve is affected by changes in congestion while portaging.

Conclusions

The overall results suggest that people have different preferences and expectations over site attributes in relation to congestion. It is important to know individuals' preferences in order to take decisions about the development of the area. All the information obtained from this study is important for optimal management of public lands subject to entry quota restrictions. The paper gives evidence that the random effects probit model is an efficient framework in determining the welfare impacts of congestion on each part of the trip.

REFERENCES

Boxall, P., K. Rollins, J. Englin, (2003): Heterogeneous preferences for congestion during a wilderness experience*, Resource and Energy Economics*, vol. 25, issue 2, pp.177-195.

Dumitras, D.E., K. Rollins, S. Balazs, (2005): The Contingent Valuation Method. Case study on a System of Recreation Areas, *Symposium Bulletin Prospects for the Agriculture of the 3rd Millennium*, Cluj-Napoca, vol.62.

Greene, W. H., (2003), *Econometric Analysis,* 5th Edition, Prentice Hall, New Jersey

Hanemann, W.M., (1984): Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses, *American Journal of Agricultural Economics*, vol. 66, n° 2, pp.332-341.

Mitchell, R.C., R.T. Carson, (1993): Using Surveys to Value Public Goods: The Contingent Valuation Method, *Resource for the Future*, Washington, D.C.

Rollins, K., (1997): Wilderness Canoeing in Ontario: Using Cumulative results to Update Dichotomous Choice Contingent valuation Offer Amounts, *Canadian Journal of Agricultural Economics*, vol.45, pp.1-16.

STUDIA UNIVERSITATIS BABEŞ-BOLYAI, OECONOMICA, L, 2, 2005

LE CANAL DIRECT DE LA DISCIPLINE DE MARCHÉ - UN MODÈLE D'INCITATION -

ADRIAN POP*

* *University of Orleans, France. Correspondance: Université d'Orléans, Laboratoire d'Économie d'Orléans, Rue de Blois, BP 6739, 45067 Orléans, Cedex 2, France.*
E-mail : adrian.pop@univ-orleans.fr

ABSTRACT. A straightforward method to enhance market discipline in banking, as it has been discussed in the US since the early 80s, is the *Mandatory Sub-debt Policy*. This policy proposal suggests a *modus operandi* of market discipline based on two main channels. First, a *direct* channel, activated *via* the cost of issuing sub-debt securities, which theoretically should be sensitive to the bank risk profile. Second, an *indirect* channel, effective so long as the supervisor imposes constraints on bank behavior in function of the prices formed in the *secondary* sub-debt market. The article proposes an illustration of the direct channel transmission mechanism based on a simple incentive model of the banking firm.

JEL Classification: G15; G21; G28

Keywords: Banking regulation; Market discipline; Subordinated debt; Incentive model

1. Introduction

L'idée particulièrement attractive d'intégrer la discipline de marché dans la réglementation prudentielle des banques n'est pas entièrement nouvelle. En répondant à une exigence du Congrès américain, la FDIC[1] se proposa déjà en avril 1983 d'évaluer les différentes options de réforme visant à renforcer la discipline de marché dans l'industrie bancaire. Dès lors, la discipline de marché et surtout la dette subordonnée – l'une de ses instruments privilégiés – ont suscité un intérêt croissant dans le milieu professionnel bancaire, mais aussi académique.[2] De manière plus précise, nous considérons dans cet article seulement l'une des nombreuses modalités par lesquelles la discipline de marché pourrait être introduite dans les dispositifs réglementaires, *viz*. la *Politique de Dette Subordonnée* (PDS). Nous entendons par ce terme une exigence réglementaire formelle par laquelle certaines grandes banques sont obligées d'émettre *régulièrement* et de maintenir un montant *minimal* de dette subordonnée sous forme de titres

[1] *Federal Deposit Insurance Corporation* (FDIC) représente l'organisme gouvernemental chargé de l'assurance des dépôts aux États-Unis.

[2] Pour un tour d'horizon et une vue d'ensemble des différentes propositions de valoriser ces concepts, le lecteur peut se référer avec profit aux synthèses exposées dans BGFRS (1999), BGFRS&DT (2000), Evanoff et Wall (2000, p. 67 *et passim*) ou Pop (2003).

homogènes. La dette subordonnée représente une dette bancaire (*i*) *non-couverte* par le système d'assurance des dépôts, (*ii*) *non-garantie* par un droit spécial sur l'actif bancaire, une ligne de crédit etc., (*iii*) ayant une certaine *maturité initiale minimale* et (*iv*) un *statut subordonné* par rapport à toute autre forme de dette contractée par la banque, spécialement les dépôts. La clause de subordination signifie naturellement qu'en cas de faillite de la banque émettrice, les porteurs de titres de dette subordonnée sont prioritaires uniquement par rapport aux actionnaires.

Les propositions de dette subordonnée obligatoire visent à créer une classe distincte d'investisseurs - les créanciers subordonnés - dont les incitations viennent s'aligner naturellement sur celles des autorités de tutelle. Des tâches qui étaient accomplies jusqu'à présent prioritairement par les autorités publiques de régulation pourront être partagées avec les acteurs privés du marché. Ce partage de la fonction de *monitoring* des organisations bancaires pourrait se traduire, par exemple, par des économies considérables en termes de coût de surveillance.

Récemment, deux tentatives ont émergé pour mieux valoriser ces concepts.

Une première tentative est constituée par la nouvelle loi américaine sur la modernisation financière – le ***G****ramm-****L****each-****B****liley (GLB) Financial Modernization Act* – considérée par certains auteurs comme le changement le plus important survenu dans le cadre législatif bancaire américain depuis le ***G****lass-****S****teagall Act* de 1933. Bien qu'il ait suscité de nombreuses controverses au sein du Congrès américain, le *GLB Act* a été finalement signé, en novembre 1999, par le président de l'époque, W.J. Clinton. Cette loi, par son titre 1, section 108, intitulée de manière très suggestive "*Use of subordinated debt to protect financial system and deposit funds from «too big to fail» institutions*", demande une étude concernant la faisabilité et la désirabilité de la PDS. L'étude en question (BGFRS&DT, 2000), élaborée à la suite d'une fructueuse collaboration entre le Conseil des Gouverneurs de la Fed et le Département du Trésor américain, est parue un an après, en décembre 2000. Même si la mise en œuvre *immédiate* d'une telle politique n'est pas recommandée en conclusion de cette étude, sa désirabilité est vivement soutenue:

> «*The Board* [of Governors of the Federal Reserve System] *and the Secretary* [of the Treasury] *believe that existing evidence supports efforts to use subordinated debt as a way to encourage market discipline.* […] *A policy of mandatory subordinated debt issuance may potentially enhance market discipline and safety and soundness. Nonetheless, the Board and the Secretary believe that additional evidence must be gathered before they can support a request for legislative authority to impose a requirement that large insured depository institutions or they holding companies maintain some portion of their*

> *capital in the form of subordinated debt.* […] *the evidence supporting a fairly straightforward mandatory subordinated debt policy with modest objectives is sufficiently strong that continued research and evaluation seem warranted.*»
>
> BGFRS&DT (2000, pp. *iv*, 56-57)

Le troisième pilier du nouvel accord sur les fonds propres, proposé par le Comité de Bâle (BIS, 2004) et intitulé «*discipline de marché*», constitue une seconde tentative. En intégrant la discipline de marché dans la structure de son nouveau triptyque de régulation, le Comité reconnaît, de manière explicite, l'importance des forces de marché dans ses efforts destinés à renforcer la stabilité et la solidité du système bancaire international:

> «[…] *the thinking behind prudential policies experienced a paradigm shift. This has been crystallised in increasing efforts to work with, rather than against, the grain of market forces.* […] *As a result, market discipline has come to play a greater role in ensuring financial stability.* […] *more can and should be done to strengthen market discipline.*»
>
> Andrew Crockett,
> Directeur Général de la BIS

Le Comité de Bâle propose un schéma fondé sur trois principaux piliers se consolidant mutuellement afin de contribuer au renforcement de la sécurité du système financier international. Ces piliers sont: (I) les exigences minimales de fonds propres, (II) le processus de surveillance prudentielle et (III) la discipline de marché. Le troisième pilier concerne l'amélioration du processus de divulgation publique de l'information dans le but de réduire l'opacité inhérente à l'industrie bancaire et de faciliter l'exercice de la discipline par les acteurs privés du marché.

Certes, la transparence représente une condition *nécessaire* pour une meilleure évaluation des conditions financières et du profil de risque des organisations bancaires. Néanmoins, cette condition n'est pas *suffisante*. Il faut également créer une structure appropriée d'incitations, capable d'assurer l'usage de ces informations supplémentaires par les forces de marché. Le nouvel accord de Bâle prévoit la *possibilité* pour les banques, mais non l'*obligation*, d'émettre des titres de dette subordonnée.

L'approche par la dette subordonnée constitue à notre avis une vision alternative plus précise de la discipline de marché. Une PDS explicite propose une source concrète de discipline, ainsi que des canaux réalistes de transmission dans l'industrie bancaire. Tout d'abord, le canal *direct* se manifeste *via* le coût d'émission qui théoriquement est sensible à un changement du profil de risque. Ainsi, la tarification appropriée de la dette bancaire sur le marché *primaire* pourrait dissuader *ex-ante* les organisations bancaires d'adopter

des stratégies aventureuses en matière de risque. Deuxièmement, le canal *indirect* est effectif dans la mesure où le superviseur et les autres contreparties privées observent les prix formés sur le marché *secondaire* et infèrent des signaux pertinents quant au risque de défaut des banques émettrices. Pour que le canal indirect soit effectif, une condition *nécessaire*, même si *non suffisante*, est que les prix formés sur le marché secondaire soient sensibles au risque bancaire.[3]

L'objectif de cet article est de présenter un modèle intuitif, inspiré des travaux menés par les économistes du Conseil des gouverneurs de la Fed, capable d'illustrer de manière très simple la transmission de la discipline de marché *via* le canal *direct*. Nous commençons par esquisser la logique sous-jacente au canal *direct* de la discipline de marché, ensuite nous présentons les hypothèses du modèle, les parties impliquées et les principaux résultats.

2. Quelques intuitions de base

En théorie, les porteurs de titres de dette subordonnée sont considérés comme les acteurs du marché les plus adaptés pour surveiller les profils de risque des banques. En effet, dans des conditions normales de solvabilité, leurs intérêts sont alignés sur ceux du régulateur et fortement opposés à ceux des actionnaires.[4] Pour que le marché de dette subordonnée exerce son rôle disciplinaire, un certain nombre de conditions favorables devraient préalablement être réunies. Tout d'abord, les investisseurs doivent impérativement être *capables* d'évaluer correctement les organisations bancaires. Cela ne représente pas un défi trivial. Selon la théorie moderne des intermédiaires financiers, la banque crée de la valeur en mettant en place des procédures de sélection et de contrôle des emprunteurs. À la suite de ce processus de *monitoring*, la banque acquiert des informations privées relatives à la qualité de son portefeuille de crédits, ce qui *a priori* rend l'actif bancaire 'opaque' et donc difficile à évaluer de l'extérieur.

Deuxièmement, la divulgation publique d'une *information* financière régulière et pertinente représente une autre condition indispensable d'une évaluation efficace des établissements bancaires. Le Comité de Bâle confirme

[3] Pour une revue de la littérature concernant le contenu informationnel des prix formés sur les marchés de la dette bancaire dans les pays développés, voir Pop (2004). Nous proposons une analyse empirique détaillée de l'efficience informationnelle du marché européen de la dette subordonnée bancaire dans Pop (2005c). Pour un élargissement de perspective à l'échelle internationale et un examen plus approfondi de l'articulation entre (*i*) le renforcement de la discipline de marché et (*ii*) la préservation d'une concurrence équitable entre les grandes banques internationales ('*level the playing field*'), voir Pop (2005b).

[4] Les intérêts des créanciers privés subordonnés sont similaires à ceux du régulateur, mais non identiques. Nous proposons dans un autre article (Pop, 2005a) une analyse formelle des incitations caractérisant les détenteurs de titres de dette subordonnée dans un cadre unifié et cohérent: la théorie des options.

l'importance de cette condition. Les normes de communication financière détaillées dans le 3e Pilier du nouvel accord aideront sans conteste à réduire l'opacité inhérente aux asymétries d'informations existant entre la banque et ses investisseurs.

Enfin, les *incitations* des porteurs de titres de dette jouent un rôle crucial dans la théorie de la discipline de marché. En particulier, si les créanciers sont convaincus d'être protégés *de facto* par un 'filet de sécurité' étendu au-delà de ses contours définis *ex-ante*, leurs incitations à surveiller seront irrémédiablement compromises. Dans cet esprit, le Comité de Bâle reconnaît que «*la communication financière peut avoir une efficacité limitée en termes de discipline de marché si les acteurs du marché se sentent protégés par un 'filet de sécurité' officiel*» (*apud* BIS, 1998, p. 7, §27).

Les économistes du Conseil des gouverneurs de la Fed ont proposé récemment une définition alternative du canal *direct* de la discipline de marché, légèrement différente de celle que nous avons proposée dans l'introduction:

> *«Le canal direct de la discipline de marché est effectif si le niveau de risque choisi par une banque lorsqu'elle anticipe une émission de dette subordonnée est inférieur au niveau de risque qu'elle aurait choisi si une telle émission n'avait pas eu lieu et le portefeuille d'actifs qu'elle aurait financé avait été ainsi moins important».*
>
> Covitz et al. (2000, p. 29)

Supposons par la suite que les banques ne substituent pas de la dette subordonnée aux fonds propres. Cette hypothèse n'est pas si héroïque, étant entendu que la plupart des grandes banques ciblent en effet des ratios de capital au sens strict (i.e. *Tier 1*). Dans ces conditions, l'essentiel de la discipline de marché consiste en la capacité des créanciers subordonnés à sanctionner les prises de risque excessives. Par conséquent, tout au long de cet article nous supposons que l'alternative à l'émission de dette subordonnée est de ne rien émettre et de financer ainsi un portefeuille d'actifs moins important.

Il convient de souligner que la définition énoncée précédemment ne compare pas:

-le niveau de risque que la banque aurait choisi lorsqu'elle anticipait une émission de dette subordonnée

avec

-le niveau de risque choisi dans le cas où, au lieu d'émettre de la dette subordonnée, la banque aurait financé son actif à l'aide d'un instrument de dette *insensible* au risque (e.g. dépôts assurés).

En effet, cette comparaison aurait été satisfaite dans tous les cas où les *spread*s de la dette subordonnée aurait été sensibles au risque (le degré de sensibilité n'aurait pas été très important dans ce cas!). Bien évidemment, dans la réalité, le canal *direct* de la discipline de marché n'est

pas efficace si la sensibilité des *spreads* au risque bancaire est trop faible. En effet, une sensibilité relativement faible permet aux actionnaires de bénéficier d'un accroissement du risque dans la mesure où la prime de risque anticipée exigée par les créanciers est positive, mais réduite.

De plus, une dette subordonnée faiblement sensible au risque impose une discipline *directe* moins efficace même si on la compare avec les dépôts. En effet, en période de difficultés, les déposants peuvent retirer simplement leurs fonds. Cela constitue alors un mécanisme disciplinaire extrême, bien plus «efficace», i.e. susceptible de freiner *ex-ante* les prises de risque excessives (*cf*. Carletti, 1999).

En revanche, la définition proposée par Covitz *et alii* est satisfaite uniquement lorsque le prix auquel le marché de dette subordonnée tarife le risque est supérieur à celui exigé par un créancier *rationnel* et *neutre au risque* sous l'hypothèse d'*absence de garanties gouvernementales* (explicites ou implicites).

3. Parties impliquées, hypothèses et principaux résultats

Les parties impliquées sont:

-une banque

et

-ses potentiels créanciers obligataires subordonnés, supposés *neutres au risque*.

La banque détient au passif une quantité fixe de dépôts D et fonds propres K, en montant suffisant pour satisfaire au minimum réglementaire:

$$k = K/s,$$

où k représente le ratio minimum de capital (supposé constant) et s l'actif total.

Pour simplifier les choses, les déposants sont rémunérés au taux sans risque R_f, normalisé à zéro. De même, les primes d'assurance de dépôts sont normalisées, elles aussi, à zéro. La banque a accès à un ensemble P de portefeuilles d'actifs. Chaque portefeuille comprend deux dimensions: une probabilité de défaut ρ et un volume s,

$$P = (\rho, s); \rho \in [0, 1]; s \in [D + K, s^{max}]$$

La maturité de la dette subordonnée et celle des portefeuilles sont supposées identiques. De plus, la banque dispose de l'option d'accroître la taille de son portefeuille en émettant de la dette subordonnée, rémunérée au taux 'brut' r^{DS}. Les trois situations suivantes sont possibles:

a. Si la banque décide d'émettre de la dette subordonnée **et** ne fait pas faillite, alors le portefeuille génère un taux de rentabilité[5] $[h + \rho]/[1 - \rho]$, où $h > 0$ afin d'éliminer les situations triviales où le risque bancaire est si faible

[5] En effet, dans un monde *neutre au risque*, si l'on note h la rentabilité espérée du portefeuille en l'absence de défaut, la relation suivante est vérifiée:
[Proba. de survie]·[1 + le taux de rentabilité recherché] = [1 + h]
De plus, la rentabilité totale du portefeuille est égale à $s[1 + h]$.

que la banque est rationnée sur le marché primaire. De plus, $h'(\rho)>0$ car les portefeuilles plus risqués proposent une rentabilité espérée plus élevée. Dans cette première situation les déposants reçoivent le taux sans risque $R_f = 0$, tandis que les créanciers subordonnés sont rémunérés au taux r^{DS}.

b. Si la banque émet de la dette subordonnée **et** fait faillite, on suppose que la valeur du portefeuille se réduit à néant. De plus, il existe des coûts espérés subis par les actionnaires à la suite d'un accroissement du risque. Ces coûts (ou désutilités) sont *idiosyncratiques*, i.e. spécifiques à chaque banque *i*. Leur valeur espérée par unité de fonds propres est notée $C_i(\rho)$. La fonction C_i est de classe **C²** avec:

$$C_i'(\rho) > 0,\ C_i''(\rho) > 0 \text{ et } \lim_{\rho \to 1} C_i'(\rho) = +\infty .$$

c. Si la banque décide de ne pas émettre des titres de dette subordonnée, alors elle financera un portefeuille de crédits plus restreint, i.e. de taille $s = D + K$ au lieu de $s = D + K + \boldsymbol{DS}$, conformément à la définition énoncée à la section précédente.

L'hypothèse selon laquelle les dépôts sont constants est forte, mais reflète le fait que «*les banques ne considèrent pas les dépôts, disponibles sur demande au gré des déposants, comme des substituts parfaits à la dette sensible au risque*» (*cf*. Covitz *et al*., 2000). Si c'était le cas, les banques relativement risquées seraient toujours dissuadées d'émettre de la dette subordonnée. En effet, toutes les banques (sans discrimination) proposent un prix de marché pour attirer les dépôts, alors que les banques risquées offrent des rendements plus élevés afin de placer leurs titres de dette subordonnée.[6]

Considérons dans ce qui suit les deux cas de figure suivants:

Cas I. Le modèle est résolu sous l'hypothèse que les créanciers subordonnés sont *neutres au risque* et *entièrement compensés* pour le risque encouru. Notons que, dans ces conditions, le taux d'intérêt ajusté pour le risque (relatif à la dette subordonnée) doit être nécessairement égal à R_f:

$$[1 - \rho]\cdot[1 + r^{DS}] = 1 + R_f = 1,$$

ce qui implique $r^{DS} = \rho/[1 - \rho]$.

La proposition suivante a lieu:

Proposition n.1. *Sous l'hypothèse de neutralité au risque, le fait d'anticiper des paiements d'intérêt plus importants aux créanciers subordonnés ne suffit pas à discipliner effectivement les banques.*

[6] Cette hypothèse, selon laquelle les dépôts et la dette subordonnée ne sont pas des substituts parfaits, est empiriquement validée par Covitz *et al* (2004). En effet, ces auteurs ne parviennent pas à prouver que les banques relativement risquées sont toujours plus enclines à émettre de la dette subordonnée.

Démonstration

Le programme de la banque, en l'absence de conflit d'intérêt entre les dirigeants et les actionnaires, est de maximiser la rentabilité des fonds propres:

$$ROE = \frac{return}{equity} = \frac{1}{K}\left\{(1-\rho)\left[s\left(1+\frac{h+\rho}{1-\rho}\right) - D(1+0) - DS\left(1+\frac{\rho}{1-\rho}\right)\right] + \rho\cdot 0\right\} - C_i(\rho)$$

taux de rentabilité du portefeuille — R_f — r^{DS} — par unité de fonds propres

Après avoir réarrangé les termes et tenu compte de la relation $k = K/s$ on obtient:

$$ROE = (1-\rho)\left[\frac{1}{k}\left(1+\frac{h+\rho}{1-\rho}\right) - \frac{D}{k\cdot s} - \frac{DS}{k\cdot s}\left(1+\frac{\rho}{1-\rho}\right)\right] - C_i(\rho)$$

Les créanciers privés subordonnés (neutres au risque) exigent un taux de rentabilité r^{DS} qui ne dépend pas du niveau de risque choisit à l'équilibre par la banque: le canal direct de la discipline de marché n'est pas effectif. Pour s'en persuader, il suffit de reprendre la définition de ce taux, $r^{DS} = \rho/[1-\rho]$.

Le programme de la banque s'écrit dans ces conditions de la manière suivante:

$$\max_{\rho} ROE = (1-\rho)\cdot\frac{1}{k}\cdot\frac{1+h(\rho)}{1-\rho} - (1-\rho)\cdot\frac{D}{k\cdot s} - (1-\rho)\cdot\frac{DS}{k\cdot s}\cdot\frac{1}{1-\rho} - C_i(\rho),$$

où *ROE* et donnée finalement par:

$$ROE = \frac{1}{k}\cdot[1+h(\rho)] - (1-\rho)\cdot\frac{D}{k\cdot s} - \frac{DS}{k\cdot s} - C_i(\rho).$$

À l'optimum,

$$\frac{\partial ROE}{\partial \rho} = 0 \quad \Leftrightarrow \quad \frac{1}{k}h'(\rho) + \frac{D}{k\cdot s} = C_i'(\rho).$$

Bien évidemment, tenant compte de l'expression de cette condition du premier ordre, la probabilité de défaut choisie par la banque à l'équilibre ρ^* ne dépend ni du montant de dette subordonnée, ni du taux promis aux créanciers, *Q.E.D.*

En résumé, le simple fait d'émettre des titres de dette subordonnée sur un marché où les investisseurs sont *neutres au risque* est insuffisant pour activer le canal direct de la discipline de marché. En outre, la décision d'émission de la banque s'avère *insensible* au risque choisi.

Proposition n.2. *Lorsque le canal direct de la discipline de marché n'est pas effectif, il n'existe pas de relation entre la décision d'émission et la probabilité de défaut choisie par la banque à l'équilibre, ρ^*.*

Démonstration

Une banque a intérêt à émettre de la dette subordonnée lorsque le taux de rentabilité du portefeuille est supérieur au taux promis aux créanciers subordonnés:

$$\frac{h+\rho}{1-\rho} > \frac{\rho}{1-\rho} \quad \Leftrightarrow \quad h > 0, \quad \forall \rho \in [0,1] .$$

Bien évidemment, cette décision est indépendante du niveau de risque choisi par la banque à l'équilibre ρ^*, *Q.E.D.*

Cas II. Revenons maintenant au cas plus intéressant où le marché de la dette subordonnée tarife le risque de manière plus sévère que les créanciers neutres au risque étudiés précédemment.

Par conséquent, on suppose que le marché tarife le risque de telle sorte (i.e. *si sévèrement*) que la définition du canal direct de la discipline de marché énoncée initialement soit satisfaite. Une sensibilité plus accrue au risque peut être due, par exemple, à l'aversion pour le risque des investisseurs. Au lieu d'introduire des effets de réputations ou une fonction d'utilité concave pour caractériser le comportement des créanciers, on suppose tout simplement que le marché exige une *prime de risque*, notée $\varphi(\rho)$, avec $\varphi(\rho) > 0, \forall \rho$. Cette prime de risque est incorporée dans le taux d'intérêt de la dette subordonnée, qui devient désormais

$$r^{DS} = \varphi(\rho) + \frac{\rho}{1-\rho} .$$

Afin de s'assurer que le canal direct de la dette subordonnée est effectif quelle que soit la forme de la fonction de coût $C_i(\rho)$, on ajoute la restriction suivante: *la rentabilité espérée des créanciers subordonnés est fonction croissante du risque pris par la banque*. Cette rentabilité est donnée par

$$(1-\rho)\cdot(1+r^{DS}) + \rho \cdot 0 = (1-\rho)\cdot\left[\varphi(\rho) + \frac{1}{1-\rho}\right] = 1 + (1-\rho)\cdot\varphi(\rho)$$

Comme la fonction $f(.)$ définie par $f(\rho) = 1 + (1 - \rho)\varphi(\rho)$ doit être croissante en ρ, on obtient la relation suivante:

$$f'(\rho) = -\varphi(\rho) + (1-\rho)\cdot\varphi'(\rho) > 0 ,$$

ou encore

$$(1-\rho)\cdot\varphi'(\rho) - \varphi(\rho) > 0, \forall \rho .$$

Dans ce deuxième cas, le risque pris par la banque (*conditionné* par l'émission de dette subordonnée) devrait être inférieur au niveau choisi si celle-ci s'abstenait d'émettre de la dette subordonnée et finançait par conséquent un

portefeuille de crédits moins important (*i.e.* la définition proposée au début est satisfaite!).

L'introduction d'une prime de risque positive $\varphi(\rho)$ implique l'existence d'une corrélation inverse entre le profil de risque de la banque et sa décision d'émettre de la dette subordonnée. Cette intuition peut être exprimée à l'aide de la proposition suivante:

Proposition n.3. *Étant donné un ensemble quelconque de N banques* $\{C_1, C_2, \ldots, C_N\}$, *si* $(1-\rho)\cdot\varphi'(\rho)-\varphi(\rho)>0$, *alors les banques les moins risquées sont plus probablement celles qui émettent des titres de dette subordonnée. De plus,*

A. *À chaque fois qu'une banque (ayant choisi un certain niveau d'équilibre du risque ρ*) émet de la dette subordonnée, des banques caractérisées par des niveaux de risque* plus faibles *émettent également de la dette subordonnée;*
B. *À chaque fois qu'une banque (ayant choisi un certain niveau d'équilibre du risque ρ*) émet de la dette subordonnée, il se peut que des banques caractérisées par des niveaux de risque* plus élevés *émettent ou non de la dette subordonnée.*

*Démonstration (*per reductio ad absurdum*)*

Si on suppose que le canal direct de la discipline de marché est effectif et que les banques risquées ne sont pas *moins enclines* à émettre de la dette subordonnée, la proposition n.3 est-elle vérifiée? Autrement dit, est-il possible de repérer en même temps une banque *i* à risque *élevé* qui *émet* de la dette subordonnée et une autre banque *j* à risque *faible* qui *n*'en *émet pas*?

Le contraire de cette troisième proposition implique le respect simultané des trois conditions suivantes:

1. $\rho_i^* > \rho_j^*$
2. $h > \varphi(\rho_i^*)\cdot(1-\rho_i^*)$
3. $h < \varphi(\rho_j^*)\cdot(1-\rho_j^*)$

$\Rightarrow \varphi(\rho_i^*)\cdot(1-\rho_i^*) < \varphi(\rho_j^*)\cdot(1-\rho_j^*)$

Les deux dernières conditions impliquent évidemment $\varphi(\rho_i^*)\cdot(1-\rho_i^*) < \varphi(\rho_j^*)\cdot(1-\rho_j^*)$. D'un autre côté, si l'on note $Z(\rho) = \varphi(\rho)(1-\rho)$ et on remarque que la fonction Z est croissante en ρ (en effet $\partial Z/\partial\rho = (1-\rho)\,\varphi'(\rho) - \varphi(\rho) > 0$, $\forall\ \rho$), alors $\varphi(\rho_i^*)\cdot(1-\rho_i^*) < \varphi(\rho_j^*)\cdot(1-\rho_j^*)$ implique que $\rho_i^* < \rho_j^*$. Ce résultat contredit bien évidemment la première condition, *Q.E.D.*

4. Remarques finales

Les propositions n.1,2&3 suggèrent la façon dont on pourrait inférer le degré de discipline imposée par le marché de dette subordonnée à partir de l'impact du risque choisi en interne par la banque (ρ^*) sur sa décision d'émission. Les principales implications en termes d'efficience du canal *direct* de la discipline de marché peuvent être résumées de la manière suivante:

- Si les banques relativement risquées sont *plus* enclines à émettre de la dette subordonnée, alors les prix de marché ne sont pas suffisamment sensibles pour compenser *entièrement* les investisseurs pour le risque encouru;
- Si les banques relativement risquées ne sont *ni plus ni moins* sujettes à émettre de la dette subordonnée, alors il se peut qu'il existe une certaine sensibilité des créanciers subordonnés au risque, mais cette sensibilité n'est pas suffisante pour activer le canal *direct* de la discipline de marché;
- Enfin, si les banques relativement risquées sont *moins* susceptibles d'émettre de la dette subordonnée, le canal direct de la discipline de marché est réellement effectif.

Nous avons essayé de tester sur des données réelles l'intuition sous-jacente au modèle d'incitation qui vient d'être présenté. L'objectif était donc de fournir un support empirique adéquat aux arguments théoriques développés dans cet article.[7]

L'échantillon sur lequel s'appuie notre analyse empirique comprend plus de 500 grandes banques européennes, pour lesquelles les rapports annuels de 1996 à 2003 étaient disponibles dans *Fitch-IBCA BankScope Database*. La stratégie empirique consiste à comparer, à l'aide d'une analyse *bi-variée* (fondée sur le test de la somme des rangs de Wilcoxon), les caractéristiques de deux populations distinctes d'organisations bancaires:

- les banques caractérisées par des ratios *dette subordonnée*/*actif total* relativement réduits (échantillon **A**)

et

- les banques détenant des montants très importants de dette subordonnée dans leurs bilans (échantillon **B**).

Plus précisément, les comparaisons ont été réalisées à partir de trois dimensions essentielles de l'activité bancaire, à savoir la *profitabilité*, la *capitalisation* et le *profil de risque*. Si la décision d'émettre de la dette subordonnée est *endogène* – i.e. conforme au modèle d'incitation décrit dans cet article –, alors on devrait déceler (*i*) une performance meilleure, (*ii*) des ratios d'adéquation du capital plus élevés et (*iii*) un profil de risque moins dégradé, *ceteris paribus*, pour les banques faisant partie de la deuxième catégorie.

En ce qui concerne la première dimension considérée, nous révélons en effet que les banques appartenant à l'échantillon **B** sont significativement plus profitables en moyenne que leurs concurrentes. Quant au degré de capitalisation, ce sont toujours les banques de l'échantillon **B** qui affichent des ratios d'adéquation des fonds propres (définis *lato sensu*) plus importants

[7] Pour une présentation *in extenso* des principaux résultats empiriques obtenus, voir Pop (2005d).

en moyenne que ceux de leurs homologues incluses dans l'échantillon **A**. Enfin, les résultats empiriques relatifs à la dernière dimension prise en compte sont malheureusement moins clairs. D'une part, le profil de risque s'avère significativement moins dégradé dans le cas des banques comprises dans l'échantillon **A**. En d'autres termes, les banques détenant des montants très importants de dette subordonnée dans leurs bilans présentent des encours de crédit douteux relativement plus élevés en moyenne. Néanmoins, ces dernières banques rapportent en même temps des réserves pour le risque de crédit plus appréciables, destinées notamment à absorber en partie les pertes futures sur leurs portefeuilles de crédit. Au bout du compte, ce dernier résultat pourrait très bien être interprété comme étant l'indice d'un comportement prudent en matière de gestion des risques.

En conclusion, les émissions *volontaires* de dette subordonnée ne sont pas *neutres*! Elles ont un impact favorable sur la profitabilité, la capitalisation et le degré de couverture du risque de crédit. Ce résultat original est cohérent avec les principes défendus par les théoriciens de la discipline de marché en général et par les apologistes des propositions de dette subordonnée obligatoire en particulier.

REFERENCES BIBLIOGRAPHIQUES

Bank for International Settlements (BIS) (1998): *Enhancing Bank Transparency*, Basel Committee on Banking Supervision, september

Bank for International Settlements (BIS) (2004): *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*, Basel Committee on Banking Supervision, june;

Board of Governors of the Federal Reserve System (BGFRS) (1999): Using Subordinated Debt as an Instrument of Market Discipline, *Staff Study*, No. 172, december;

Board of Governors of the Federal Reserve System and the Secretary of the U.S. Department of Treasury (BGFRS&DT) (2000): The Feasibility and Desirability of Mandatory Subordinated Debt, Report submitted to the Congress pursuant to the section 108 of the *GLB Act*, december;

Carletti, E. (1999): Bank Moral Hazard and Market Discipline, Financial Markets Group, London School of Economics, *Discussion Paper*, No. dp326, may;

Covitz, D.M., Hancock D., Kwast M.L., (2000): Mandatory Subordinated Debt: Would Banks Face More Market Discipline?, Board of Governors of the Federal Reserve System, *mimeo*, june;

Covitz, D.M., Hancock D., Kwast M.L., (2004): Market Discipline in Banking Reconsidered: the Roles of Funding Manager Decisions and Deposit Insurance Reform, Board of Governors of the Federal Reserve System, *Finance and Economics Discussion Series*, No. 2004-53, october;

Evanoff, D., Wall L.D. (2000): Subordinated Debt and Bank Capital Reform, Federal Reserve Bank of Chicago, *Working Paper*, No. 2000-07, august;

Pop, A. (2003): Dette subordonnée, discipline de marché et réforme réglementaire, *Revue d'Économie Financière*, vol. 71, pp. 261-276, juillet;

Pop, A. (2004): Beyond the 3rd Pillar of Basel 2: Taking Bond Market Signals Seriously, communication au LIIIe *Congrès annuel de l'AFSE*, Paris – La Sorbonne, septembre;

Pop, A. (2005a): Mandatory *Sub-Debt Policy* as an Alternative to the Third Pillar of Basel II: Is It Feasible?, *Louvain Economic Review*, vol. 71, no. 2, pp. 193-221, 2nd quarter;

Pop, A. (2005b): "Market Discipline in International Banking Regulation: Keeping the Playing Field Level", *in Proceedings of the 12th Global Finance Conference*, C/O School of Business Studies, Trinity College, Dublin, august;

Pop, A. (2005c): The Indirect Channel of Market Discipline in Banking: Testing the '*Sine Qua Non*' Hypothesis, under revision for *Annals of Economics and Statistics*;

Pop, A. (2005d): Why a *Sub-Debt Policy* Must Be Mandatory? A *Pro Domo* Apology, University of Orléans, *mimeo*, february.

STUDIA UNIVERSITATIS BABEŞ-BOLYAI, OECONOMICA, L, 2, 2005

CHALLENGES IN IMPLEMENTING INFORMATION TECHNOLOGY STANDARDS – RESULTS FROM A CASE STUDY

RALUCA BUNDUCHI*

**.University of Aberdeen, United Kingdom,*
Business School, Edward Wright Building,
Aberdeen, AB24 3QY, r.bunduchi@abdn.ac.uk

ABSTRACT. Standards in the area of information and communication technologies are claimed to bring a number of benefits for the organisations that adopt them. However, there is little research that addresses the factors that shape the implementation and use of standards at the level of individual organisations. This paper discusses the implementation of standards for clinical data messaging in the health sector in Scotland. The results of the case study show that technical factors play only a minor role in shaping the implementation process The characteristics of the organisational culture, and the fragmented structure of health organisations are found to hamper the implementation of standardised information systems. Local and generic knowledge at the level of the local implementation teams appear to facilitate the implementation and use of the standards.

JEL Classification: I1, O33

Keywords: information systems, standards, health

1.Introduction

Before the emergence of the Internet, Electronic Data Interchange for Administration, Commerce and Transport (EDIFACT) and was the dominant standard in the area of Electronic Data Interchange (EDI). However, the EDI systems based on EDIFACT standards were characterised by high implementation and maintenance costs, inflexibility and complexity in use which made them expensive and difficult to use (Medjahed & Benatallah & Bouguettya & Nhu & Elmagarmid, 2003; Turban & Lee & King & Chung, 2000). With the advent of the Internet, the major efforts in the standardisation of information exchange either in the private or public sector concentrate around the eXtensible Markup Language (XML) related standards. Based on the Standard Generalized Mark-up Language (SGML) meta language, XML standards allow the separation of content from presentation, enable users-defined tags, and are based on an open development process under the auspices of the World Wide Web Consortium (W3C). Such features are claimed to enable a cheap and flexible approach to the standardisation of

information exchange, with XML fast emerging as the dominant standard for data representation and exchange.

This study explores the implementation of XML based standards for information exchange in the area of health informatics. The paper discusses a case study of the standardisation of clinical data messaging in the health sector in Scotland. The analysis focuses on implementation side of the standardisation process and aims to identify and explain the factors that facilitate and/or inhibit the implementation and use of the standardised systems.

2.Theoretical background

Most of the studies of standardisation in the area of Information and Communication Technologies (ICTs) focus on the development side rather than the implementation and use of the standards. Some of the earliest accounts of standardisation attempt to understand the relation between the economic conditions and the evolution of standards by explaining the emergence of standards in terms of path dependencies and network externalities (David, 1985; Farrell & Salloner, 1985). Drawing from a large range of theoretical perspectives, the literature on standardisation has inquired into the factors that affect the choice of firms between various modes of standardisation (Austin & Milner, 2001; Belleflame, 2002; Farrell & Saloner, 1988), has analysed the different forms of standard settings (Bensen & Farrell, 1991; David & Shurmer, 1996, Hawkins, 1999; Weiss, 1993) and has highlighted the politics and interests involved in standardisation committees that shape standard development (Graham & Spinardi & Williams & Webster, 1995; Schmidt & Werle, 1998).

The use of standards in the area of ICTs is documented in the existing literature as leading to a number of positive implications for the organisations adopting the standards. The adoption of standards is claimed to reduce entry barriers arising in production of networks products and services, to diminish the asymmetries in the costs of access to information, to reduce the scope for a supplier to impose switching costs upon its customers, to create economies of scale and networks externalities, and to reduce risks and transaction costs (Antonelli, 1994; David & Steinmueller, 1994; Damsgaard & Truex, 2000). However, as Damsgaard and Truex (2000) found in their analysis of EDI standards, such benefits are rarely achieved at an organisational level. The ability of standards to achieve their benefits depends not only on the process through which they are developed (see for example David and Steinmueller, 1994), but also on the way in which they are implemented in organisations. Therefore, to explain the success or failure of a standard, one has to understand the factors that shape the implementation and use of that standard within organisations.

In the mainstream literature, there is only a limited number of studies that addressed the factors that influence the way organisations adopt, implement and use a particular standard. For example, Chen (2003) identifies

a number of factors that influence the adoption of web service standards in organisations including company size, organisational culture and the organisation's specific IT architecture. Moreton, Sloane and Simon (1995) and Sloane (2000) also suggest that there are a number of factors that influence the change in the standards used in organisations such as costs, training requirements and compatibility with existing systems. However, such studies regarding standard adoption are generic, and do not provide an in depth analysis of the factors that shape the implementation of a particular standard in an organisation based on empirical research.

This study provides an account of the implementation of a standard for information exchange in the health sector in Scotland in order to explain the factors that influence the standard implementation process. The empirical research is based on a single instrumental case study (Stake, 1994) for two reasons. Such a research design allowed (1) to gain in depth insights into the standardisation process (Stake, 1994), and (2) to understand the dynamics present within a single settings (Eisenhardt, 1998). Rather than attempting to identify general patterns that apply across a multitude of contexts and technologies, this research aims to understand the dynamics involved in the standard implementation process within a particular context: how does it occur and what are the factors that shape it.

Semi structured interviews were used as the main method for data collection, complemented with secondary data sources such as the analysis of internal documentation, policy documents and other information publicly available. In total, 14 people were interviewed over a period of 6 months including 4 policy makers (PM), 7 respondents involved in developing and coordinating the implementation of the clinical data standards and the information systems products (NHS IT), as well as 3 people involved in the implementation of the information systems products in two of the trust board in Scotland (Local TB). To protect the identity of the interviewees, the respondents in this paper will be identified only as belonging to one of these 3 categories. Following Stake (1994) recommendation, the analysis of the case was based on making a detailed description of the case and its settings and followed the iterative model proposed by Miles and Huberman (1994): after data collection, the data was reduced through open coding, followed by displaying the data using descriptive and explanatory data displays, and finally conclusion drawing and verification.

3. Case study

3.1. Context

In 2000, the National Health Service (NHS) in Scotland announced radical changes in its approach to information technology. The aim of the new strategy was to take advantage of the promised benefits of the new

Internet technologies by developing an integrated electronic patient care system, where the IT systems would be designed and delivered around the needs of the patients, and not around the NHS institutions. Such an integrated electronic system for patient care was seen as a significant effort to increase the efficiency of the health service.

NHS Scotland followed an incremental approach to the development of the integrated electronic patient care record. The electronic patient records are kept in regional repositories at the level of the health boards. The patient record repository called SCI (Scottish Care Information) store is developed in partnership between NHS Scotland and a system vendor - SemaSchlumberger. SCI Store is located at the level of the health boards and captures data from the existing legacy systems within that particular health board. Other 6 additional SCI products facilitate the exchange between SCI Store and the existing systems within and between the local NHS organisation:

- SCI Integration - middleware interfacing tool (E-biz 2000) which extracts data from the legacy system and place it in the repository.
- SCI Gateway - a national communication portal that handles the clinical messages (e.g. referral, discharge letters, and appointments results) between primary and secondary care.
- SCI Discharge – generates the discharge letter based on the data captured from hospital IT systems.
- SCI Outpatient – a scheduling system for the management of appointments, bookings and waiting lists.
- SCI Clinical - an operational system to support hospital clinicians, but not widely developed.
- SCI prescription – not yet developed at the time of the study.

SCI products are presented in the figure below.

Only two of the SCI products, SCI Store and the communication portal part of the SCI Gateway are mandatory national products. For the rest of the products, the local boards have the liberty to choose between SCI and commercial products. The exchange of information between the various components in the integrated patient care system requires the existence of common standards for clinical data messaging throughout the NHS. Such standards have to specify the structure and content of the clinical data exchanged between the various IT systems in the NHS. The standards cover not only the network protocols (e.g. HTTM or SMPT) and the XML messaging (e.g. SOAP),

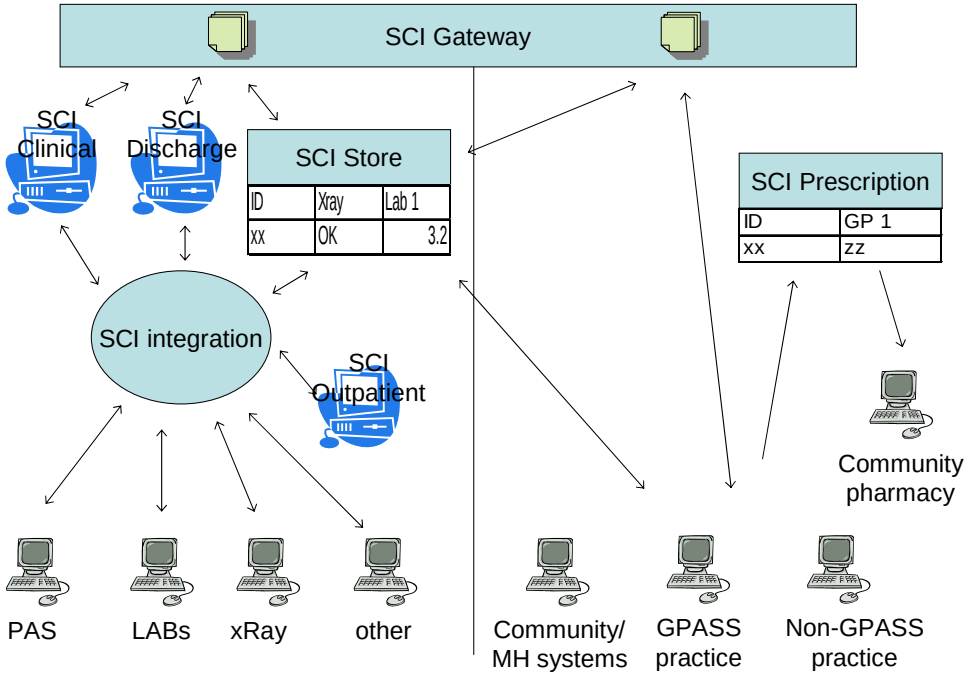


Figure 1. SCI products and their role in the NHS Scotland

but also the information flows between the various components of the NHS system (e.g. referral and discharge letter, clinical letter, appointment booking, lab results, test ordering). Such standards embedded in the SCI products prescribe the content and structure of the data exchanged between clinicians from the different parts of the healthcare system, when such data is to be exchanged, and in what manner. Therefore, the standards affect not only how the IT systems/products will be designed and will interoperate, but also the work of nurses, junior doctors, consultant, lab people and GPs who exchange such data.

Following the general approach to IT provision, the development of clinical data messaging in Scotland is done incrementally. Rather than investing

the time and money in participating in an international standardisation effort (e.g. CEN or HL7), Scotland preferred to develop its own standards based on XML, and using the existing work from other standardisation initiatives in healthcare such as CEN ENV13606, HL7, NHS data dictionary, SNOMED and SIGN guidelines. Development is approached in a pragmatic manner, where standard development runs in parallel with system development and implementation. Such an approach allows for a constant refinement of the standard as a result of successive implementations within the system. The emphasis is on developing a "*good enough standard*" that "*works fast*" and "*might be changed along the way*". Speed of development and implementation and not the technical quality of the standard is the focus of the standardisation process.

As mentioned above, the standards embedded in the IT products affect significantly the working practices in the health sector. Such an incremental approach to standard and system development allows the time for a gradual change in the working practices in the health sector, as well as for an incremental replacement of the existing IT legacy systems that have to interface with the new products. At the same time, containing standard development within NHS Scotland allows for a speedy process and easier resolution of conflicts, while at the same time saves the time and financial effort of getting involved in an international standard organisation. Such a rapid and relatively inexpensive standardisation process allows NHS Scotland to comply with the strong political commitment for a rapid system development and to save financial resources allocated for IT provision.

This case study focuses on the implementation of the two mandatory products, SCI Gateway and SCI Store in the Scottish health organisations.

3.2.Implementation of standardised systems

The parallel synchronicity between standard and systems leads to a constant feedback between standard and system development and implementation. At the same time, the pragmatic approach to standard development allows changes or refinements in business requirements that appear during implementation to be readily incorporated into the standard. Such an approach allows, in theory, to accommodate the needs of the various users and was seen by the policy makers and SCI developers as a way to facilitate the implementation of the standardised systems into the local NHS organisations. However, the implementation of the standards for clinical data exchange embedded in the new IT systems has raised a number of problems. These problems as well as the factors that were found to facilitate the implementation process are discussed below.

In a first stage, the implementation of the standards for clinical data exchange in the local trust was hampered by two major technical problems.

First, the health sector (in general) is populated by a large number of stand alone IT systems that do not interact with one another and where there is almost no reuse of data or software components. In Scotland, such a fragmentation is less visible in the primary care sector as GPASS (a GPs IT system) has achieved around 80% representation. However, in secondary care, the diversity of existing IT systems means that their integration with the standardised products is highly problematic. According to one of the interviewees, *"[in] one of the Trusts there [are] 65 different cancer systems, which is actually more than there are clinicians. So they've got a different system for each clinician and each different cancer specialty. None of them joined up, none of them sharing any information, all of them basically access databases, very small, thin bits of product, all of which do what the clinician wants but don't do anything else."* (NHS IT1).

A second concern is related to the need to build generic standards that aim to bring together the interests of various categories of users throughout the NHS system. Such a requirement leads to a lack of customisation of the standard to the various working practices and the local systems within the different trusts. At the same time, the format and content of a number of the standardised documents (e.g. referral letters) are designed to serve both clinical and administration needs. This leads to an overload of administrative and clinical data within the same document, which translates in lengthy documents, difficult to read by either one of the two categories of users.

Although the technical problems were perceived as significant, it was the social issues that were stresses as the major factor in shaping the implementation of the standardised technology. According to one the SCI developers, *"interfacing legacy systems is always a bit of a challenge but it hasn't been one that's really stopped anything … so the challenge may often, or sometimes, be perceived as being old technology has prevented things from happening, whereas in reality it tends to be the inability to implement new business processes that caused the challenge"* (NHS IT2).

First, the health service in general is characterised by a clear demarcation between primary and secondary care, between GPs and hospital clinicians. Although NHS Scotland has made significant effort to close the gap between primary and secondary care through programmes such as the Manage Care, the demarcation is still visible. According to one of the interviewees, *"[there is] quite a big separation between the way that the GPs work and the way the clinicians work, and what their view is on what an appropriate patient is … there's not really perhaps a lot of trust between the two sides"* (NHS IT3). The segregation between primary and secondary care consists not only in different working practices and different perspectives, but also in a lack of communication and trust. Such segregation makes it difficult to promote the benefits of standardisation when they are unevenly split between the two sectors. For

example, standardising discharge letters means that all the efforts are made in the secondary care (by junior doctors who have to take time to fill in the standardised forms), whereas all the benefits are transferred in the primary care (better and rapid information for general practitioners).

A second problem is linked with the characteristics of clinical culture. In general, clinicians are seen to belong to a very diverse and highly independent community, and Scotland is no exception: *"if you get 100 clinicians in a room you'll get 101 different views. I guess they're quite sort of independent-minded 'I have a way of doing things and it's the right way' kind of mentality, which means that it's quite hard to extract the common approach to something"* (NHS IT3). At the same time, diversity and variety is seen as strength rather than a weakness: "*doctors don't like doing things by rote, they don't like protocols, they don't like checklists, it's an art rather than a science. So the more you say you need to specify this, you need to revise this data, then the more resistant they are*" (NHS IT1). In this context, the introduction of a product that requires standardised working practices (for example filling in a standard discharge letter) is likely to meet a significant degree of resistance. Such a resistance is exacerbated by the fact that clinicians perceive the standardisation of their working practices as imposed by a central authority. They therefore see this as a breach in their right to be involved in the decisions that affect their work. As one of the SCI managers in one of local NHS organisations mentioned: *"[there is a] feeling it's a national product, somebody somewhere's decided we're going to have this and it's being imposed on us"* (LTB1). On the positive side, clear communication between the SCI team and the users regarding the benefits of the standardisation process was found to facilitate the implementation of the standards embedded in the new SCI products. Such benefits however had, in most of the cases, to be perceived by users as linked with their own, individual "self interest" rather then with the clinical community as a whole. The reason for such an individualistic approach is not only the nature of clinical culture - "*clinicians work for themselves, they don't work for each other*" (NHS IT1) – but also the significant pressures that are presently shifting the health system. As one of the interviewees mentioned when asked to discuss the factors for supporting a successful implementation of the standards: *"Self-interest. Two word answer. … the conditions round about the people locally, things like the new GMS contract, the pressures on hours, the pressures on their need for audit of their clinical practice, the pressures for information to support clinical practice, pressures from patients all create an atmosphere of self-interest."* (PM)

Two other factors were found to facilitate the implementation of the standards within the local NHS organisations: (1) experience of working in the clinical environment and (2) knowledge of the local people and of the existing systems and working processes for the members of the local team responsible for coordinating the implementation of the SCI products.

Previous experience in the health service helped the local implementation team to understand the clinical environment and the clinical culture. As a respondent from one of the local trusts mentioned: "*I'm too long in the tooth in the health service … so I knew what they would say, that's why there was no way I was going to ask them. You avoid the problem, but we've no complaints. People, discussion and communication's a wonderful thing, and it works good most of the time, but there's a time when you don't want to get involved in it*" (LTB2). Such generic knowledge regarding the clinical environment helped the local team to understand how clinicians think and act, and consequently to manage better the implementation process.

Knowing the people and the processes that were affected by the standardization process helped the local team to coordinate better the implementation. For example, the fact that one of the members of the local team has worked within the department where the standardised systems were introduced was found useful: *"I worked in the 3 labs in [this trust board] at some point or another, so I knew the people, I knew what they were talking about and I also knew the objections they would come up with and could counter with reasonable arguments, without much hassle."* (LTB2). Such local knowledge facilitated mutual understanding which helped the implementation team to deal with the users' concerns.

4. Discussions

The table below summarises the factors that were found to facilitate and hamper the implementation of the standardised SCI products in NHS Scotland.

Table 1:
Factors shaping the implementation of the standardised IT systems

FACTORS (+/-)	EXPLANATION
(-) Integration problems with existing IT systems	Fragmented IT structure in secondary care
(-) Information overload in the standardised document	Need for generic standard
(-) Lack of customisation to local working practices and/or existing systems	
(-) Perception of unbalanced distribution of benefits resulting from standardising information exchanges	Segregation, lack of trust and communication between and within the various parts of the health care sectors
(-) Resistance to change existing, non-standard working practices	Independent oriented organisational culture which need to retain ownership over changes that affect their work
(-) Changes perceived as imposed by the central authority	
(+) Demonstrate clear benefits	Required due to overload of pressures/new reforms in the health care + history of IT failure
(+) Experience in the health service	Generic knowledge
(+) Knowing the people & processes	Local knowledge

Only two of the factors identified during the interviews refer to technical issues (integration with existing systems and information overload of the standardised documents. Technical issues regarding standardisation thus appear to play only minor role in shaping the implementation process in comparison with the social aspects.

Unbalanced distribution of benefits, resistance to adopt standardised procedures, and the perception of the users that such changes are imposed from outside the local organisation obstruct the implementation of standardised technology. At the same time, the outcome of the process depends to a large extent on the characteristics of the organisational context where the implementation takes place. The independent oriented organisational culture which praises diversity and individualism, as well as the strong divide between the various parts of systems translate in a number of challenges for the implementation of common standards for information exchange.

On the positive side, the ability of the standard developers to clearly define the benefits of adopting the standards, together with generic and local knowledge at the level of the local teams that coordinate the implementation process facilitate the process and smoothes the adoption of the standards within local organisations. Such knowledge facilitates an understanding of the minds and expectations of the people, as well as of the local, un-standardised practices and help in managing the change process.

5. Conclusions

The analysis of the implementation of Scottish NHS standard for clinical data communication has shown that the implementation of standards raises a number of concerns which greatly exceeds the technical issues. As one of the technical members of a local team mentioned: *"The technology's easy, it's the people issues that are the difficult to tackle. I've got the easy job here*" (LTB3).

The Scottish case showed that whereas the technical problems can be overcame, it is the "people issues", the social side of the implementation process that matters. Although standardisation is seen as a generic process, which aims to bring commonality between diverse locales, it appears that the outcome of the standardisation process depends to a large extent on the nature of the organisational context in which such standards are implemented. The organisational culture, the fragmentation of the system, as well as the ability of the standardisers to understand this context, and the people and processes within its shape standards implementation.

REFERENCES

Antonelli, C. (1994). Localized technological change and the evolution of standards as economic institutions. *Information Economics and Policy*, 6 (3-4), 195-216.

Belleflamme, P. (2002) Coordination on formal vs. de facto standards: a dynamic approach, *European Journal of Political Economy*, 18(1), 53-176.

Besen, S.M., and Farrell, J. (1991). The Role of ITU in Standardization - Pre-Eminence, Impotence or Rubber Stamp. *Telecommunications Policy*, 15(4), 311-321.

Damsgaard, J., and Truex, D. (2000). Binary trading relations and the limits of EDI standards: the Procustean bed of standards, *European Journal of Information Systems*, 9(3), 173-188.

David, P.A. (1985). Clio and the Economics of QWERTY, *The American Economic Review*, 75(2), 332-337.

David, P.A., and Shurmer, M. (1996). Formal standards-setting for global telecommunications and information services. Towards an institutional regime transformation?. *Telecommunications Policy*, 20(10), 789-815.

David, P.A., and Steinmueller, W.E. (1994). Economics of compatibility standards and competition in telecommunication networks. *Information Economics and Policy*, 6(3-4), 217-241.

Farrell, J., and Saloner, G. (1988). Coordination Through Committees and Markets. *The RAND Journal of Economics*. (19(2), 235-252.

Farrell, J., and Saloner, G. (1995). Standardization, Compatibility, and Innovation. *The RAND Journal of Economics*, 16(1), 70-83.

Graham, I., Spinardi, G., Williams, R., and Webster, J. (1995). The Dynamics of EDI Standard Development. *Technology Analysis & Strategic Management*, 7(1), 3-20.

Hawkins, R. (1999). The rise of consortia in the information and communication technology industries (emerging implications for policy. Telecommunications Policy, 23(2), 159-173.

Moreton, R., Sloane, A. and Simon E. (1995). Implementing information management and technology standards: a framework. *Technology Management*, 2(6), 275-288.

Medjahed, B., Benatallah, B., Bouguettya, A., Nhu, A.H.H., and Elmagarmid, A.K. (2003). Business-to-business interactions: issues and enabling technologies. *The VLDB Journal*, 12(1), 59-85.

Schmidt, S.K., and R. Werle. (1998). *Co-ordinating technology: Studies in the international standardization of telecommunication*. MIT Press, Cambridge.

Sloane, A. (2000). The standards process: tools and methods for standards tracking and implementation. *Computer Standards & Interfaces*, 22(1), 5-12.

Turban, E., J. Lee, D. King, and H.M. Chung Electronic Commerce. (1995). *A Managerial Perspective*. Prentice Hall International, New Jersey.

Weiss, M.B.H. (1993) The Standards Development Process: A View from Political Theory. *Standard View*, 1(2), 35-41.

STUDIA UNIVERSITATIS BABEŞ-BOLYAI, OECONOMICA, L, 2, 2005

MONETARY POLICY RULES IN A CONVERGING SMALL OPEN ECONOMY*

ÉVA KATALIN POLGÁR**

** *Corvinus University of Budapest, Hungary and European Central Bank*

ABSTRACT. The analysis of monetary policy rules has been confined to models not capable of examining situations where an economy is converging to a higher balanced growth path. For the small open economies having entered the European Union recently this is however a very relevant question. The main aim of their integration is convergence and most of the criteria they have to fulfil to become members of the euro zone are of monetary nature. It is of special interest for them whether the chosen monetary strategy aiming at the fulfilment of the requirements affects transition. In this paper a first attempt is made to compare monetary policy rules in a monetary model of small open economies, which builds essentially on the convergence literature. The results show that the economy behaves very differently in the transition under the different monetary policy rules examined.

JEL Classification: E42, E52

Keywords: monetary policy, economic transition, converging economy

1. Introduction

The "eastern enlargement'" of the European Union (EU) took place in May 2004, when ten countries became new members, mostly small open economies approaching the end of a transition process from their centrally planned economies to market economies. On accession they also became candidates for membership in the Economic and Monetary Union. The entry, however, requires the fulfilment of different eligibility criteria (the *convergence criteria of Maastricht*) most of which are of monetary nature. Thus monetary policy has a very important role to play in these countries. What strategy would be preferable for them?

* I wish to thank János Vincze, Dietmar Meyer, Tamás Bánfi and Balázs Szentes for helpful comments and discussions. Special thanks go to Horst Tomann who was supervising my work for one year with much care and attention. For the same reason I am particularly indebted to Ákos Valentinyi whose constant help in the main phase of the research was invaluable. All the remaining errors are mine. Financial support of the German Academic Exchange Service (DAAD) and the Marie Curie Fellowship scheme of the European Commission (HPMT-CT-2001-00353) are gratefully acknowledged. The views expressed here are solely the responsibility of the author and should not be interpreted as reflecting the views of the European Community or the European Central Bank.

If we look at the current design of monetary policy in the ten countries, we find a very diverse picture. From giving up monetary autonomy completely and relying on a currency board to a completely free float combined with inflation targeting different strategies are being followed. Theory does not give too much guidance, either. This may seem surprising, since there are many monetary general equilibrium models analysing the optimality of monetary policy rules in several different set-ups, also for the case of a small open economy.[1] But most of them abstract from production or include production using labour as the only input. Moreover, they examine economies being close to their steady states in low inflation environments. Almost none of these assumptions are a good approximation in the case of the new members. They are not supposed to be close to steady state, but on their transition path towards it or in a catch-up process converging to a higher balanced growth path. This convergence is one of the goals of their integration.

Using models of convergence could be another approach. These are growth models where different frictions are built into an otherwise neoclassical framework to get a speed of convergence in the range of empirical estimates. But these are mostly non-monetary models.[2] There are of course models including both capital and money explicitly, but they mainly deal with closed economies and more importantly the basic model describing the real economy is not a convergence model.[3]

The interaction between transition and monetary policy is not at all impossible. Already in 1979 Stanley Fischer showed that the results of e.g. Sidrauski (1967) about the superneutrality of money hold true only in the steady-state. This already relies on rather special assumptions, but even in models where it holds, money affects transition: the speed of convergence is not independent of the growth rate of the nominal money supply. For the transition money is superneutral only if the utility function is logarithmic, but for the constant elasticity of substitution form most commonly used in empirical applications it is not.

I believe that all three basic elements mentioned (*small open economies, convergence and monetary nature*) are essential for an analysis of optimal monetary policy in the case of these countries. So as a contribution to the

[1] For closed economy see Galí (2002), Woodford (2001); for two-country models Obstfeld-Rogoff (1995 and 1998), a model for a small open economy is presented in Galí-Monacelli (2002).

[2] See e.g. Barro–Mankiw–Sala-i-Martin (1995), Chatterjee–Sakoulis–Turnovsky (2001), Lane (2001).

[3] The dynamic stochastic general equilibrium approaches (DSGE-models) became the primary tools for empirically oriented modelling, such as e.g. Smets–Wouters (2003) or Del Negro et. al. (2004). These models are not aiming for modelling convergence either and for better empirical results they are assuming many different types of nominal rigidities. To my knowledge only Benczúr (2003) and Benczúr–Kónya (2004) use certain elements necessary for modelling convergence, but their goal differs from mine: instead of concentrating on the interactions of monetary policy and convergence, they are examining the real effects of different nominal shocks.

existing body of literature this study tries to combine the three. I will try to choose the simplest setup as basis (this is a convergence model that builds on the assumption of *constrained capital mobility*) and introduce money into it (applying the *money in the utility function approach*). I assume *flexible prices* (in part for simplicity, but also because for sticky prices monopolistic competition should be assumed, which is questionable for small open economies).

Due to space limitations I give a summary of the paper here, concentrating on the setup and the results only, while leaving the equations and calculations out. The basic model equations are summarised in an appendix and a full version of the paper is available from the author upon request. Here section 2 discusses the model, section 3 the results and section 4 concludes.

2. The model

My goal is to develop a model for analysing monetary policy in converging small open economies. Here *convergence* means the process of reaching the steady state from the starting point: the path from the state given by the initial conditions to the one characterised by a balanced growth path. I believe that for this we need a monetary model of a small open economy in transition. In this attempt I chose the model setup used in Lane (2001);[4] a dynamic general equilibrium model of a small open economy in continuous time, where *capital mobility is constrained* enabling a straightforward handling of the model structure and slowing down convergence. I "monetise" the model economy using the *money in the utility function approach* following Stanley Fischer.[5]

The economy is populated with a large number of infinitely lived identical competitive households owning capital in the economy. They are operating the production technology themselves, having access to the same production technology (representative agent). They are selling their final product on a competitive market having no market power there. The production process has two phases. First intermediate products are being produced used later as inputs for producing the final products. There are traded and nontraded inputs. Both are produced using different physical capital goods. No labour input is employed, and the technology of producing the intermediate products is Cobb–Douglas (where I assume the same exogenous productivity parameter and the same capital intensity for both sectors – thus ruling out growth effects resulting from the variation in productivity between the two sectors for now). The final output is produced combining the intermediate products through a similar Cobb–Douglas technology. The exponent of the traded input expresses the degree of openness. Output can be consumed or invested.

[4] This is a modified version of the one by Barro–Mankiw–Sala-i-Martin (1995).
[5] Fischer (1979). This approach was introduced in Sidrauski (1967).

We impose a borrowing constraint: the country can borrow at perfect capital markets but only limited amount not exceeding the value of capital in the traded sector because this is the only acceptable collateral.[6] Practically optimality will require then the equality of traded capital and foreign debt. Hence, the net return on capital in the traded sector has to equal the world real interest rate. Using this, the aggregate production function can be rewritten, so that it only depends on the non-traded capital good. The assumption of constrained capital mobility is useful as it slows down convergence in an open economy and also makes the analysis simpler by reducing the number of variables in the model.[7]

Households have a dynamic constrained utility maximisation problem. Money is introduced through lump sum transfers. Households make two decisions: first about the ratio of their consumption and saving, and secondly a portfolio choice (what to use savings for: accumulating capital stock or real money balances). The government plays a very simple role here: it provides economic agents with money.

The solution results in the *equilibrium conditions:* the transversality condition, the budget constraint of the government, the resource constraint of the economy, a static and a dynamic equilibrium condition. The static one states that the marginal rate of substitution between real balances and consumption equals the relative price of these two. According to the dynamic condition (the equivalent of an intertemporal Euler equation in the model) the optimal growth path of consumption and real balances depends on the distance from the steady-state: if the difference in the returns is big, than the transition is faster. To solve the system we also need a policy function.

I examine four different monetary policy strategies. The first one, money supply rule means nominal money supply targeting. The national bank can determine the rate of monetary expansion. In the second case the monetary authority controls the nominal interest rate: it chooses a level regarded as adequate and then keeps it constant. In the third case of inflation targeting the monetary authority controls inflation directly. Again, it decides on an adequate level and then keeps it constant.[8] In the fourth case the nominal interest rate is controlled again, but here it is changed in line with inflation. The central bank has a reaction function: if inflation is higher, the nominal interest rate is raised and vice versa.

[6] Several reasons could be mentioned for this: it is easier to impose sanctions on this sector or this capital is useful also to foreign investors. See Lane [2001] p.224. for more details.

[7] Introducing explicit markets for all types of goods (thus including the relative prices in the model) would lead to an equivalent formula for the real interest rate and thus would not change the results.

[8] The definition used here is a very simple, special definition of inflation targeting. We could not include forms taking expectations into account, however, as the model is deterministic.

To investigate on the stability I log-linearise the system around the steady state in all cases.[9] The steady state is always a saddle point here; hence there is a unique (converging) transition path. Then I analyse the impact of monetary policy on the speed of convergence along this path.

3. Summary of the results

The goal of the model is the examination of the question whether it matters for convergence what monetary policy rule is followed. On the basis of the above analysis the answer is definitely yes: the model economy behaves very differently in the transition under the different monetary policy rules assumed, although the equilibrium values of the real variables are the same. Now I shortly summarize the main results.

	Steady-state	**Transition: for faster convergence…**
Money-supply rule	money is superneutral	faster monetary expansion
Interest rate targeting	money is superneutral	money is superneutral here too!
Inflation targeting	money is superneutral	lower inflation rate (not a general result!)
Feedback–rule	money is superneutral, nominal variables are not uniquely defined!	lower inflation rate (not a general result!)

With a *money supply rule* the steady state inflation rate (and through this the group of nominal variables) is determined through the chosen value for the growth rate of the nominal money supply, while the real ones are determined independently of money by the model's parameters (so as in all of the cases, money does not matter in the steady state). Money is not superneutral during transition: after controlling for openness, the speed of convergence is higher, the higher is the growth rate of the money supply. This calls for increasing the nominal money supply, which however results in higher steady state inflation.

The *interest rate rule* stands out leading to very different results from the others. Money is superneutral also in transition. What matters is the degree of openness: the greater it is, the faster convergence will be (we get this result in all the examined cases).[10]

Under *inflation targeting* the monetary authority controls inflation directly and through the choice of its steady state level the monetary variables are also determined in the steady state. We do not have clear-cut results here,

[9] This way we approximate values close to the steady-state. Values far from it can only be examined numerically. This paper is confined to the analytical examination now.

[10] Nominal variables are determined by the choice of the nominal interest rate.

though money matters in transition: under certain conditions a higher inflation rate will reduce the speed of convergence (when controlled for openness).

For the *feedback-rule* results are similar: one stable saddle-path and uncertainty about the effects of monetary policy in transition (that are however present). Again under certain conditions a higher inflation rate results in slower convergence with openness kept constant. Both cases are contradicting the results under the money supply rule, where a higher money growth rate (and so higher inflation) can lead to faster convergence. The only difference from inflation targeting is that monetary variables remain undetermined in the steady state in this case.

In all three cases where monetary policy affects convergence, for a special case of the utility function (logarithmic form) the superneutrality of money holds also in the transition. Also, in the three cases examined openness always furthered, speeded up convergence, but under the last two rules this relationship was not generally true (it held only for certain parameter values). The result is consistent with the general view of economists: a higher degree of openness enables a faster adjustment.

4. Conclusion

The analysis of this simple model illustrates that the dynamic behaviour of the economy can differ to a great extent under different monetary policy rules already in a very simple setup. The model presented combined the elements regarded as inevitable, but other important elements have still been missing. The most relevant for a small open economy is probably the exchange rate. An obvious extension is thus the analysis of similar problems in models where the exchange rate can also explicitly be taken into account. There are other possible extensions. It can be worth modifying the model by assuming different productivity parameters for the production technology of the sectors and examine whether the Balassa–Samuelson effect is valid in this two-sector model. An important extension would be the numerical examination of the model to see the complete transition path. But for this it could be justifiable to set up and solve a stochastic version of the model first.

Nevertheless, the presented model highlights the differences in the behaviour of the model economy under different monetary strategies and due to the simple structure a local analysis is also analytically possible. Since the exchange rate is not explicitly included, it can be used to characterise e.g. the situation after the introduction of the euro, where the currency used by the main trading partners is not different from the one used in the respective country – so exchange rate does not matter for the economic decisions. The analysis of the model is therefore despite all the listed shortcomings in any case theoretically interesting and justifiable.

Appendix

The production functions of the intermediate goods: $y_T = Ak_T^{\alpha}$ and $y_N = Ak_N^{\alpha}$. Production function of final output: $y = y_N^{1-\theta} y_T^{\theta}$.

The borrowing constraint leads to: $k_T (= d) = \dfrac{\alpha\theta y}{r^w + \delta}$.

The final form of the production function accordingly: $y = Bk_N^{\eta}$, where $B = A^{\frac{1}{1-\alpha\theta}} \left(\dfrac{\alpha\theta}{r^w + \delta} \right)^{\frac{\alpha\theta}{1-\alpha\theta}}$ and $\eta = \dfrac{\alpha - \alpha\theta}{1-\alpha\theta}, 0 < \eta < 1$.

Households' problem:

$$\max \int_0^{\infty} e^{-\rho t} \frac{(c^{\beta} m^{\gamma})^{1-\sigma} - 1}{1-\sigma} dt \text{ , where } \beta, \gamma, \sigma > 0, \beta + \gamma \leq 1 \text{ and}$$

$$\dot{k}_N + \dot{m} = (1-\alpha\theta) Bk_N^{\eta} + x - c - \delta k_N - m\pi$$

$$k(0) = k_N(0) + k_T(0) > 0$$

$$c, m, k_N, k_T, d \geq 0; r^w, P \geq 0, x = given$$

$$k_N(0) + k_T(0) - d(0) < k_N^*$$

$$\lim_{t \to \infty} (k_N + k_T - d) e^{-r^w t} \geq 0$$

Government's budget constraint: $\dot{m} + m\pi = x$.
Conditions of equilibrium:

$$\lim_{t \to \infty} \lambda (k_T + k_N - d) e^{-\rho t} \leq 0$$

$$\dot{m} + m\pi = x$$

$$(1-\alpha\theta) Bk_N^{\eta} - c - \delta k_N = \dot{k}_N$$

$$(1-\alpha\theta)\eta Bk_N^{\eta-1} - \delta + \pi = \frac{\gamma c}{\beta m}$$

$$(\beta - 1 - \beta\sigma)\frac{\dot{c}}{c} + (\gamma - \gamma\sigma)\frac{\dot{m}}{m} = -(1-\alpha\theta)\eta Bk_N^{\eta-1} + \delta + \rho$$

REFERENCES

Barro, R., Mankiw, G., Sala-i-Martin, X. (1995): Capital Mobility in Neoclassical Models of Growth, *American Economic Review*, Vol. 85, 1 (March), pp.103-115.

Benczúr, P. (2003): Nominális sokkok átmeneti reálhatása egy kétszektoros növekedési modellben, *MNB Füzetek, National Bank of Hungary*, 9, November

Benczúr, P., Kónya, I. (2004): Nominal growth in a small open economy, manuscript, for *the October 2004 conference titled Economic Growth and Development by the Institute of Economics of the HAS*[11], September

Chatterjee, S., Sakoulis, G., Turnovsky, S. (2001): Unilateral Capital Transfers, Public Investment, and Economic Growth, *European Economic Review*, Vol. 47, No. 6, pp.1077-1103.

Del Negro, M., Schorfheide, F., Smets, F., Wouters, R. (2004): On the Fit and Forecasting Performance of New Keynesian Models, *Federal Reserve Bank of Atlanta, Working Paper*, 37, December

Fischer, S. (1979): Capital Accumulation on the Transition Path in a Monetary Optimizing Model, *Econometrica*, Vol. 47, 6 (November), pp.1433-1439.

Galí, J. (2002): New Perspectives on Monetary Policy, Inflation, and the Business Cycle, *NBER, Working Paper*, 8767, 2002

Galí, J., Monacelli, T. (2001): Monetary Policy and Exchange Rate Volatility in a Small Open Economy, *NBER, Working Paper*, 8905

Lane, Ph. (2001): International trade and economic convergence: the credit channel, *Oxford Economic Papers*, 53, pp.221-240.

Obstfeld, M., Rogoff, K. (1995): Exchange Rate Dynamics Redux, *Journal of Political Economy*, Vol.103, 3, pp.624-660.

Obstfeld, M., Rogoff, K. (1998): Risk and Exchange Rates, *NBER, Working Paper*, 6694

Sidrauski, M. (1998): Rational Choice and Patterns of Growth in a Monetary Economy, *American Economic Review*, Papers and Proceedings, 57, pp.534-544.

Smets, F., Wouters, R. (2003): An estimated stochastic dynamic general equilibrium model of the euro area, *Journal of Economic Association*, Vol. 1, No. 5 (September), pp.1123-1175.

Walsh, C. (2003): *Monetary Theory and Policy*, second edition, MIT Press, Cambridge, MA

Woodford, M. (2003): *Interest and Prices, Foundations of a Theory of Monetary Policy*, Princeton University Press, Princeton

[11] Hungarian Academy of Sciences

STUDIA UNIVERSITATIS BABEŞ-BOLYAI, OECONOMICA, L, 2, 2005

MACROECONOMIC IMPACT OF GDP COMPOSITION CHANGES IN THE CROATIAN ECONOMY

VLADIMIR ČAVRAK, IVO DRUŽIĆ, DOMINIK PRIPUŽIĆ*

* *University of Zagreb, Croatia*

ABSTRACT. In this article the share of the various sectors in GDP and some newer trends in the sectoral composition of Croatian GDP were analyzed. Methodological changes that happened during the transition period disable simple construction of longer time-series data. Even by limiting the observed period to a methodologically consistent time interval, the divergent results of the present structure of Croatian economy arise. These results indicate that the structural changes in the transition period are very dynamic and that they will probably continue to happen, possibly even to a higher scale. This means that there is a need for a continued and more detailed analysis of relative and absolute importance of different sectors. However, even with the present time-series the serious problem of the long-term deindustrialization of Croatian economy is evident. The deindustrialization denotes the phenomenon of decline of the structural share of industry and the absolute fall of the physical volume of industrial production. Most transition economies have experienced the dynamic growth of industrial production after restructuring. This growth provides an additional impulse to overall GDP growth. Contrary to them, Croatia still experiences opposite tendencies, which is partially reflected through the inconsistency of the GDP calculation methodology.

JEL Classification: E01

Keywords: GDP, structural changes, transition

1. Transition and structural changes

It has been empirically known and confirmed that changes in the structure of national economy occur together with the process of economic growth and development of society. Numerous works have been written in this field due to the scientific importance of the subject (Chenery 1960; Kuznets 1966; Chenery and Syrquin 1989).

Economic transition is a process of transition from planned to market economic conditions, i.e. to a market as a mechanism of allocation of scarce resources. Under such conditions, a change in the economic structure does not only appear as a result of economic development, but of the transition itself. Large and sudden changes in the economic structure of most transition countries, such as the increase of unemployment, decrease of production, instability etc., to pre-transitional economic conditions, which are characterized by strong centralization and competition-free business, full employment,

almost without any private ownerships, show the power of many intertwined factors which all influence the structural changes of gross domestic product.

There are many reasons why the economic structure in planned economies differs from the economic structure in market economies; for example, the preference given to industry (esp. hard industry), very high investment rates in the heavy industry, the lack of a financial system that would efficiently allocate savings into investments and the absence of required institutional and legal regulation that operate in market economy (Gross and Suhrcke, 2000, p.1.).

On the other hand, it was convenient that the Soviet type planned economies were suitable for the mobilization of resources necessary for the expansion of the production activities during World War II and the after-war reconstruction. Accelerated growth as a direct consequence of forced industrialization initially resulted in high growth rates of planned economies, which enabled industrialization and the insertion of the majority of planned economies among middle-class developed industrial countries, and also the reduction of a development gap between eastern and western Europe in the period of 1950s and 1960s of the twentieth century. (Svejnar 2002., p. 4.).

The fall of Communism created such expectations as that the former planned economies, when transitioning to market economic conditions, will achieve accelerated economic growth and gradually approach to the developed countries. It could be argued that the process of general, especially economic transition in ex-planned economies was not part of the general knowledge at the time. It is sometimes assumed, however in some reduced approach that literature on transition before the breakdown of planned economies did not exist at all (Gregory P.; Stuart. P. 2000, p 266). However, it turned out that the transformation from planned to market economy was far too complicated than it was initially expected. The reasons why performances of transition economies were less than expected were identified.

In the first place, it was considered that, when all obstacles of planned economies were removed, it will be possible for market forces to move quickly and spontaneously. But this was not the case. Events show us that the creation of efficient market frame is a far more complicated process than initially thought, which in developed market economies took place over a long period. One of the obstacles, the absence of market institution and infrastructure, was initially not considered significant. Secondly, western economies have achieved very good results in the 1990s, which raised the economic efficiency scale and the level of expectation. Thirdly, economic problems connected with transition were under-rated, and policy makers in many countries did many wrong steps.

It is clear how a change in output structure is common to all transition countries (see table 1.).

Table 1.

Composition of output (percentage of GDP), 1990-91and 1997-98

Regions and periods	Agriculture	Industry	Services
CSB			
1990-91	13,7	45,1	41,2
1997-98	13,9	33	53,1
CIS			
1990-91	27,5	39,7	32,8
1997-98	18,7	31,2	50,1

Source: World Bank, 2002, p. 6.

During the transition, both in countries of Central, Southeastern Europe and Baltics (CSB) and in countries of ex USSR i.e. Commonwealth of Independent States (CIS) share of industry in GDP fell on the third approximately, while share of the services' sector rose on half of total GDP. The main difference between these two groups of countries was the fact that the rise in share of the services in CSB countries fell on the burden of industry share shrinking, while in CIS countries share of the services rose equally due to the fall of industry and agriculture as well.

As opposed to the majority of other centrally planned economies, Croatia operated in a specific, so called market-planned environment, defined as "market socialism". That special system of so–called workers' self-management was not able to respond to the new development challenges at the beginning of the 1980s. In that period big oscillations occurred, and massive industrial plants were not successfully restructured neither adapted to the new technologies and the age of computers, need for higher labor mobility, better motivation of the employees, and contemporary business organization etc.

In the 1990s Croatia had to deal with burden of war, large human losses, destroyed infrastructure, problem of hundreds of thousand displaced persons and refugees, occupation of some part of its territory etc, not to mention all the complexity of transition process. So, additional complex structural changes could not be interpolated by using current economic models of transition. Namely, in practice these operating models of transition can be divided into two major groups.

First group of transition models is most frequently linked with the process of creation a new employment by economic growth in so-called new sector. (for details see Blanchard, 1997). Second group of models emphasize previous restructuring and stabilization as a condition for the efficient transition, where we achieve desirable macroeconomic effects in a short-run by market automatism, leaving off the rigidity of previous system. In both of these groups

theoretical supposition emphasize linkage of the efficient transition with indispensable changes in structure of GDP.

2. GDP composition changes in Croatian economy from 1995 to 2004.

For the analysis of the structural change during transition, a structural change needs to be defined first. Structural change can denote different changes, i.e. the change in the structure of employment by activities, the change in the export structure, the change in the structure of financial sector etc. In this article, the term structural change will indicate the change of different sectors share in total gross value added of the economy. Sector classification will follow the NACE – NKD standard. NKD or National Classification of Economic activities (in Croatian: Nacionalna klasifikacija djelatnosti) is the official Croatian classification of economic activities. It is worked out on the basis of statistical classification of economic activities that is used in the EU (NACE Rev. 1) in order to ensure that the statistical data can be compared, and, indirectly, to ensure coordination with the international statistical standards like the System of National Accounts, issued by the UN.

In the pre-transition period the key macroeconomic indicator, used as an indicator of the level of economic activity, was social gross product (DBP – društveni bruto proizvod) and social product (DP – društveni proizvod). Social gross product and its primary distribution were calculated according to real (or productive) method, or so called material production concept. Starting in 1976 the results of DBP and DP calculations are constructed and calculated according to old national classification UCEA or Unified Classification of Economic Activities (in Croatian: JKD or Jedinstvena Klasfikacija Djelatnosti), so the comparison to earlier data is not possible.

Since the conversion of the data classified according to the old classification (JKD) has not been made yet, our analysis will rely on the consistent time-series data, beginning in 1995. The data in this interval is methodologically consistent and can be compared. The choice of this specific period was, by our beliefs, necessary in order to get the more reliable results, since conceptual, methodological and technical problems make the detailed analysis of structure, dynamics and measurement of the GDP very complex.. By concentrating on the consistent time series, it is possible to gain better understanding about the way Croatian economy changes. These insights can prove helpful later, when the analysis is extended over the longer but inconsistent time series. Besides, for the simplicity reasons, even though the analysis will rely on GVA or gross value added, instead of gross domestic product as a measure of economic activity, we will refer to it as GDP.

As every other classification, the NKD is divided into different levels. First level consists of 17 sections. Each section is marked with its own alphabetical code. Since the goal of analysis is broad patterns of changes in economic structure, we will group some of these sections into 8 different categories - sectors. The first category "agriculture" incorporates two sections, namely the section "A – Agriculture, hunting and forestry" and section "B – Fishing". The next three sections, "C – Mining and quarrying", "D – Manufacturing" and "E – Electricity, water and gas supply" are joined into one broad category "industry". Next few sections remain ungrouped. These are "construction" ("F – construction by NKD), "trade and repairs" ("G – Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods"), "hotels and restaurants" ("H – hotels and restaurants") and "traffic" ("I – transport, storage and communication"). Next category "finance" consists of two related sections "J – Financial intermediation" and "K – Real estate, renting and business activities". Finally we have one very wide category that includes all remaining sections. It is called "public sector" and it includes following sections: "L – Public administration and defense, compulsory social security", "M - Education", "N – Health and social work", "O – Other community, social and personal activities", "P – Private households with employed persons" and "Q – Extra-territorial organizations and bodies".

GDP in current prices grows through the whole observed period. Categories that have the largest share in value added are "industry" and "public sector". Besides that, from Table 2 one can see that, in the observed period, the GDP expressed in current prices has more than doubled, from its initial level of almost 80 billion HRK, to almost 180 billions HRK in the last observed year.

3. Sectoral analysis of structural changes in current prices

The sectoral analysis enables a precise insight into the dynamics of the structural changes in the Croatian economy (See Table 3). Sectors can be divided into few distinctive groups, depending on the movement of their share in the GDP in current prices.

The first group consists of two sectors with the falling share. The first sector from this group is "agriculture". Its share in GDP in current prices fell from 10,37% in 1995 to 7,88% in 2004. The fall of the next sector share, "industry" was much more pronounced, from 27,66% to 22,23%.

There are three sectors whose share in GDP rose during the observed period. The first one is "hotels and restaurants". Its share rose from 2,53% to 3,57%.Next we have "traffic". At the beginning its share in GDP was 9,62%. In the last year it was 10,85%.The third sector with the raising share is "finance". In 1995 it's share was 14,50%, and at the end of the observed

period, in 2004, it was 16,90%. However, it should be noted that for the two sectors, "traffic" and "finance", the share in GDP stagnated or fell in the first half of the observed period.

The third group includes remaining sectors, which displayed non-uniform movements in the observed period. These are "construction", "trade and repairs" and "public sector".

Table 2. GDP in current prices in billion HRK	NKD CODE	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Agriculture	A+B	8.413	9.059	9.688	10.901	11.501	11.240	12.482	13.038	13.260	14.051
Industry	C+D+E	22.439	23.306	27.113	29.074	29.314	31.539	33.858	35.157	37.054	39.645
Construction	F	4.625	5.965	7.437	7.732	6.346	5.876	6.832	8.546	10.732	11.854
Trade and repairs	G	9.502	11.122	13.051	13.788	11.892	13.160	15.955	18.347	20.039	21.297
Hotels and restaurants	H	2.050	2.732	3.266	3.581	3.614	4.217	4.761	5.261	5.974	6.371
Traffic	I	7.805	8.022	9.092	9.777	10.315	12.425	13.821	15.759	17.754	19.353
Finance	J+K	11.765	13.315	14.694	16.387	17.770	18.864	21.260	24.355	27.493	30.148
Public sector	L+M+N+O+P	14.524	17.359	20.367	25.053	28.464	30.340	30.090	31.574	33.323	35.647
Total		**81.123**	**90.880**	**104.707**	**116.293**	**119.216**	**127.661**	**139.060**	**152.037**	**165.629**	**178.366**

Source: DZS-SYC, various years

Table 3. Share of different sectors in GDP (current prices)	NKD CODE	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Agriculture	A+B	10,37	9,97	9,25	9,37	9,65	8,80	8,98	8,58	8,01	7,88
Industry	C+D+E	27,66	25,65	25,89	25,00	24,59	24,71	24,35	23,12	22,37	22,23
Construction	F	5,70	6,56	7,10	6,65	5,32	4,60	4,91	5,62	6,48	6,65
Trade and repairs	G	11,71	12,24	12,46	11,86	9,98	10,31	11,47	12,07	12,10	11,94
Hotels and restaurants	H	2,53	3,01	3,12	3,08	3,03	3,30	3,42	3,46	3,61	3,57
Traffic	I	9,62	8,83	8,68	8,41	8,65	9,73	9,94	10,37	10,72	10,85
Finance	J+K	14,50	14,65	14,03	14,09	14,91	14,78	15,29	16,02	16,60	16,90
Public sector	L+M+N+O+P	17,90	19,10	19,45	21,54	23,88	23,77	21,64	20,77	20,12	19,99
Total		**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Source: Authors' calculation

Table 4. GDP in constant 1997 prices in billion HRK	NKD CODE	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Agriculture	A+B	9.367	9.473	9.688	10.230	9.962	10.083	10.308	10.487	9.946	10.359
Industry	C+D+E	24.598	25.452	27.113	28.682	29.445	30.406	31.592	32.992	34.490	35.847
Construction	F	5.540	6.581	7.437	7.474	6.719	6.242	6.404	7.290	8.761	9.254
Trade and repairs	G	9.973	11.529	13.051	12.946	11.509	12.016	14.010	15.668	16.774	17.473
Hotels and restaurants	H	2.271	2.766	3.266	3.289	3.286	3.802	3.832	4.119	4.459	4.599
Traffic	I	8.210	8.784	9.092	9.483	8.999	9.404	10.485	11.242	12.275	13.124
Finance	J+K	13.632	13.928	14.694	15.288	15.845	15.883	16.003	17.308	18.556	19.659
Public sector	L+M+N+O+P	19.401	19.771	20.367	20.887	21.502	21.860	21.811	21.838	21.857	22.148
Total		**92.991**	**98.284**	**104.707**	**108.280**	**107.267**	**109.696**	**114.446**	**120.944**	**127.118**	**132.463**

Source: Authors' calculation

Table 5. Share of different sectors in GDP (const. 1997 prices)	NKD CODE	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Agriculture	A+B	10,07	9,64	9,25	9,45	9,29	9,19	9,01	8,67	7,82	7,82
Industry	C+D+E	26,45	25,90	25,89	26,49	27,45	27,72	27,60	27,28	27,13	27,06
Construction	F	5,96	6,70	7,10	6,90	6,26	5,69	5,60	6,03	6,89	6,99
Trade and repairs	G	10,72	11,73	12,46	11,96	10,73	10,95	12,24	12,95	13,20	13,19
Hotels and restaurants	H	2,44	2,81	3,12	3,04	3,06	3,47	3,35	3,41	3,51	3,47
Traffic	I	8,83	8,94	8,68	8,76	8,39	8,57	9,16	9,30	9,66	9,91
Finance	J+K	14,66	14,17	14,03	14,12	14,77	14,48	13,98	14,31	14,60	14,84
Public sector	L+M+N+O+P	20,86	20,12	19,45	19,29	20,05	19,93	19,06	18,06	17,19	16,72
Total		**100,00**	**100,00**	**100,00**	**100,00**	**100,00**	**100,00**	**100,00**	**100,00**	**100,00**	**100,00**

Source: Authors' calculation

The share of the "construction" exhibits a wave-like pattern, with highs in 1997 (7,1% of share in GDP), and in 2004 (6,65% of share in GDP), and lows in 1995 (5,7%) and in 2000 (4,6%). The share of the second sector in this group, "trade and repairs", also oscillated, even though with the growing trend. Movement of its share was as follows: 11,71% of share in GDP in 1995, 12,5% in 1997, 9,9% in 1999 and 11,94% in 2004. The strong growth of the share of third sector in this group, "public sector" in the first half of the nineties was a result of the newly born Croatian state and defense expenditures during the war conflict. The further growth, in the second half of the nineties, could be caused by the expanding bureaucracy, which also means that its efficacy was reduced. The share in GDP of this sector was 17,9% in 1995, peaked at 23,9% in 1999 and ended at 19,99% in 2004. It should be noted that the structure of GDP in current prices reflects the current state of an economy. Current prices are the prices that are valid in the moment in the time in which an economic activity takes place. However, "The GDP at constant prices is an extremely important measure of dynamics and the level of economic development, deprived from the influence of the price movements. The GDP at current prices disables the analysis of economic growth and structural changes in an economy, since it displays not only the dynamics of physical amount of output and real structural changes, but the changes in price movements too." (CSY-94, p 155.)

4. Analitycal discrepancies

When the focus of the analysis switches from GDP at current prices to GDP at constant prices, some interesting results are obtained. The movements of some sector shares at constant prices diverge significantly from the movement of the same sector in GDP at current prices. These sectors are "industry" and "public sector" (See Table 4 and 5). The main characteristic of the "industry" sector share in GDP at current prices is a constant decline, from 27,66% of share in 1995 to 22,23% of share in 2004. However, the same share, but now in GDP at constant prices, displays a very different behavior. It rises from 26,45% in 1995 to 27,06% in 2004, taking the highest value in 2000, 27,72%.The share of "public sector" in GDP at current prices, as mentioned earlier, oscillates (17,90% in 1995., 23,88% in 1999, 19,99% in 2004), but the same indicator, now in GDP at constant prices declines from 20,8% of share in 1995 to 16,72% in 2004.

If we take the shares in GDP at current prices, the most prominent feature is the continuing fall of the "industry" sector. Simple change of analytical approach, by taking the shares in GDP at constant 1997 prices reveals significantly different results. Now the share of "industry" is stagnant. One possible interpretation is that the reason lies in the disparate movement of the relative prices in different sectors, namely the higher inflation in services

sector. It is obvious that the analysis of the national economy structure is a highly complex task. Even though the analysis was limited to a consistent time-series data, using two simple methods of calculating the shares of various sectors in GDP, resulted in divergent outcomes.

Anyhow, without regarding to different outcomes, the questionable status of Croatia as a middle developed industrial country is evident. This questionable status, which cannot be hidden through the statistics, is a result of the deindustrialization process that Croatia experienced during the transition period. Deindustrialization is not the process of decreasing relative share of industry in GDP. This phenomenon is a natural outcome of economic development because of growing importance of tertiary sector and leading role of service sector. The relative share of industry in GDP may be decreasing, but the industrial production is still growing in absolute terms. The slower dynamics of its growth reduces its share in total output, compared to the overall level of GDP growth. The term deindustrialization refers to the process that took its place in transition economies, where it denotes the phenomenon of decline of the structural share of industry and the absolute fall of the physical volume of industrial production. These deep and significant changes experienced in Croatia and other transition economies may be one possible cause of the drastic disparities in shares of industrial sector in GDP at constant and current prices.

5. Concluding remarks

Propositions for further research include: comparative analysis of GDP structure and employment movement in the same sectors, comparison of the data on relative price movements in each sector with the shares of the same sectors in GDP at the current prices, definition of the appropriate conversion method for older data classified in accordance to JKD in order to enable comparison with newer data, and comparative analysis of the structural changes in other transition and post-transition countries.

However, even at the present stage of research the serious problem of the long-term deindustrialization of Croatian economy is evident. Most transition economies have experienced the dynamic growth of industrial production after restructuring. This growth provides an additional impulse to overall GDP growth. Contrary to them, Croatia still experiences opposite tendencies, part of which is possibly reflected through the inconsistency of the GDP calculation methodology.

REFERENCES

Blanchard, O. (1997): *The Economics of Post-Communist Transition*, Oxford: Oxford University Press

Chenery, H.B. (1960): Patterns of Industrial Growth, *American Economic Review*, 50 (September): pp. 624-654

Chenery, H.B., and Syrquin, M., (1989): Patterns of Development, 1950-1983., *World Bank Discussion Papers*, WDP41, March 1989

Družić, I., (2000): Makroekonomska ograničenja tranzicijske strategije, *Ekonomski pregled*, 51 (7-8): pp. 613-632

Družić, I., (2003): *Hrvatski gospodarski razvoj*, Zagreb: Politička kultura

Družić, I., (ed.), (1997), *Razvoj i tranzicija hrvatskoga gospodarstva*, Zagreb: Politička kultura

DZS: *Statistical Yearbook of Croatia*, various years.

Gregory P., Stuart R. (2000), *Russian and Soviet Economic Performance and Structure*, Addison Wesley

Gros, D., Suhrcke M. (2000): Ten years after: what is special about transition countries?, *EBRD Working paper*, WP56, August 2000.

Kuznets, S. (1966): *Modern Economic Growth*, New Haven: Yale University Press.

Svejnar, J. (2002): Transition economies: Performance and Challenges, *Journal of Economic Perspectives*, 16 (1): pp. 3-28

World Bank, (2002): *Transition – The First Ten Years: Analysis and Lessons for Eastern Europe and the Former Soviet Union*, Washington D.C.: The World Bank

STUDIA UNIVERSITATIS BABEŞ-BOLYAI, OECONOMICA, L, 2, 2005

COMPARISON OF CROATIAN AND EUROPEAN TAX AND ACCOUNTING REGULATIONS

IVANA DRAŽIĆ LUTILSKY*

**University of Zagreb, Croatia, Graduate School of Economics and Business, Phone: 00 385 98 43 03 11, e-mail address: idrazic@efzg.hr*

ABSTRACT. In every country, accounting and tax system are connected. Harmonization of Croatian tax regulations with the tax regulations of the European Union demands certain adjustments of accounting system. Tax system should be transparent and simple to enable accounting system in providing all the necessary information. Harmonization of Croatian tax and accounting systems opens up numerous opportunities to Croatia in achieving free movement of capital and competitiveness on the European market. The aim of research is to determine Croatian potential position in tax and accounting regulations in comparison with the regulations in the European Union.

JEL Classification: M41

Keywords: accounting, taxation, harmonization, Croatia, European Union

1. Introduction

"In order to join European Union, every country must fulfil certain conditions and criteria among which is the adoption of the *acquis communautaire*."[1] They must adopt, implement and enforce all parts of the *acquis* in order to be allowed to join the EU. The *acquis* includes all primary legislation (treaties), secondary legislation (Regulations, Directives, Decisions, Recommendations, etc.) and case law (judgments of the European Court of Justice and European Court of First Instance). Hence, since EU rules are constantly changing (new Directives are enacted, Regulations are amended, judgments are handed down), the *acquis* is not a static document, but one that is in constant evolution.[2]

For enlargement negotiations, the *acquis* has been divided into 31 chapters.[3] Chapter 5 (Company Law) has greatest relevance to private sector accounting and auditing; Chapter 2 (Freedom of Movement for Persons), Chapter 3 (Freedom to Provide Services), Chapter 4 (Free Movement of Capital), and Chapter 6 (Competition Policy), also have implications for private

[1] Source: www.mei.hr; (official site of Ministry for European integrations visited on 16 of April 2004.

[2] Gielen F., Barros Hirata A.C.,:„Corporate sector accounting and auditing within the Acquis Communautaire“ Draft for discusion only, February 8,2005.

[3] A summary of each chapter of the acquis, entitled, “Enlargement of the European Union Guide to the Negotiations Chapter by Chapter,” is available at: http://europa.eu.int/comm/enlargement/negotiations/chapters/negotiationsguide.pdf

sector and auditing. When applying for membership, an applicant country will receive a roadmap from the European Commission tracing its progress in adopting the *acquis*. Accession negotiations may be concluded even if the *acquis* has not been fully adopted, as transitional measures may be introduced after accession. However, transitional periods must be as short as possible and cover as few sectors as possible.[4]

Signing the Stabilization and Association Agreement Pact, Croatia has committed on adoption, implementation and enforcement off all parts of the *acquis* in order to be allowed to join the EU.

Regarding the accounting and auditing regulations, harmonization is achieving a lot more success then harmonization of tax regulations. Especially after June 2000, when the Commission adopted a Communication (COM/2000/359) entitled "The EU's Financial Reporting Strategy: The way forward." In it, the Commission suggested that all EU companies listed on a regulated market (i.e., publicly traded companies) should be required from 2005, to prepare their consolidated accounts in accordance with International Accounting Standards (IAS). A single set of standards would make it easier to compare companies' performance figures and raise capital, as well as provide better protection for investors. Member States would also be allowed to extend the application of IAS to unlisted companies and to individual accounts. In addition, the Communication stated that a proper enforcement infrastructure was to be developed to ensure that accounting standards are applied in the same way in all Member States. The Communication also called for the modernization of the Accounting Directives.[5]

Subsequent to this Communication, Regulation (EC) No 1606/2002 of The European Parliament and Council was issued, which requires listed companies, including banks and insurance companies, to prepare their consolidated accounts in accordance with International Accounting Standards beginning 2005.[6]

Harmonization of tax regulations between Member States will be difficult to endorse. Some economics even consider standardization rather then harmonization because the tax regulations are specific within Member States.[7] Those specifics are a result of national differences. In the last few years, trends regarding tax reforms in the European Union (meaning here

[4] Ibid.

[5] Gielen F., Barros Hirata A.C.,:„Corporate sector accounting and auditing within the Acquis Communautaire" Draft for discusion only, February 8,2005.

[6] Ibid.

[7] Communication from the Commission to the Council "Towards an internal market without tax obstacles:A strategy for providing companies with consolidated corporate tax base for their EU-wide activities" (com(2001.)582 final), 23. October 2001., www.europa.eu.int; official site visited 15. September 2003.

on the Council of the European Union) are deduction in tax burden and in increasing tax relief. That reform also means less tax income in States budget. Many of Member States are fighting with unemployment and social contributions which are just too high, so it is difficult to believe that they will even consider in gathering less budgetary revenue.

1.1. Harmonization of direct and indirect taxes

Harmonization of Croatian tax systems with tax systems of the European countries has started after Yugoslavia break down. Institutional tax framework in Croatia was made according to German tax system model.

Because of that reform Croatian tax system is more or less harmonized with EU demands regarding Directives and Communications which are related to taxation. The situation is similar in all 25 Member States and each of them has some specific characteristic in their own tax system so it is very hard to even speak off complete harmonization.

Harmonization of Croatian tax regulations is in its final faze and still there are some differences regarding indirect taxation.

Individual income tax regulation is in compliance with regulations of Member States, except because of the war there are some specifics in tax relief (incentives) like non – populated areas or regions under special government protection like Vukovar.

Table 1.

Individual income tax brackets and tax rates

Tax rates	Tax brackets	
	Month (HRK)	Year (HRK)
15%	till 3.200,00	38.400,00
25%	4.800,00	57.600,00
35%	14.400,00	172.800,00
Sum	22.400,00	268.800,00
45%		

Since, there was no similar situations in European countries it is logical that there is none specific tax relives.

Regarding corporate tax regulations the situations is similar. For regions under special government protection there is reduced tax rate and the purpose is to boost production and employment. Also the Croatian tax knows employment incentives and occupational rehabilitation and employment of disabled persons. All of that is because of the specific situation that Croatia was in – war. The standard corporate income rate is 20%.

Value added tax is partially harmonized between countries and that is because of the VI. Directive 2001/41/EC, which was bring by the Council

of the European Union. Hence that, harmonization of value added tax regulations among Member States is possible.

Croatian value added tax regulations or it's better to say institutional framework for VAT is definitely harmonized with the Council VI. Directive. The main differences are some exemptions which include services of banks, savings and loan institutions, and insurance and reinsurance companies, renting of housing premises and medical care and exc. The standard VAT rate is 22%. The zero rate applies to bread and milk, certain types of books, scientific journals and medicines. That is also different because Member states (some of them) have more than 2 rates and they include preference rates.

The greatest challenge for Croatia will be harmonization of excise duties. The reasons are lower rates and the categorizations of products which are taxable aren't harmonized with Member States (especially alcohol and mineral oils). Also Croatia need better administration for excise duties and for better application of regulations and better control.

1.2. Harmonization of accounting regulations

Accounting Law is the basic institutional framework which regulates accounting area. In accordance with Accounting Law, companies are obligated to use International Accounting Standards and International Financial Reporting Standards. The proposal of new Accounting Law, which is now under Government consideration and soon it will be adopted, has include all Directives, Regulations, Decisions, Recommendations, Interpretative Communications and Commission Comments. This is all part of *acquis communauitaure.* In regards, Croatia has to enhance institutional framework for consistent applying and interpreting IFRS.[8] According to the new Accounting law Croatian accounting standards board[9] stops existing and his role should take Financial Reporting Council.[10] That body will be established by Croatian Government. Also, regarding Croatian National Stabilization and Association Agreement Pact, they are obligated to prepare and develop Professional Accounting Licence Programme.

Audit regulations are mostly in compliance with *acquis communauitaure.* But nevertheless, it is important to enhance internal audit but also "super audit" meaning on the control of audit companies. Also, regarding Croatian National Stabilization and Association Agreement Pact, they are obligated to prepare and develop Professional Auditing Licence Programme.[11] Croatia

[8] Source: www.mei.hr; „National Stabilization and Association Agreement Pact"(official site of Ministry for European integrations) visited on 16 of April 2004.
[9] His role was to interpret and translate IAS.
[10] Ibid.
[11] Source: www.mei.hr; „National Stabilization and Association Agreement Pact"(official site of Ministry for European integrations) visited on 16 of April 2004.

is also obligated to prepare Code of ethics for professional auditors and accountants in accordance with IFAC Ethic Code.

The mainstream and trend in EU is that all countries are leaving (especially transition countries) the concept of Accounting Law and preparing for SME's National Accounting Standards. Croatian accounting regulations is still based on Accounting Law (by the Law everybody are obligated on IFRS) and it is very difficult for SME's to apply all the IFRS. So, Croatian SME's are mostly using Corporate income tax Law and VAT Law because it is simpler and then they are not obligated to prepare two sets[12] of financial statements but only one which can be used in both purposes: state purpose and owner purpose.

1.3. Significant results

When we are talking about tax systems of Member States and harmonization of the same, we can come to conclusion that there are some similarities between countries and that each and every one tax system has some specific characteristics which can be observed in that specific country. So, there can't be just one tax system that could apply on every country but there could be only some guidelines regarding rates, taxable income and transactions and maybe on double taxation treaties. So, what is the main reason for the harmonization in tax regulations of the EU (meaning the European Institutions)? Is behind all of that political power and desire in justification of existing institutional labyrinth and bureaucracy in the EU or better yet, Brussels? If the objective isn't harmonization we all wonder how they would justify existence of all advisories and technical groups in which are mainly experts working for big salaries.

United States of America are functioned just fine even though they don't have harmonized and comparable tax systems among States, not to mention VAT which is a precondition even to become a candidate for membership.

So, the process of harmonization is not completed in Internal Market of the EU. Croatia and every other country applying for membership can only try to harmonize all legislation with *acquis communauitaure* because some Directives and Recommendations are not applicable even in all Member States (they are still not implementing the VI. Directive and some recommendation regarding corporate income tax). Tax regulations of Member States are just partially harmonized. Comparing tax and accounting regulations Croatia is definitely keeping up with all actions and happenings on the European scene in the last ten years.

But in Croatia is still important political corruption which is partially breaking country progress in socio-economics, political and market economy sense. The reason for this statement can be found in correlation between

[12] Dual concept of financial reporting.

political elections and enhancing the tax incentives because in that way they are manipulate with voters.

Tax regulations should represent the economy in some countries and not the political idea if ruling party. Stimulating with tax regulations they should prevent tax evasions (“grey economy”) and they should improve free movement of capital and make sure that Croatian corporations are competitive in Internal Market even before becoming a Member State.

2. Conclusion

Croatia is one of those countries which is conceptually using dual concept if financial reporting. Nevertheless, tax regulations are much more detailed then accounting regulations and SME's are rather using tax regulations when they are preparing financial statements. So, we can say that tax regulations are dominating over accounting regulations. So, it is the author opinion that it would be better for Croatian SME's to start working on National Accounting Standards instead making such authority out of IFRS.

Also, Croatia is trying to stimulate investors and to draw foreign capital by eliminating withholding tax or better yet by not paying tax (corporate or individual) on dividends and shares. Taking in consideration tax regulations in Member States and wider that definitely isn’t a practise because in this way Croatia is giving up a part of budgetary revenue.

BIBLIOGRAPHY

Gielen F., Barros Hirata A.C. (2005): Corporate sector accounting and auditing within the Acquis Communautaire, *Draft for discusion only*, February 8

IBFD (2004): *European tax handbook 2004*. Amsterdam: International Bureau of Fiscal Documentation.

Sixth Council Directive (2001), 41, EC

Fourth Council Directive 78/660 EEC, CONSLEG: www.fee.be;

Seventh Council Directive 83/349/ EEC, CONSLEG: www.fee.be;

Structures of the taxation systems in the EU; www.europa.eu.int .

Chairas, I.Y., Wirawan E. D. (2002): *Accounting harmonization in ASEAN 2002*. www.lexusnexus.com

Messere, K.C. (1998): *The Tax System in Industrialized Countries*, Oxford University Press

www.mei.hr; *National Stabilization and Association Agreement Pact*, (official site of Ministry for European Integrations) visited on 16 of April 2004.

Communication from the Commission to the Council, (2001): Towards an internal market without tax obstacles:A strategy for providing companies with consolidated corporate tax base for their EU-wide activities" (com(2001.)582 final), 23. October 2001., www.europa.eu.int; official site visited 15. September 2003.

STUDIA UNIVERSITATIS BABEŞ-BOLYAI, OECONOMICA, L, 2, 2005

HIERARCHY OF DECISION MAKING PROCESS

MARIA PARLIŃSKA*

**Warsaw Agricultural University, Poland,*
Department of Econometrics and Computer Science,
Faculty of Agricultural Economics Nowoursynowska
street 166, 02-787 Warsaw, tel/fax +48-22-5934129,
e-mail: parlinska@alpha.sggw.waw.pl

ABSTRACT. AHP method is useful in solving problems connected with choose of the best solution, especially when the choice of certain solution is based on subjective preferences of examined group. In the article is shown the using of the method for choosing the best source of information on the wholesale market.

JEL Classification: C44

Keywords: decision making, AHP

Hierarchy **-** it is organization of data or processes according to the relative importance of elements such as in agent-principal sample. This system is extremely popular and useful in databases held in computer systems to project an authentic dependencies being created in the network.

The most popular computer hierarchies are file system hierarchy standards (FHS) and hierarchy of the network addresses on Internet. The definition was composed on the bases of the dictionary: Płoski Z. *Słownik Encyklopedyczny – Informatyka* published by Europa, 1999.

Hierarchical data model, broaden *flat-file database concept* by records' types and subordination relations. The definition was composed on the bases of the dictionary: Płoski Z. *Słownik Encyklopedyczny – Informatyka* published by Europa, 1999.

Hierarchical, related to hierarchy, for example hierarchical system of files or hierarchical system of transactions. The definition was composed on the bases of the dictionary: Płoski Z. *Słownik Encyklopedyczny – Informatyka* published by Europa, 1999.

AHP (Analytical Hierarchy Process)

AHP method is useful in solving problems connected with choose of the best solution, especially when the choice of certain solution is based on subjective preferences of examined group.

Defining the magnificent criteria for the certain group of users is the first step of the analysis. It is followed by identification of their weight. Usage of the AHP method leads to ranking of importance of certain criteria. That allows choosing the solution, which fulfils the most important criteria on the highest level.

The theoretical sense of the AHP method can be presented as followed:

There is known matrix (n x n) of criteria' preferences

$$A = [a_{ij}]$$

with diagonal with value 1, characterised by

$$a_{ji} = 1 / a_{ij}$$

Vector of priorities:

$$\begin{bmatrix} w_1/w_1 & w_1/w_2 \ldots & w_1/w_n \\ w_2/w_1 & w_2/w_2 \ldots & w_2/w_n \\ w_n/w_1 & w_n/w_2 \ldots & w_n/w_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_n \end{bmatrix} = n \begin{bmatrix} w_1 \\ w_2 \\ w_n \end{bmatrix}$$

The principle of determination of the vector of priorities is to consider each element of the matrix of criteria' preferences

$$a_{ij} = \frac{w_i}{w_j}, w_i, w_j$$

"W" – It is the weight of the certain estimation relative to i^{th} and j^{th} criteria. The demand of transitivity is fulfilled by cohesion of estimators. It means: if the element "i' is preferred against "j" and element "j" against "k", element "i' is preferred against "k'.

$$Aw = nw$$

$$Aw = \lambda w$$

$$(A - \lambda I)w = 0$$

$$\det(A - \lambda I) = 0$$

$$\lambda_{max} = n$$

$$CI = \frac{\lambda_{max} - n}{n - 1}$$

$$CR = \frac{CI}{R}$$

The indicators of random conformities are as follow:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
R	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.48	1.56	1.57	1.59

Saaty's algorithm can be described in few steps:

(1) The results of estimation of criteria are put into matrix A

(2) The estimators a_{ij} are summed for each column in the matrix A

$$k_j = \sum_{i=1}^{n} a_{ij}$$

(3) Standardisation:

$$b_{ij} = \frac{a_{ij}}{k_j}$$

(4) Vector of part sums:

$$s_i = \sum_{j=1}^{n} b_{ij}$$

(5) Weights:

$$w_i = \frac{s_i}{n}, i = 1,\ldots,n$$

Correspondence analysis of weights

(6) Aw=n*w

(7) u=A*w

$$u_i = nw_i, i = 1, \ldots, n$$

(8) $u_i = \lambda_i w_i$

λi is calculated

$$\lambda_{\max} = \frac{1}{n}\sum_{i-1}^{n} \lambda_i$$

Elimination of non-correspondence should be done. The matrix of preference is raised to the high enough power to ensure the correspondence of the estimators in order to stabilize the value of vector of priorities with assumed in advance toleration.

If deviation between ***λ****.max* and value n estimated by CR pass the possible limits, the estimation of validity of criteria must be done again. T.L.Saaty recommends giving to the critical value CR value 0.1. If CR is less or equal 0.1, it can be said that correspondence exists. If CR is more than 0.1, so then the analyse must be repeated***.**

Example:

The choice of the best source of information

Assume that there are 4 criteria of information: TV, radio, Internet, daily newspaper. Rules of preference and scale of marks or preferences (dominations) are as follows:

1 – In case of lack of domination, when criteria have the same importance
2 – In case of slight advantage of one of the criteria against the second one
3 – Reasonable domination,
4 – Reasonable strong domination,
5 – Magnificent domination,
6 – Strong domination,
7 – Very strong domination,
8 – Very strong but not absolute domination
9 – Absolute domination.
For example:

- If the information received from TV are considered as more important then those received from the radio on the "reasonable" level, value of preference is "3"
- If the information received from TV are considered as more important then those received from the daily newspaper on the "highest" level, value of preference is "9"
- If the information received from the radio are more important then received from the daily newspaper on the "high" level, value of preference is "7"
- The values of preference in opposite relation are reciprocal to calculated numbers

The matrix of importance of criteria (preferred sources of information) is:

	TV	Radio	Internet	Daily newspaper
TV	1	3	9	7
Radio	1/3	1	5	7
Internet	1/9	1/5	1	5
Daily newspaper	1/7	1/7	1/5	1
Total	1.6	4.34	15.2	20

The rules of setting the importance of criteria are taken into consideration:

- Sum of values of preferences in each column is calculated (2)
- Each value in the column of the matrix of importance of criteria is dived by the total value (3)
- The average of the row as the weight of particular criteria is calculated (5)

If heuristics mentioned above are used, for given example, the matrix with following normalised values will be received:

	TV	Radio	Internet	Daily newspaper	Weight
TV	0.625	0.691	0.592	0.350	0.565
Radio	0.209	0.230	0.329	0.350	0.280
Internet	0.069	0.046	0.066	0.250	0.108
Daily newspaper	0.089	0.033	0.013	0.050	0.047

The average values in particular rows of normalised matrix give the approximate characteristic vector of weights' values for particular criteria:

w = (0,565;0,280;0,108;0,047).

In this example was shown, that the most important as a source of information was TV, when daily newspapers have the smallest weight value, so the importance of that source of information is very small.

REFERENCES

Gruszecki T., (2001): *Nagrody Nobla w ekonomii*, Wydawnictwo Verba, Lublin

Parlińska M., (2002): *Dystrybucja informacji w wirtualnym środowisku*, Wydawnictwo SGGW 2002

Saaty T.L., (2004): *http://www.katz.pitt.edu/fac_pages/Saaty.htm*, marzec

Wellisz S., (1997): Gdy Brak Pełnej Informacji, *Wiedza i Życie*, nr 2

THE TECHNOLOGICAL DYNAMISM OF CHINESE MOBILE COMMUNICATIONS

JING CAI*

**University of Aberdeen, United Kingdom, Business School, Edward Wright Building, Aberdeen AB24 3UD, e-mail: Jing.cai@abdn.ac.uk*

ABSTRACT. The development of technological capability in Chinese mobile firms is currently by no means as rapid as the development of the Chinese economy, but in the long run will largely determine its pace and nature. This paper, from a historical perspective, presents development of Chinese mobile communications service sector. It tackles two key issues – regulation and finance which inhibit mobile carrier's technological dynamism.

JEL Classification: O14, O33, O53

Keywords: mobile communication, technological dynamism, China

1. Introduction

For the last decade the Chinese telecommunication industry has undergone rapid growth with an average annual rate of 37%, far above the GDP growth rate for the same time period (see Figure 1).

From September 2002 China has been the world largest telecom service market in terms of both mobile subscribers and fixed-line users. By the end of 2003 the nation's average penetration rate for fixed-line has reached 21.2% with total subscribers of 263 million.

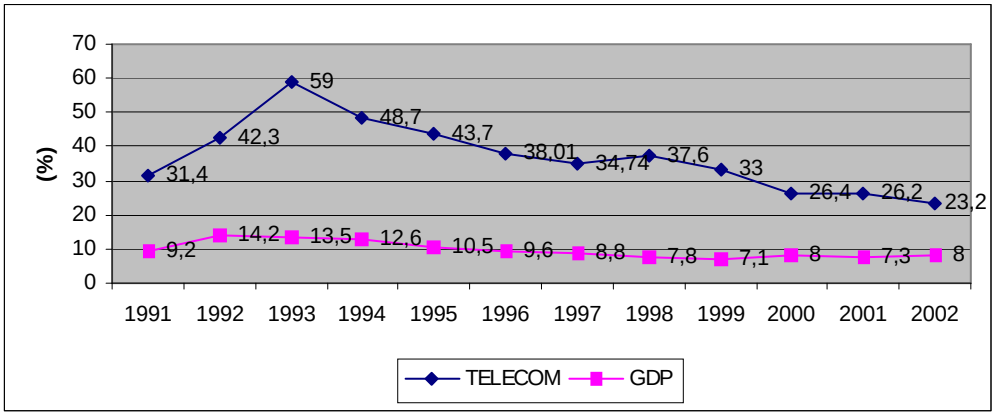


Figure1. Comparison of growth rate between telecom service sector and GDP

Source: www.mii.gov.cn

The penetration rate for the mobile phone has reached 20.92% with total subscribers of 269 million (see Figure 2). The network capacity for the fixed line by 2003 was 363 million and the capacity for mobile services was 336 million. The total length of optical cable was 2.71 million kilometres (www.mii.gov.cn).

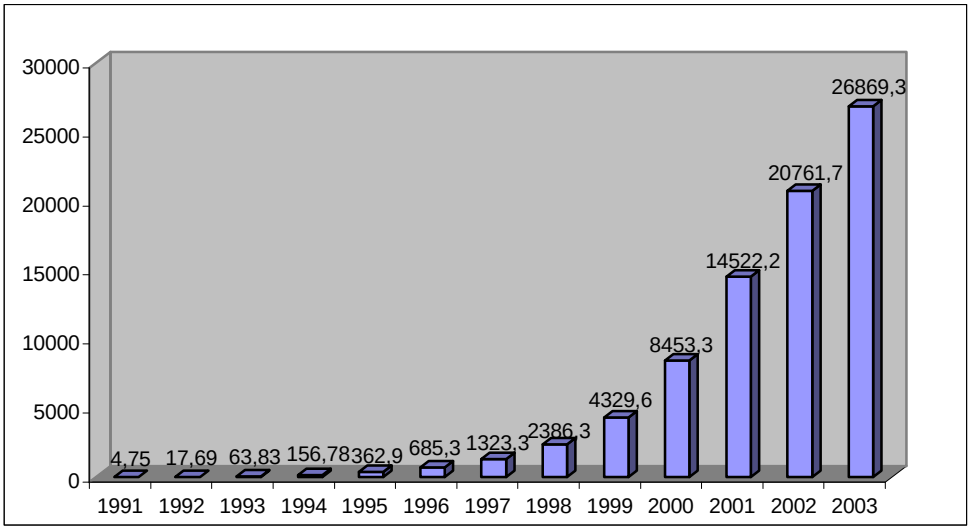


Figure 2. The total number of mobile subscribers (1991-2003) (unit:10,000)
Source: www.mii.gov.cn

A key issue for Chinese economy is how far its strategic industry - telecommunications (particularly mobile sector) will continue to focus on relatively high-technology development, taking a strategy of a rapid move 'up-market' in terms of technological dynamism. It is widely agreed that one of the main obstacles to such a move up-market is the country's system of regulation and finance. In China's drive to catch up with the developed countries, the state has shown its will to take the strategy of promoting the state-owned economy as its main vehicle in its strategic industries such as mobile communications. Tylecote and Cai (2004), and Cai and Tylecote (2005) have shown that the failure of most Chinese State-Owned-Enterprises (SOEs) to develop dynamic technological capability can be well explained by deep-seated faults in their corporate governance.

In a recent empirical work Cai studied the effects of relationship with government in mobile equipment sector. It was found that in the Chinese context, the relationship with government rather than the corporate governance or ownership type in purely quantitative terms, determines firm's technological development. In this article the same pattern can be traced in the mobile

service sector. Section 2 first introduces restructures of the regulation system in the Chinese telecommunications industry in which the mobile sector operates. Then it proceeds to reforms in financial system. Section 3 discusses the effect of relationship with government on mobile carrier's technological development. Section 4 is the conclusion.

2. Historical development of Chinese mobile communications

The Chinese mobile service was first introduced in 1987 in Guangdong Province by the then Ministry of Posts and Telecommunications (MPT). At that time, mobile services were analogue but, starting from 1994, China Telecom (the first mobile carrier) established an extensive GSM digital mobile communications network and began to expand its digital service. China Unicom (the second mobile carrier) which was established in 1994 also began to construct a nation wide GSM system. In 1997 digital subscribers surpassed the number of analogue users and the analogue system was finally abandoned in 2001 (http://www.oecd.org/dataoecd/52/27/2500835.pdf). In 1999 China Unicom began building a CDMA1X network and pilot services were offered in Beijing, Shanghai, Xian and Guangzhou in 2001. The deployment of the nation wide commercial services started from January 2002. By the end of 2003, CDMA1X subscribers reached 16.9 million. As the only competitor, China Mobile launched GPRS in May 2002. Within two months the GPRS subscribers have reached 1.3 million (http://tech.sina.com.cn/it/t/2002-07-11/125383.shtml).

In order to explore fully the potential of network resources and to generate more revenue from current subscribers, both China Mobile and China Unicom have introduced a variety of value-added services in the past few years, which include SMS and Voice-over-IP long distance calls. In an effort to keep up with mobile commerce developments world wide, both China Mobile and China Unicom launched WAP services on May 17th 2000. Services now available through WAP include mobile banking, stock trading, news, weather reports and e-mail (Wu 2001).

From 2001 mobile business has surpassed the fixed-line business. In 2002 mobile services shared nearly half of the total telecom revenue and became the major revenue contributor (See Table 1).

Table 1:

The proportion of telecom business revenue

Year	Local fixed-line business	International fixed-line business	Data communication business	Wireless communication business	Others
2000	47.3%	4.13%	2.39%	41.69%	4.49%
2001	41.47%	2.56%	3.71%	45.74%	6.52%
2002	41.1%	2.1%	4.4%	47.5%	3.9%

Source: www.ccidconsulting.com

Despite all these achievements, in general, the mobile penetration rate in China is still low compared with those in advanced countries, which have an average penetration rate of more than 50%. Apart from that, the unbalanced penetration distribution along different geographical locations reflects the unbalanced development of the Chinese economy. This imbalance exists not only between developed or less developed regions (eastern vs central or western regions) but also within the rich regions. By 2003 the average penetration rate of eastern regions (27.61%) was about three times those of the central (10.9%) and western regions (10.91%). However in 2002 even within the eastern region, some cities have extremely high penetration rates such as Guangzhou 83.89%, Shanghai 60% and Beijing 55.47% (China Statistical Year Book 2003).

Reforms in this sector start off with restructures in regulations, which can be roughly divided into four phases.

2.1. Reforms in regulatory structure

Phase 1: before 1994:

Prior to 1994 the construction of a basic telecommunications network and the provision of telecommunications services were all under the unified management of the central government through the Ministry of Post and Telecommunications (MPT).

At the central level the State Council together with a number of other authorities has very wide discretionary powers to formulate and implement regulations. The State Council has the supreme administrative right in the setting of national policy. Any important strategy and policy made by the MPT should be approved by the State Council. The State Development Planning Commission (SDPC) is the largest comprehensive organisation that is responsible for the development of economic, scientific/technological and social development strategies, as well as key economic technology policies. In addition, the Commission formulates national economic plans and provides financial support for large infrastructure projects, including those relating to telecommunications. In general, projects exceeding CNY 30 million must first get the approval from the SDPC (Wu 2000, OECD 2003). The Ministry of Commerce is in charge of regulations related to technological import/ export policies, goods as well as their quota and licences (OECD 2003). The Ministry of Post and Telecommunications (MPT) was the sole provider of public telecommunications services in China up until then. It was a functional organ under State council that was responsible for posts and telecommunications policies, regulations, technical standards, management planning and operations. Under the MPT, the Directorate General of Telecommunications (DGT) took charge of the telecommunications industry (www.mii.org.cn).

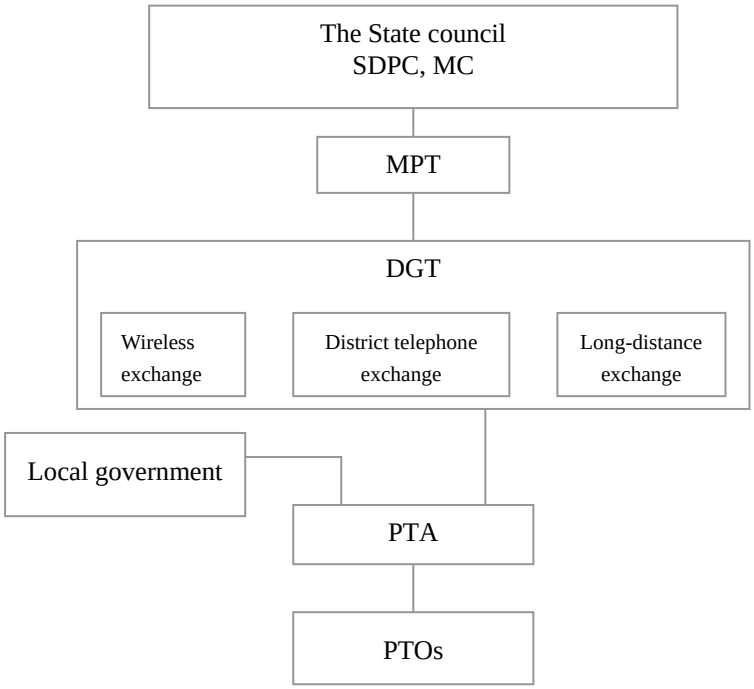


Figure 3. The Chinese telecom regulatory structure (before 1994)

Sources: OECD 2003, China Academy of Social Science 1953
SDPC: State Development Planning Commission
MC: Ministry of Commerce
MPT: Ministry of Post and Telecommunications
DGT: Directorate General of Telecommunications
PTA: Post and Telecommunications Administration
PTO: Post and Telecommunications Office

At local level, the Posts and Telecommunications Offices (PTOs) were the operating organs that provided mobile services. In each province there was a PTA who administrated PTOs within that province. Local government had supervision rights over the performance of PTAs, and could also inquire into such issues as development plans, telecom tariff and even staff wages of PTOs (Wu 2001, OECD 2003, www.mii.org.cn).

Phase 2: 1994-1997

Regulation reforms began in 1994 when the Chinese State Council finally realized the inefficiency and poor management of the MPT. New and higher requirements for the management of the service sector led to the separation of the governmental regulatory functions from that of the operating

functions. The DGT detached from the MPT and developed into a new type of state enterprise, which was called China Telecom (see Figure 4). China Telecom, the first and the dominant operator, engaged in international, inter-city and mobile communication services. Posts and Telecommunications Administrations operated the local services. In the same year China Unicom was established jointly by the Ministry of Electronics Industry (MEI), the Ministry of Electricity and Power (MEP) and the Ministry of Railways (MR). The main business of China Unicom was mobile phone services (Lee 1999, Wu 2000). Ji Tong has emerged as the third telecommunications operator in China. In September 1995, the People's Liberation Army (PLA) established the Ji Tong aiming to compete with the MPT in radio, paging and other value-added networks services by utilizing the facilities of its own networks (Wu 2001).

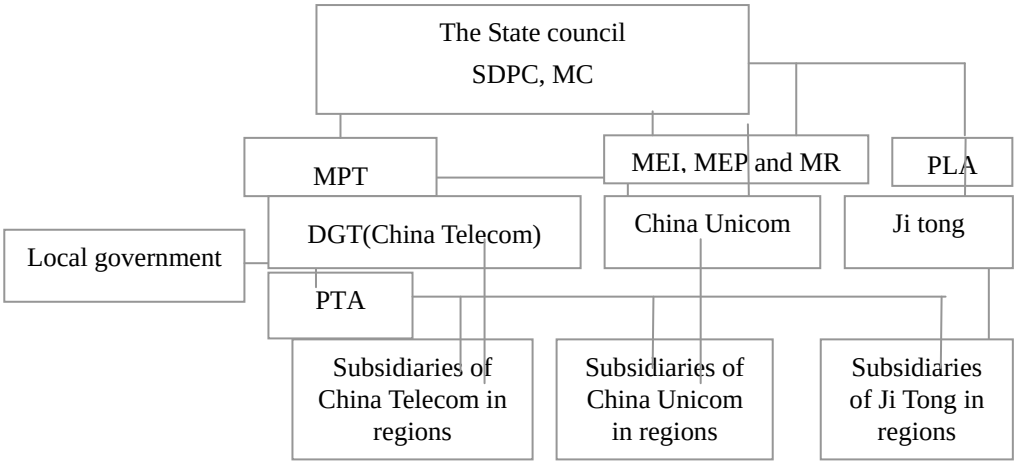


Figure 4. The Chinese telecom regulatory structure (1994 to 1997)

Sources: Lee 1999, Wu 2001
MEI: Ministry of Electronics Industry
MEP: Ministry of Electricity and Power
MR: Ministry of Railways

Phase 3: 1998-2002

In 1998 the state council integrated the former MPT, the MEI and part of the Ministry of Radio, Film and Television into a new Ministry – the Ministry of Information Industry (MII). MII is in charge of formulating telecom laws and regulations, managing mid-to long-term development strategies, allocating spectrum sources and investment, administering technology importation, utilisation and supervising telecom service carriers. After the

establishment of the MII, each provincial PTA was required to completely peel off its operating loads and change into a pure administrative organisation - Provincial Communication Administration (PCA). Its operating loads were taken by China Telecom and thus China Telecom became a nation-wide service carrier who had the sole responsibility for its own profits or losses (see Figure 5) (Wu 2000, www.mii.gov.cn).

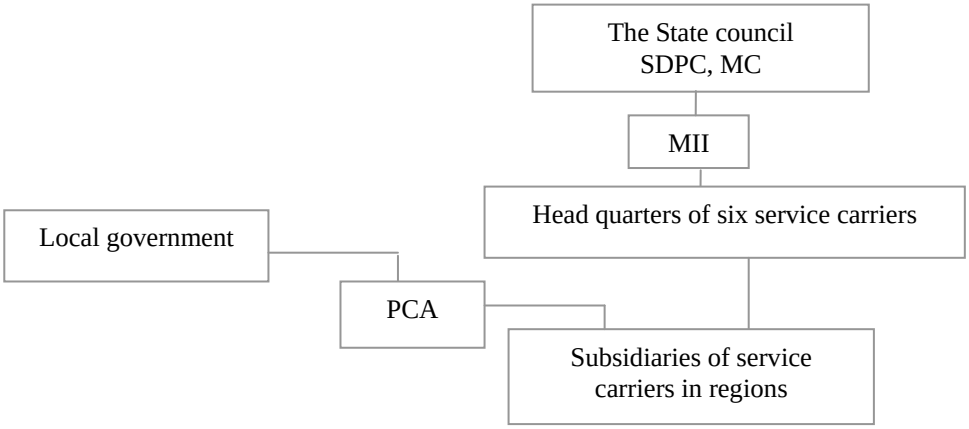


Figure 5. The Chinese telecom regulatory structure (1998 to 2002)

Sources: www.mii.org.cn

From 1999 to 2000, in the process of preparing the entrance into the WTO, there was a dramatic restructuring in the service market structure (see Table 2). In April 2000, the Chinese government broke up the former monopolist China Telecom into four divisions. Three of them become independent companies. The fourth division, Guoxin Paging (the former paging branch of China Telecom) was merged with China Unicom. The three other divisions of the split now are China Telecom, mainly operating fixed lines services and the data transfer networks; China Mobile Communications, operating mobile phones and the cordless transfer of data; and China Satellite operating satellite communications and high-speed Internet delivery (South China Morning Post 2002, May 20th).

At the end of 2000, two new companies entered the telecommunications service sector. One is China Netcom and the other, Railcom. The owners of China Netcom are the Chinese Academy of Sciences (CAS), the Ministry of Railways (MR), and the State Administration of Radio, Film and Television (SARFT). China Netcom is involved in Internet access, fixed-line telephone services and high-speed data transfer. By contrast, Railcom is backed by the Ministry of Railways with its independent communications network. Its main business is the fixed lines service, data transfer, paging and Internet telephone services (Wu 2001, People's Daily 2002, May 17th).

Table 2:
The market structure of Chinese telecommunications service sector

Year	Organisations/ Company	Owners/ Shareholders	Services	Market shares (2000)	Market shares (2002)
1994	Unicom	MII	Cellular phone service	7.7%	12.1%
2000	China Mobile	MII	Cellular service	34.5%	37.4%
2000	China Satellite	MII	Satellite communications and high-speed internet service	-	1.2%
2000	Railcom	MII	Fixed lines, data transfer, paging and Internet telephone services	-	
2002	China Telecom (new)	MII	Fixed line service in 21 southern and western provinces	57.7%	32.5%
2002	China Netcom	MII	Fixed-line service in 10 northern province; Internet access and telephone services and high-speed data transfer; Radio, paging and other value-added	-	16.8%

Sources: South China Morning Post 2002, May 20th, Wu 2001, www.ccidconsulting.com

In May 16th 2002, China Telecom has been split into the north and south sectors. The south sector retains China Telecom's activities in 21 southern and western provinces. Its activities in ten northern provinces were merged with China Netcom Group and Ji Tong by establishing a new company (China Netcom Group Corp.) (South China Morning Post 2002, May 20th). Currently there are in total 6 operators offering services from fixed lines services, cordless transfer of data to high-speed Internet delivery.

Phase 4: 2003 - present

In 2003 in the process of reforming SOEs, the state council defined 196 national enterprise groups (including all telecom service carriers) directly controlled by the central government (see Figure 6). A new organisation the State-owned Assets Supervision and Administration Commission (SASAC) was established and delegated as the owner of these national enterprise groups, to exercise shareholder's rights (URL: http://www.sasac.gov.cn/zyzz.htm).

2.2. Reforms in the financial system

The technological competence of Chinese mobile carriers is not simply a matter of the various organs of the state with responsibility for regulating them. It is also very much influenced by financial institutions and markets. Reforms in the methods of carriers' finance have undergone two stages: from heavily depending on government subsidies and bond financing, to the current stock financing.

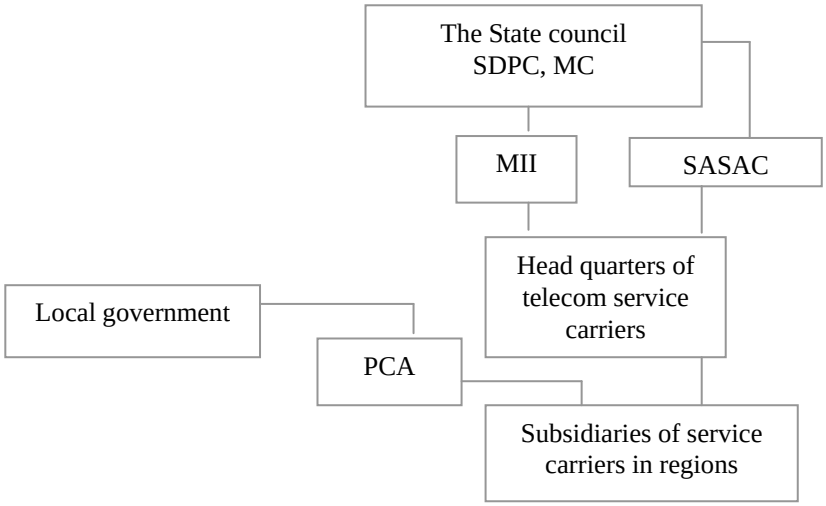


Figure 6. The Chinese telecom regulatory structure (2003 to present)
Source: http://www.sasac.gov.cn/zyzz.htm

Stage 1: subsidies and bond financing (1949-mid 1990s)

Before 1980 the bulk of investment funding for the telecom service sector came mainly from the government, with limited investment from the then Directorate General of Telecommunications (DGT). Over 31 years from 1950 to 1981, the total fixed asset investment was only CNY 6.4 billion Yuan (Wang 2001).

In the 1980s, the Chinese government put forward a series of aggressive plans for the development of the service sector. From 1984, the DGT (or China Mobile and China Unicom after 1994) could retain 90% of the total profits. In the same year the MPT adjusted the depreciation rate for telecom fixed assets. The average depreciation rate for telecom equipment was 20%. For some key equipment, the depreciation rate could reach 40 %[1]. Later, 90% of bank loans were exempted from the DGT (or China Mobile and China Unicom after 1994) during the time period of 1985 and 1999 (Yang 1993, Lee 1999, Wang 2001 and MII website: http://www.mii.gov.cn).

Charging a connection fee was another effective way to speed up the construction of the mobile networks. MPT allowed its PTOs to charge CNY 3000-5000 Yuan from 1992, which was about twice the average annul income of a city inhabitant. With the increase in mobile penetration, connection fees began to drop and were finally abolished in 2001. From 1992 to 1995, approximately CNY 62 billion Yuan was collected through the connection fee (Wang 2001).

[1] All these policies were abolished in the year 1995.

Apart from the mobile connection fee, another main source for mobile investment was the installation fee, which came from fixed-line business (Wang 2001). During 1986 and 1990, the total installation fee collected by the DGT was CNY 66 billion Yuan accounting for 32.7% of the total telecom investment in that period. During 1991 and 1995, the total installation fee collected was CNY 1110 billion which accounted for 45.8% of the total telecom investment (see Table 3).

Table 3.

Percentage of installation fee to the total telecom investment in fixed assets (1986-1995)

Years	The sum of installation fee (CNY billion)	Total telecom investment (CNY billion)	% of installation fee of total telecom Investment
1986-1990	66	201.98	32.7%
1991-1995	1110	2423.89	45.8%

Source: Wang, 2001

Except for internal sources, external sources have been also encouraged since 1986. Though foreign investments were prohibited from the direct participation in the basic mobile services, they were participated in way of investment from foreign governments, World Bank and Asia Development Bank. Service carriers were allowed to buy telecom equipment with low-interest loans from the above sources, and customs tariff was totally exempted[2]. Foreign investments gradually played an important role, particularly when the connection fee (or installation fee) started to decrease or was eventually abolished (Wang 2001) (see Table 4)

Phrase 2: stock financing (mid 1990s to present)

As mobile services started to meet market demands, preferential policies in tax and bank loans began to fall. Some of them were terminated in the mid-1990s. Consequently, carriers started to shift from bond financing to stock financing. China Telecom (Hong Kong) Ltd. (the former China Mobile) was first listed on the New York Stock Market Exchange on October 22nd, 1997. On the following day the company was also listed on the Hong Kong Stock Market. USD 40 billion in total was raised in its first public offering (www.Chinamobile.com). On June 2000 China Unicom (Hong Kong) Ltd. was listed on both the New York Stock Market Exchange and Hong Kong Stock Market. USD 56.5 billion in total was raised (www.Chinaunicom.com). It was a very significant step which means that the Chinese mobile sector has begun to enter international capital markets for finance.

[2] This policy was abolished in the year 1996.

Table 4.
Sources of investment in fixed assets of the telecom industry1981-1995[3]
(Unit: CNY billion)

Years	Total	Government investment	Bank loan	Foreign investment	Investment by firms themselves	Others
1981	7.1	3.6 (50%)	0	0	3.5 (49.6%)	0
1982	9.2	3.6(39%)	0	0.1(0.5%)	5.5(59.9%)	0
1983	9.5	4.1(43%)	0.1(0.6%)	0	5.3(55%)	0.1(1%)
1984	13.5	6.4(47%)	0.4(2.7%)	0.1(0.8%)	6.6(48.8%)	0.1(0.4%)
1985	19.3	6.7(35%)	2.1(10.7%)	0.2(1%)	9.6(49.4%)	0.8(4.1%)
1986	24.5	6.6(27%)	3.1(12.5%)	0.4(1.6%)	13.1(53.6%)	1.3(5.3%)
1987	30.6	6.9(22%)	3.7(12.2%)	1.2(3.8%)	16.9(55.3%)	1.9(6.3%)
1988	36.7	4.5(12%)	5.7(15.5%)	3.1(8.5%)	21.8(59.4%)	1.5(4.2%)
1989	50.4	5(9.9%)	5(9.8%)	5.7(11.3%)	32(63.5%)	2.8(5.5%)
1990	59.8	4.8(8%)	3.9(6.6%)	6.6(11.1%)	40.2(67.2%)	4.3(7.2%)
1991	86.1	5.4(6.2%)	5.8(6.7%)	13.5(15.6%)	60.9(70.7%)	0.7(0.8%)
1992	162.5	4.8(2.9%)	10.9(6.7%)	21.5(13.2%)	122(75.1%)	3.3(2%)
1993	404.2	6.2(1.6%)	51.9(12.8%)	41(10.1%)	299.3(74.1%)	5.5(1.4%)
1994	775.8	4.5(0.6%)	51(6.6%)	124.4(16%)	576.4(74.3%)	19.5(2.5%)
1995	995.2	4.6(0.5%)	55.4(5.6%)	156.7(15.7%)	754.5(76.5%)	24(2.4%)

Source: Statistical Year Book on transportation, posts and telecommunications (1996)**.** Figures in the bracket are the percentage of individual item out of total investment in each given year.

3. Discussion

There are two challenges of technological change to the relationship with government: incentive within the structure of control and informal opportunities to gain. As introduced in section 2.1 both the regulatory organisation (the MII) and two service carriers evolved from the same system. There are various ties and connections among them. The MII hires, fires and promotes management in these two carriers. The top manager is a member of the political elite and can hope to be promoted into senior administrative positions for what is seen as good performance. Many managers only regard business as a route to a higher position. In these two carriers the relationship with the state 'controllers' generally leads management to look to external, often foreign sources for discrete packages of equipment and technology – which leaves them still dependent on external sources for the next 'upgrading', since they do not really acquire technological competence.

[3] Since the first development stage, the mobile sector hasn't been separated as an individual accounting unit in Chinese statistics, this table refers to the investment of the whole telecom service industry.

In order to strengthen the system of monitoring, the state took the strategy of establishing a special supervisory organ - the SASAC - rather than deregulating its power to MII. The efficiency of the new structure is very doubtful. Currently the SASAC is in charge of 196 national groups that cover almost every industry (for example, telecom, oil, steel, electrical power, automobile, spinning and wearing and even gramophone record and tourism). To supervise them effectively, it requires industry-specific expertise that will inevitably increase the costs of the regulation. It is also very doubtful whether the SASAC really could dismiss those top executives who perform badly, given that top executives have very high positions in the Chinese Communist Party (CCP).

Ambiguous legal provisions create informal social network '*guanxi*' which offers firms and managers the chance to override the government regulations. '*guanxi*' can be used by firms to gain new project permission and avoid time-consuming approval of the process; to escape some social responsibilities such as environmental issues; or to maintain industrial protection by keeping a good relationship with government authorities. The manager may simply use his *guanxi* with branches of state-owned banks to cover operating losses out of new loans; alternatively the money may be used for real investment capital. As a direct consequence two service carriers care little about the quality of their investments. Most money has been spending on duplicative and excessive constructions. Between Inner Mongolia and Baotou (north of China where the economy is less developed), there are 2 microwave channels, 1 optical cable, 1 coaxial cable, 1 balanced cable, 3 overhead cables and 3 sets of short wave communication equipment (Wang 2001). The utilisation rate of China Mobile' network in 1997 was just 63%. Some up to date information showed that the utilisation rate of transportation cables from Huhehaote, Inner Mogolia to Beijing was 58% and that for the cable from Beijing to Guangzhou was only 29% in 2003 (URL: http://www.cnii.com.cn/20030218/ca139073. htm). Up to 2003, China Unicom and China Mobile have borrowed USD57.55 and USD23.67 billion respectively from banks (www.chinamobile.com; www.chinaunicom.com), despite the fact that the cash flow in China Unicom has been negative up until the end of 2003 (www.chinaunicom.com). From 2000 on, China Mobile has spent an average of USD 25.93 billion per year on its network expansion and upgrading, compared with USD 19.02 billion spent per year by China Unicom (See Table 5).

Competition between these two carriers is simply based on price. As shown in Figure 7 and 8, ARPU (Average Revenue Per User) value for both China Mobile and China Unicom has been decreasing since 2000, despite the increase in MOU (Minutes of Use) value.

Table 5.
Annual investment on mobile network by China Mobile and China Unicom
(USD, billion)

Year	2000	2001	2002	2003
CM	17.92	21.51	37.14	44.65
CU	-	25.34	15.3	16.41

Source: www.chinamobile.com; www.chinaunicom.com

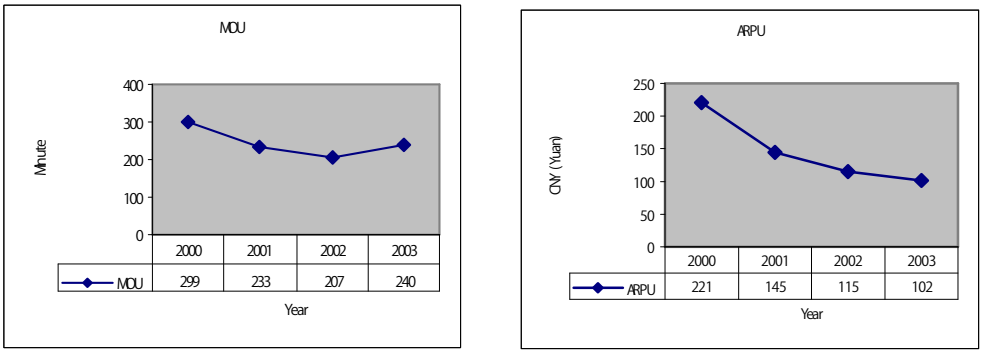


Figure 7. MOU and ARPU values for China Mobile from 2000-2003
Source: www.chinamobile.com

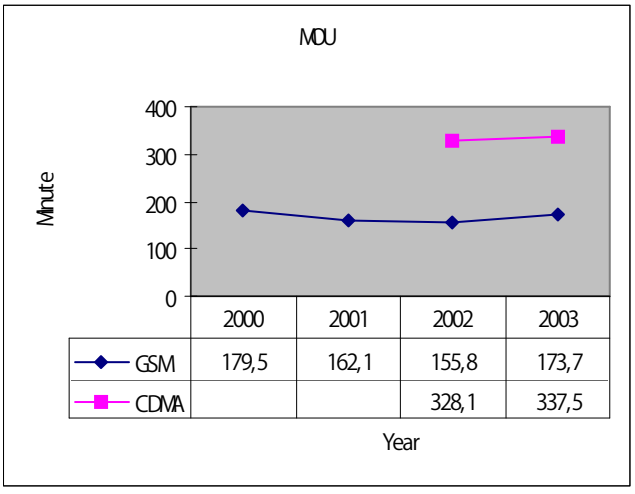
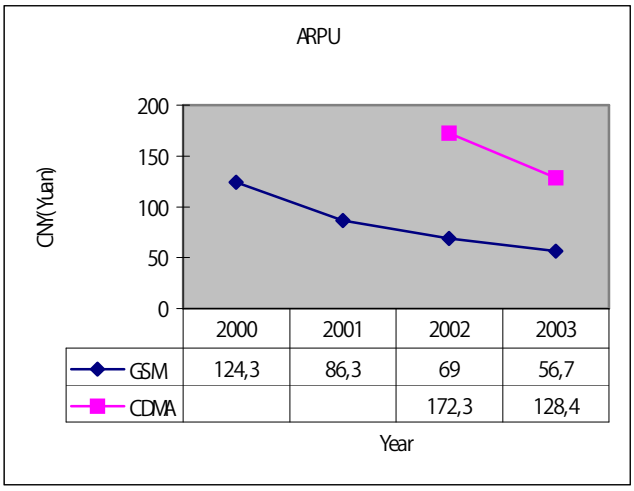


Figure 8. MOU and ARPU values for China Unicom from 2000-2003
Source: www.chinaunicom.com



Source: www.chinaunicom.com

Table 6.

Network capacity and utility rate of China Mobile and China Unicom

			2000	2001	2002	2003
Network capacity (million)	CM		110	159	182	222
	CU	GSM	19	23	44	73
		CDMA	-	-	4.5	27
Utility rate of network capacity	CM		59.7%	65.5%	75.7%	79.7%
	CU	GSM	66.66%	87.55%	87.23%	87.3%
		CDMA	-	-	57%	62.3%

Source: www.chinamobile.com; www.chinaunicom.com

Apart from this, revenue generated by these two carriers is mainly from voice services - calling and voicemails. In 2003 revenue generated from value-added services by China Mobile only accounted for 10.2%, and less than 5% of China Unicom's (www.Chinamobile.com; www.chinaunicom.com). Taking into account the continuous decrease in ARPU value and the reduction in revenue generated from value-added services, it is apparent that, in order to make a profit, the two carriers have to subscribe more and more potential customers through network expansion. Table 6 shows the network capacity expansion of these two carriers during the last four years. Within just four years the network capacity for China Mobile has doubled from 110 million to 222 million in 2003. By comparison the network capacity for China Unicom has increased almost four fold from 19 million to 73 million for their GSM system, and by around six times from 4.5 million to 27 million for their CDMA system. However by 2003, the utility rate of network capacity

for China Mobile was less than 80%, compared with 87.3% for China Unicom's GSM system and 62.3% for CDMA system (www.chinamobile.com; www.chinaunicom.com).

4. Conclusion

Findings

This paper has reviewed the historical development in Chinese mobile service sector, in particularly regulation structure and finance system. Although some progress has been made, two service carriers remain inefficient and unable to compete in a market environment. It has argued that the relationship with government, manifested through such factors as inadequate regulatory and financial mechanisms divert carrier's strategy away from anything like an optimal long-term policy for technological development. This paper has shown how this tends to distorts the upgrading of technology in such a way that the service carriers remain dependent on external sources.

Policy Prescriptions

What can be done to remedy the situation? Clearly the central state is at best a remote owner and as such has similarities to the "outsider" institutional shareholders of the US and UK. It has much to learn from these systems in containing the relationship with government.

For example, to increase the transparency of the regulatory system, it is requisite that the regulatory authority - MII - should be completely separated from the service carriers. Top managers of service carriers should come from the labour market rather than the MII. In addition the tenure of top managers should be left to the market since it is the way for service carriers to have sole responsibility for their own profit or loss. It is of equal importance that the enactment of laws will guarantee the fairness and justness of the regulatory process.

The current protection in mobile service sector has proved to be highly unsuccessful. Effective infant industrial protection requires a good understanding of the market situation and high skills of policy formulation. For Chinese government with little experience accumulated so far (of how to regulate a market economy), it is crucial that government can implement and correct policies objectively and impartially over time, in case of inefficiency. It involves substantial communication and cooperation among various government authorities. One of the best ways of offsetting some of the distortions that have been created by policy is vibrant competition. An alternative way to increase competition is to separate the ownership from the operation of the physical network. Put in another way, the ownership of the physical network could belong to the state while there can be several carriers which rent and run the network.

Research limitations and recommendations

Findings here are based on general literature and aggregate data rather than on specific case studies of service carriers. Such studies can be extremely useful - if they are directed at elucidating the links between relationship with government and technological development. Paired case studies of success and failure would be valuable.

REFERENCES

Cai, J. and Tylecote, A. (2005): A healthy hybrid: the technological dynamism of minority state-owned firms in China, *Technology Analysis and Strategic Management*, Vol.17, issue 3, forthcoming, 2005

China Statistical Year Book (2003) in State Statistical Bureau (Ed.), State Statistical Publishing House, Beijing.

Lee, Z. (1999). *Telecommunications industry in China: Its impact on the economic growth and democratisation*., URL: saturn.ewha.ac.kr/publication/thesis/thesis.htm

OECD (2003) *Review of the development and reform of the telecommunications sector in China*, Directorate of Science Technology and Industry/ Committee for Information, Computer and Communications, 6/FINAL, JT00140818.

Tylecote, A.B. and Cai, J. (2004). China's SOE reform and technological change: a corporate governance perspective, *Asian Business and Management*, vol 3, No.1: 57.

Wang, O. (2001): *The development of restructuring of Chinese telecommunication industry 1994-2001*. Ph.D thesis of Chinese Academy of Social Science.

Wu, J.C. (2000). *The development of China's telecommunications industry*, Xinhua Press, China

Wu, Y. R. (2001). *Growing through deregulation: a study of China's telecommunications industry*. URL:www.econs.ecel.uwa.edu.au/economics/Research/2001/DP%2001.21.pdf

Yang, T.F. (1993). *The current development of posts and telecommunications industry*, China present Press: 25-26, 30-31.