The Law of Large Numbers and its Visionary Returns to Scale in Insurance

by

Yuval Shilony

Department of Economics, Bar-Ilan University Ramat-Gan, Israel 52900

e-mail: shilony@mail.biu.ac.il

Abstract

It is generally construed in the literature of business, mathematics and economics that the source of the relative advantage of the insurance firm in bearing risks is the law of large numbers. This law provides the impetus of insurance production, it is argued, by affording returns to scale in risk bearing because large scale allows more accurate prediction of average loss and higher chance of firm survival. In conjunction with scale advantages regarding other types of insurance costs, found in studies, that suggests conditions of natural monopoly in insurance. Yet, in practice thousands of insurance firms coexist and compete. This work answers that apparent discrepancy by disputing this specious argument on empirical grounds. It concentrates on firm’s survival, without invoking utilities, using a simple model of insurance with binomial risk. The law of large numbers indeed affords returns to scale in survival but only in numbers too large to be practical. In the actual market for insurance there are negative returns to scale in risk bearing as each additional customer exacerbates the hazard of financial ruin of a firm. Letting the firm select its equity and reinsurance level optimally does not change the general result. The rationale for insurance production must be tracked in another domain.

Key Words: insurance, firm survival, law of large numbers, returns to scale, natural monopoly

JEL classification: D42  D80  G22

November, 2005
Visionary-CR3.doc
1. Introduction

Risk bearing is a burden for the risk-averse and the insurance industry produces relief services for its clients. Unlike a doctor who aims at relieving a sick patient by eliminating the disease, i.e.; curing him, the insurance firm can not ‘cure’ the risk-bearer but simply shifts the risk onto itself. The fundamental questions of how insurance actually works and what is the root of the relative advantage of the insurance firm in bearing risks for its customers have been universally answered by the law of large numbers.

Laymen and academics alike support this legend. It is propounded in business texts, e.g. Rejda (2003, Ch. 2), and in mathematics, e.g., Kneale (1949, p.141), Newman (1956). This view pervades the economics literature and has found its way into many microeconomics texts in chapters on uncertainty and insurance. The New Palgrave Dictionary of Economics maintains that the law of large numbers is the basis for the existence of insurance. It even finds it appropriate to devote space, in its three-page entry on insurance, for a proof of a version of this law; see McCall (1987).

To clarify, the argument is that by selling many insurance contracts the insurance firm faces less risk because on a large scale one can fairly accurately predict the average loss. The risk of insolvency is real and troubling for an insurance firm. Therefore enhanced confidence in its underwriting results affords the competing larger firm an advantage that ought to be reflected in lower prices. In other words, with regard to solvency there are returns to scale in insurance production.

The measurement of returns to scale in insurance is beset by conceptual and definition problems, e.g., what is the right variable to represent output. While Joskow (1973) claimed that production of insurance manifests nearly constant returns to scale, later studies found minor increasing returns to scale, see Doherty (1981) and studies of several European countries in the July 1991 issue of The Geneva Papers on Risk and Insurance. The typical study regresses the expense ratio, i.e., the insurance firm’s accounting costs excluding claims payments over premium revenue, on output. The slope is negative and highly significant. These studies do not consider the peril of insolvency and bankruptcy as a cost. This last cost however, is the essence of insurance production, and as
such deserves more scrutiny. It ought to be incorporated into the study of returns to scale.

Returns to scale on all counts, including insolvency and other costs, would suggest conditions of natural monopoly in insurance. The insurance industry however, is populated by many firms, around 3000 in the U.S., of various sizes and organization types, which coexist and compete. Although regulated, entry into the industry is rather easy. This work answers that apparent discrepancy by disputing the traditional argument on empirical grounds. The insurance firm is not in the business of predictions and forecasts; it seeks survival and profits. The issue becomes how larger numbers contribute to these goals.

Note first that the law of large numbers pertains to the mean risk and its variance, while an insurance firm is concerned with the variance of the sum of all the risks it bears, which gets larger as their number increases; see Samuelson (1963). An elaboration by Ross (1999) finds conditions on the decision-maker’s utility function that assure that when the number of the risks grows, their sum eventually reaches a point of acceptance even if rejected in small numbers. In actuality, the point of eventual acceptance may be beyond the reach of the firm’s shareholders and managers. This paper does not employ utilities and addresses only the returns to scale which affect survival and cost.

As shown below, the law of large numbers affirms the eventual decay of the risk of insolvency. It is claimed, however that this result is of no consequence to the firm because at the going price of insurance the number of policies implicated is much too large to be relevant. In other words, given the demand for insurance, the law of large numbers is beyond the reach of the insurance firm. Moreover, it will not help if all insurance firms merged together. There is simply not enough demand for insurance in the world.

Section 2 presents a simple binomial model. Section 3 demonstrates the argument for a numerical example. Section 4 generalizes the result by employing the normal approximation of the binomial. Section 5 shows rudimentary data and examines real-world relationships among the variables. They imply that larger numbers exacerbate a firm’s risk of ruin. In sections 2 through 5 equity is assumed fixed and given. In section 6 the firm selects its optimal equity and reinsurance level and its behavior is analyzed. Surprisingly, larger numbers now imply less output (retention), less equity and less chance of insolvency but more reinsurance
and higher average cost. If the insurance industry is a monopoly, and reinsurance is not available, the law of large numbers is still of no help in mitigating the firm’s average cost. Section 7 is the conclusion.

2. The Model

Assume an insurance firm sells only one type of policy against a risk of, say, fire. The damage is either zero or total loss of $D$ and occurs independently for each customer with the same known probability of fire $p$. In order to focus on the fundamentals, important insurance complications like moral hazard and adverse selection are assumed not to be present. $V$ is the equity, or surplus, of the insurance firm at the beginning of the period before its insurance transactions. If the firm sells $n$ such policies at price $s$ and happens to incur $k$ losses the value of its surplus at the end of the period after the full resolution of its risks is:

$$(V + ns)(1 + r) - kD$$

where $r$ is an average return on the firm’s invested funds until indemnities will have to be paid out since fires are distributed over the period. The life of the insurance firm is stochastic. If the realization of $k$ happens to be too large, namely if

$$k > \frac{(V + ns)(1 + r)}{D},$$

the net worth turns negative and the firm is declared insolvent.

The probability of insololvency of the firm is

$$Insolv(n, s, p, V, D) = \sum_{k=\text{Min}}^{n} \binom{n}{k} p^k (1 - p)^{n-k}$$

(1)

The traditional discussion of returns to scale and natural monopoly is in terms of the average cost function. In section 6 a complete cost of insurance is assumed and analyzed. Here, as a first step to be utilized later, it is shown that not only $Insolv$, the probability of insololvency, but also $Insolv/n$ go up with $n$. So, for a given cost of insololvency, that aspect of the cost of insurance effects a negative return to scale.

Note a basic peculiarity of insurance production. Unlike the classical cost function, in insurance the cost depends, through $Insolv$, on the price of the product $s$. This is because from (1) clearly $\partial Insolv/\partial s < 0$, a higher price implies lower cost because the chance of insololvency is diminished.
In the setting of this binomial model the law of large numbers assures us the following:
For higher than ‘fair’ price, i.e., \( s > \frac{Dp}{1+r} \), \( \text{Insolv}(s,n,p,V,D) \) tends to zero as \( n \to \infty \).

To see this, note that the critical fraction of losses associated with the firm’s insolvency is \( \frac{V/n+s}{1+r} \). As losses are binomial, their fraction, \( k/n \), has mean \( p \) and variance \( p(1-p)/n \). Since the variance tends to zero as \( n \) increases the distribution concentrates around the mean and the tails shrink. So \( \Pr \left( \frac{k}{n} > \frac{s(1+r)}{D} \right) \) tends to zero since \( s(1+r)/D > p \) and the same goes, all the more so, for the probability of the smaller tail

\[
\Pr \left( \frac{k}{n} > \frac{V/n+s}{1+r} \right).
\]

3. A Numerical Example

Let us look at Figure 1 for the numerical example where probabilities are tediously calculated and not approximated: \( (D,p,V,r) = (10, 0.1, 100, 0) \) and \( s = 1.065 \). Figure 1 shows the probability of the firm’s survival to the next period, i.e., \( 1 - \text{Insolv} \), for this example. Only points five policies apart are shown. The minimum, 0.95008, is reached at 1183 policies, an enormous number for a firm with an equity of 100. Such a scale implies a ratio of sales to net worth (surplus), \( ns/V \) in the example, of over 12 while in known markets this ratio, as monitored by regulators, is below 2. To get to the range of increasing survival to scale, sales, \( ns \), would have to increase considerably.

4. Normal Approximation

To show the inconsequence of the law of large numbers more generally let us resort to DeMoivre-Laplace theorem that establishes the tendency of the binomial distribution to the normal which practically allows fairly close approximation even for moderate levels of \( n \). See, e.g., Feller (1968, Vol. 1, pp. 186, 192-3). Denote by \( \text{Insolv} \) the normal approximation of \( \text{Insolv} \) then
\[ \text{Insolv}(n, s, p, V, D) = \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{n^2}{2}}}{\int e^{-\frac{1}{2}Dp} \, dt} \]  

(2)

Claim 1:

\text{Insolv} is increasing in } n \text{ for a given surplus } V \text{ if } s < \frac{V}{n} + \frac{Dp}{1 + r}.

Figure 1
The Firm’s Chance of Survival as Function of Number of Policies Sold

\[
\frac{\partial \text{Insolv}(n, s, p, V, D)}{\partial n} = e^{\frac{1}{2}n(1-p)} \sqrt{2\pi} \left( \frac{1-p}{p} \right) \left( \frac{1}{2\sqrt{n}} \right) 
- e^{\frac{1}{2}n(1-p)} \sqrt{2\pi} \left( \frac{1+pr}{2Dn(1-p)} \right) \left( s - \frac{Dp}{1+r} \frac{V}{n} \right)
\]

(3)

The first element in (3) is positive and the second is negative if the condition holds implying (3) is positive•.
Claim 2:

\[
\frac{\partial}{\partial n} \left( \frac{\text{Insolv}}{n} \right) = \frac{1}{n^2} \left( n \frac{\partial \text{Insolv}}{\partial n} - \text{Insolv} \right) > 0 \quad \text{if}
\]

\[
2\sqrt{p(1-p)D/(1+r)} - \frac{V/n + Dp/(1+r) - s}{\sqrt{p(1-p)D/(1+r)}} < 
\]

\[
\left( \frac{1-p}{p} \right)^{2} + \frac{4}{n} e^{-\frac{1}{2} \left( \frac{\text{Insolv}}{p} \right)^{2}} \left( \frac{V/n + s - Dp/(1+r)}{4p(1-p)D/(1+r)} \right)^{2}
\]

Proof:

To prove the claim in a way amenable to empirical verification one may use a tractable representation of the standard normal c.d.f. Many handy approximations and inequalities are reported in Patel and Read (1982). The inequalities of Gordon (1941) and Birnbaum (1942) state:

If \( \Phi(x) \) is the standard normal c.d.f. and \( \phi(x) \) is its density then for \( x > 0 \)

\[
1 - \frac{\Phi(x)}{x} \leq \Phi(x) \leq 1 - \frac{\Phi(x)}{2} \sqrt{x^2 + 4 - x}
\]

(5)

Applying (5) to the integral-area of \( \text{Insolv} \) in (2) one gets

\[
\text{Insolv}(n, s, p, V, D) = \frac{1}{\sqrt{2\pi}} \int_{\Phi^{-1}(1-np)}^{\Phi^{-1}(nnp)} e^{-t^2/2} dt = \Phi \left( \frac{n - np}{\sqrt{np(1-p)}} \right) - \Phi \left( \frac{(V + ns)(1+r) - np}{\sqrt{np(1-p)}} \right) \leq
\]

\[
\frac{1}{\sqrt{2\pi}} \sqrt{\frac{p(1-p)D/(1+r)}{V/n + s - Dp/(1+r)}} e^{-\frac{1}{2} \left( \frac{\text{Insolv}}{p} \right)^{2}} \left( \frac{V/n + s - Dp/(1+r)}{4p(1-p)D/(1+r)} \right)^{2}
\]

and to demonstrate \( n \frac{\partial \text{Insolv}}{\partial n} - \text{Insolv} > 0 \) it needs to show

\[
\frac{\partial \text{Insolv}}{\partial n} = \frac{1}{\sqrt{2\pi}} \left( \frac{1-p}{p} \sqrt{n} \right)^{2} + \frac{\sqrt{n(1+r)} \left[ V/n + Dp/(1+r) - s \right]}{2Dp(1-p)}
\]

(6)

The latter inequality is implied by (4) \( \bullet \).
Claim 2 implies that under prevailing conditions, (see below) insolvency cost contributes toward decreasing returns to scale in insurance production.

5. Empirics

The crucial issue is whether the claims’ conditions hold in the relevant range that corresponds to the real insurance market. Let us add real-world insurance concepts and numbers. Denote producer price

$$s = \left( \frac{1}{1+r} + a \right) D_p$$

where $a$ is the average profit mark-up. This $s$ is the relevant price for survival since it implies the proceeds available for indemnity. The consumer price,

$$D_p = \left( \frac{1}{1+r} + a + c \right) D_p$$

is higher and includes average costs mark-up $c$. Denote the ratio of sales to equity, sometimes named the Kenny ratio and commonly monitored by regulators, as

$$K = \frac{n \left( \frac{1}{1+r} + a + c \right) D_p}{V}$$

Using these definitions the inequality

$$s < \frac{V}{n + \frac{D_p}{1+r}}$$

in claim 1 implies

$$K < 1 + \frac{1}{a} + \frac{c}{a}$$

or

$$a < \frac{1}{1+r} + \frac{c}{K-1}$$

(7)

Take for example, the figures for property and casualty insurance in the U.S. 1993-1997 as reported in Kielholz (2000). Then $c=0.37$, $a=-0.07$ and $K=1.24$ so (7) is trivially satisfied. True, the figures of $a$ and $c$ are averages of a considerable variation, but (7) is very robust. If one assumes most favorable conditions like $a= c=0.2$ then (7) requires $K<7$. Current levels of $K$ worldwide are below 2.

The empirical validity claimed here is much stronger than in conventional studies employing regressions. The inequality (7) is claimed to hold not just on average but in each and every insurance market we know and for every firm.

Note that merging firms together, combining their equity, and thereby increasing the number of clients per firm, would not help change $K$ and thereby the direction of (7). The ratio of the firm’s revenue to net worth, $\frac{\text{revenue}}{V}$ is not likely to increase much, if at all. The explanation for that rests on the following simple argument:
If for each firm $i=1,2,...,k$ \( \frac{\text{revenue}_i}{V_i} < A \) then \( \frac{\sum_{i=1}^{k} \text{revenue}_i}{\sum_{i=1}^{k} V_i} < A \).

To enjoy the survival advantages of the law of large numbers, the firm (or the industry), has to increase $K$, which means either increase $n$ for given levels of $a$ (i.e., $s$) and $V$ or increase $a$ for given levels of $n$ and $V$, as may be deduced from (7). Both avenues are closed by consumers’ limited demand. There is still another, absurd, way to increase $K$: reduce $V$. That will indeed reverse (7) but is a counter-productive way to create returns to scale in survival because it per se causes the chance of survival to dwindle. It will also shift demand down, because of the reduced quality of the product (solvency), an issue not addressed in this work, and hence afford no room for taking advantage of the returns to scale by expanding $ns$. Therefore one may conclude that in current markets for insurance, selling more policies will increase the firms’ prospect of ruin.

As to claim 2 and its condition (4), using $a,s,K$ it has, after manipulations, the form

\[
\left( \frac{1 - p}{p} \right) \frac{1}{1 + r} + a + c \left( \frac{1}{1 + r} \right) - a - n \frac{\left( \frac{1 - p}{p} \right)}{K} + a \frac{\left( \frac{1 - p}{p} \right)}{K} < e^{\frac{n}{2} \left( \frac{1 - p}{p} \left( \frac{1 + a + c}{K} \right)^2 - a^2 \right)}
\]

which is easily satisfied for all practical and reasonable figures. For example it is sufficient that the left-hand side of (8) be negative, that is

\[
\frac{2(1 - p)}{(1 + r)^2} < np
\]

which holds for the figures of Kielholz and $r=0.1$ when $2<pn$, i.e., when the expected number of claims is larger than two. If $a=c=0.2$ one needs $4<pn$. 
6. Return to Scale in Optimal Expansion

The results in section 4 were based on the assumption of a given equity (surplus). One may claim that an expanding firm, wishing to sell more policies, may secure more equity by issuing stock or retaining profits. The firm may also resort to reinsurance, a risk-reducing device neglected heretofore. This section demonstrates that even allowing such optimal adjustments the law of large numbers does not resurrect as the driving force of insurance.

Incorporating the new elements assume the following cost function of insurance.

\[ C(n,V,x) = \frac{npD(1-x)}{1+r} + nPDe + \frac{B \cdot \text{Insolv}(n,V,x)}{1+r} + gV + npDx(1+\delta) \tag{9} \]

A fraction \( x \) of the firm’s business is co-insured by reinsurance. The first term in (9) is the actuarial cost, i.e., the expected claims on the retention. The second term is underwriting, administrative and marketing costs which accrue right away and so are not discounted. In light of Doherty (1981) and other studies \( c \) may be a (mildly) decreasing function of \( n \) but here it is assumed constant as the focus is the other types of cost, and their effect on return to scale. \( B \) is the cost of reaching a state of insolvency, and \( g \) is the cost of $1 of equity. The last term is the cost of reinsurance where the loading factor is \( 1+\delta \).

Three assumptions were made here that are favorable to increasing returns to scale. First, a constant insolvency cost \( B \) was assumed, i.e., one that does not rise with the number of failed contracts, \( n \), as maybe it should. The same goes for \( g \) and \( \delta \). As the firm acquires more equity its marginal price would normally rise, see Mayshar (1983). Also, the price of reinsurance may normally rise with the amount of risk transferred. For a given \( n \) the firm would minimize the cost by selecting the optimal \( V \) and \( x \).

Denote by \( lb \) the lower bound of the modified integral in (2) incorporating the new elements:

\[ lb = \sqrt{n} \left[ \frac{V}{n} + pD(1-x + a - x\delta) \right] \left(1 + \frac{r}{1+p} \right) - \sqrt{\frac{p}{1-p}} \tag{10} \]

**Theorem:**
An insurance firm having the linear cost function \((9)\), operating in a competitive market for standard contracts where the price, that is the mark-up \(a\), is given would

1. Choose \(x=V=0\) if \(B=0\).

2. If \(B>0\) the firm would select \(x\) and \(V\) such that

\[
\frac{V}{n(1-x)} = pD \left[ \frac{r}{1+r} + \frac{\delta}{g} \right] + \frac{(\delta - a)}{(1-x)}.
\]

3. In optimal production a larger \(n\) is associated with a larger reinsurance rate and with a smaller firm’s output (retention), i.e.,

\[
\frac{dx}{dn} > 0, \quad \frac{d[n(1-x)]}{dn} < 0.
\]

4. \(\frac{dV}{dn} = pD \left[ \frac{(1-x) \left( \frac{r}{1+r} + \frac{\delta}{2g} \right)}{(1-\lambda b^2) + (\delta - a)} \right].\)

The first element in the brackets is negative for an insurer with \(\text{Insolv}<0.16\), so equity is decreasing in \(n\), unless \(\delta - a\) is positive and sufficiently large.

5. Average cost goes up with \(n\), \(\frac{dC}{n(1-x)} > 0\) for every practically imaginable set of the parameters \(a,c,g,r,\delta\).

**Proof:**

Differentiating \((9)\) partially yields for an interior solution, using \((2)\) and \((10)\):

\[
\frac{\partial C}{\partial V} = \frac{B}{1+r} \frac{\partial \text{Insolv}(n)}{\partial V} + g = -\frac{B}{\sqrt{2\pi}} e^{-\frac{\delta^2}{2}} \frac{1}{\sqrt{n(1-x)D \sqrt{p(1-p)}}} + g = 0 \quad (11)
\]

\[
\frac{\partial C}{\partial x} = npD \left( \delta +1 - \frac{1}{1+r} \right) + \frac{B}{1+r} \frac{\partial \text{Insolv}(n)}{\partial x} = npD \left( \delta +1 - \frac{1}{1+r} \right) - \frac{B}{\sqrt{2\pi}} e^{-\frac{\delta^2}{2}} \frac{\sqrt{n[V/n + pD(a-\delta)]}}{(1-x)^2 D \sqrt{p(1-p)}} = 0 \quad (12)
\]

The second order conditions are satisfied. For \(B=0\) both \((11)\) and \((12)\) are positive so cost is minimized at \(x=V=0\). By solving equations \((11)\) and \((12)\) one gets the relationship between the optimal \(V\) and \(x\).
\[ V = npD \left( \frac{(1-x) \left( \frac{r}{1+r} + \delta \right)}{g} + (\delta - a) \right) \] (13)

In a different form (13) shows optimal equity per retained contract

\[ \frac{V}{n(1-x)} = pD \left( \frac{\left( \frac{r}{1+r} + \delta \right)}{g} + \frac{(\delta - a)}{(1-x)} \right) \] (14)

Write the average cost, making use of (13)

\[ \frac{C}{n(1-x)} = \frac{pD(1+c+\delta - g(a-\delta))}{1-x} + \frac{B \cdot \text{Insolv}(n,V,x)}{n(1-x)(1+r)} \] (15)

Now differentiate (15) totally with respect to \( n \) holding \( V \) and \( x \) optimal for every \( n \)

\[ \frac{d}{dn} \left( \frac{C}{n(1-x)} \right) = \frac{pD(1+c+\delta - g(a-\delta))}{n(1-x)^2(1+r)} \frac{dx}{dn} - \frac{B \cdot \text{Insolv}}{n^2(1-x)(1+r)} + \frac{B \cdot \text{Insolv}}{n(1-x)(1+r)} \frac{dV}{dn} \]

Note that from (13) one gets

\[ \frac{dV}{dn} = pD \left( \frac{(1-x) \left( \frac{r}{1+r} + \delta \right)}{g} + (\delta - a) - n \frac{(r}{1+r} + \delta) \right) \] (16)

Substituting (11)-(13) and (16)

\[ \frac{d}{dn} \left( \frac{C}{n(1-x)} \right) = \frac{B \left( \frac{\partial \text{Insolv}}{\partial n} - \text{Insolv} \right)}{n^2(1-x)(1+r)} - \frac{pD \left( \frac{r}{1+r} + \delta \right)}{n} + \frac{pDg(\delta - a)}{n} + \left( \frac{pD(1+c+\delta + g(\delta - a))}{(1-x)^2} + \frac{B \cdot \text{Insolv}}{n(1-x)^2(1+r)} \right) \frac{dx}{dn} \] (17)

To evaluate the sign of (17) first do it for \( \frac{dx}{dn} \). Differentiate (11) totally with respect to \( n \) to arrive, after canceling out terms, at

\[ -\frac{dx}{dn} + \frac{1-x}{2n} + (1-x)b \frac{dlb}{dn} = 0 \] (18)
Differentiate $lb$ totally with respect to $n$

$$\frac{dlb}{dn} = \frac{d}{dn} \left( \sqrt{n} \left[ \frac{V}{n} + pD(1-x+a-x\delta) \frac{1+r}{(1-x)D\sqrt{p(1-p)}} - \sqrt{p \frac{1}{1-p}} \right] \right)$$

$$lb = \frac{V (1+r)}{2n} - \frac{V(1+r)}{n\sqrt{n(1-x)D\sqrt{p(1-p)}}} \cdot \frac{dx}{dn}$$

$$\frac{dx}{dn} = \frac{2n}{2n} = lb = T \sqrt{n} \quad \text{for any constant } T > 0$$

The reason for the last equality is that the other elements vanish because of (13). The coefficients of $\frac{dx}{dn}$ in (19) sum up to a multiple of the difference between the sides of (13), which equals zero, and the same goes for the sum of the other, free, elements. So when the firm selects its variables optimally

$$\frac{dlb}{dn} = \frac{lb}{2n} \Rightarrow lb = T \sqrt{n} \quad \text{for any constant } T > 0$$

Note that in a typical firm $lb > 2.33$ which is implied by $\text{Insolv} < 0.01$. If it has 100 customers then $T < 0.233$, if it has 1000 customers then $T < 0.07$.

From (17) and (20)

$$\frac{dx}{dn} = \frac{1-x}{2} \left( \frac{1}{n} + T^2 \right)$$

which is positive. Substituting (21) in (16) yields

$$\frac{d}{dn} \frac{C}{n(1-x)} = \frac{B \left( n \frac{\partial \text{Insolv}}{\partial n} - \text{Insolv} \right)}{n^2 (1-x)(1+r)} - \frac{pD \left( r \frac{1}{1+r} + \delta \right)}{n} - \frac{pDg(\delta - a)}{n} + \left( \frac{pD(1+c+\delta + g(\delta - a))}{(1-x)^2} + \frac{B \cdot \text{Insolv}}{n(1-x)^2(1+r)} \right) \frac{1-x}{2} \left( \frac{1}{n} + T^2 \right)$$

The first term of (22) is positive due to claim 2. By employing only the second, third and part of the fourth elements in (22) one gets a sufficient condition for (22) to be positive.
Indeed, (23) holds for every imaginable set of the parameters $a, c, g, r, \delta$ because they are fractions representing mark-ups, rates of return and price of equity.

As to the effect on $V$, from (18)

$$\frac{dV}{dn} = pD \left[ \frac{(1-x)(r}{1+r} + \delta) \right] - (a - \delta) - npD \frac{(r}{1+r} + \delta) dx =$$

$$pD \left[ \frac{(1-x)(r}{1+r} + \delta) \right] + (\delta - a) - npD \frac{(r}{1+r} + \delta) \left(1-x\right) \left(1+T^2\right) \left(\frac{1}{n} + T^2\right) \right) (24)$$

$$1 < \sqrt{nT} = lb \text{ if } Insolv<0.16, \text{ so } (24) \text{ is negative provided } \delta - a \text{ is not too large}$$

The economic implication of the theorem is that when a firm expands output it would optimally do so by reducing $n$ and expanding its retention $n(1-x)$. In doing so average cost goes down, manifesting returns to scale via smaller numbers. This analysis assumes that reinsurance firms are ready to carry out the implied countervailing action of contracting their output. The strain of that assumption is starker when the firm reduces output. It does so by taking more clients but reducing its retention. The reinsurance firms must be ready to take all the extra risk, and at the same price. Therefore it is illuminating to ask: what happens to the average cost of producing insurance for the entire industry when it expands?

Let us suppose that the industry is made of one firm that still acts as a competitive one under, say, price regulation. As a single firm it has no
resort to reinsurance so its cost is (9) but with \( x \equiv 0 \). Cost minimization with respect to \( V \) still holds, i.e. (11), with \( x \equiv 0 \). (15) has the form

\[
\frac{C}{n} = pD \left( \frac{1}{1 + r} + c \right) + \frac{B \cdot \text{Insolv}(n,V)}{n(1 + r)} + \frac{gV}{n}
\]

(25)

\[
\frac{d}{dn} \left( \frac{C}{n} \right) = \frac{B \cdot \text{Insolv}}{n^2(1 + r)} + \frac{B}{n(1 + r)} \frac{d\text{Insolv}}{dn} - \frac{gV}{n^2} + \frac{g}{n} \frac{dV}{dn}
\]

\[
= -\frac{B \cdot \text{Insolv}}{n^2(1 + r)} + \left( \frac{B}{n(1 + r)} \frac{\partial \text{Insolv}}{\partial V} + \frac{B}{n(1 + r)} \frac{\partial \text{Insolv}}{\partial V} \frac{dV}{dn} \right) - \frac{gV}{n^2} + \frac{g}{n} \frac{dV}{dn}
\]

(26)

The first element in (26) is positive by claim 2. The second element is the effect of optimally adjusting \( V \) and it is zero by (11). The last element is negative but is the very same one as in the case of a fixed \( V \). The aforementioned empirical studies of returns to scale concentrated on \( c \) and did not consider the (fixed) cost of equity. The sign of (26) depends on the balance between the first and last elements but certainly not on the law of large numbers.

7. Conclusion

By way of a simple model of risk and insurance, this paper examines a fundamental question: what is the engine behind the insurance industry and its relative advantage in bearing risks? Concentrating on the cost and survival of firms, and without invoking utilities, it finds that the universal answer of the literature, the law of large numbers, could not practically be true. For given equity and rate of reinsurance each additional customer aggravates the peril of financial ruin of the firm as the numbers involved are too small. Merging all firms together combing their equity will not make a difference. One must conclude that there is not enough demand for insurance in the world to reap the benefits of the law of large numbers.

Allowing the firm to select optimal equity and reinsurance at constant marginal prices, one finds increasing returns to scale associated with large numbers. A firm wishing to increase output would optimally do
it by reducing the number of contracts it sells, but increase its retention and equity. In doing so the risk of insolvency is raised; nevertheless average cost is reduced. Consequently one should not expect natural monopoly conditions in insurance, as implied by the standard prevailing legend, if the cost advantage in underwriting found in the empirical studies is more than offset by the other cost drawbacks discussed above. At any rate, large numbers have no bearing on the impetus for insurance.

The basis for insurance must be pursued and discovered in another territory. Only in a mutual should the number of owner-partners vary jointly with that of the insureds. The relative advantage of the stock insurance firm in bearing risk lies in the multitude of owners, not of clients, through the spreading of the risks among the many shareholders. Spreading a given risk over many firm shoulders tends to evaporate its burden. Samuelson (1963) wrote this conjecture but did not follow it. Arrow and Lind (1970), who analyze a large risky public project, demonstrate this for a single risk. Mayshar (1981) pointed out the relevance of their result for insurance and eloquently described it as the ‘magic’ of insurance. Shilony (2005) shows that insurance production manifests returns to scale in its fundamental and unique inputs: equity and shareholders. It also shows that reinsurance is a perfect substitute for those two inputs.
References


Shilony, Y., (2005), The Impetus of Insurance Production, Returns to Scale and Reinsurance, unpublished.