FROM RISK SHARING TO PURE PREMIUM FOR A LARGE NUMBER OF HETEROGENEOUS LOSSES

Denuit, M. and C. Y. Robert

REPRINT | 2020 / 15

UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

FROM RISK SHARING TO PURE PREMIUM FOR A LARGE NUMBER OF HETEROGENEOUS LOSSES

Michel Denuit
Institute of Statistics, Biostatistics and Actuarial Science - ISBA
Louvain Institute of Data Analysis and Modeling - LIDAM
UCLouvain
Louvain-la-Neuve, Belgium
michel.denuit@uclouvain.be

Christian Y. Robert
Laboratory in Finance and Insurance - LFA
CREST - Center for Research in Economics and Statistics
ENSAE
Paris, France
chrobert@ensae.fr

March 13, 2020

Abstract

This paper considers linear fair risk sharing rules and the conditional mean risk sharing rule for independent but heterogeneous losses that are gathered in an insurance pool. It studies the asymptotic behavior of individual contributions to total losses when the number of participants to the pool tends to infinity. It is shown that (i) insurance at pure premium is obtained for an infinitely large pool and (ii) the difference between the actual contribution and the pure premium becomes ultimately Normally distributed. The linear fair risk sharing rule approximating the conditional mean risk sharing rule is then identified, providing practitioners with a useful simplification applicable within large pools. Also, the approximate number of participants required to keep the volatility of individual contributions within an acceptable range is obtained from the established asymptotic Normality.

Keywords: risk pooling, peer-to-peer (P2P) insurance, law of large number, central-limit theorem, size-biased transform.

JEL Classification: G22

1 Introduction and motivation

Pooling of risks has received considerable attention in actuarial science. Recently, this topic has been re-visited in the context of peer-to-peer (or P2P) insurance systems. However, many papers are restricted to the case of independent and identically distributed risks. Whereas independence can be considered as reasonable, at least as an approximation, for many insurance risks, homogeneity is highly questionable. In this paper, we consider an insurance pool where the participants are exposed to independent but heterogeneous losses and study the asymptotic behavior of risk sharing rules as the number of participants increases.

Linear risk sharing rules are often applied in P2P insurance systems. In life insurance, several proposals for mutual inheritance schemes have be made in the literature and highlight such risk sharing rules. We refer the reader e.g. to Donnelly and Young (2017). As shown by Donnelly (2015), the notion of fairness (i.e. the expected gain is zero for all participants) is very relevant for this kind of pooling scheme. Schumacher (2018) discussed financial fairness in the context of linear risk exchanges. In this paper, we compare the asymptotic behavior of linear fair risk sharing rules to that of the conditional mean risk sharing rule proposed by Denuit and Dhaene (2012), that has been sucessfully applied to P2P insurance by Denuit (2019).

Nonlinear risk sharing rules have not received much attention so far. The reason may be that they are not perceived as relevant for practice. The conditional mean risk sharing rule defined by Denuit and Dhaene (2012) is however transparent and relatively easy to communicate to participants (being based on the familiar concept of averaging), despite being generally nonlinear. According to this rule, each participant contributes the conditional expectation of the loss brought to the pool, given the total loss experienced by the entire pool. The properties of the conditional mean risk allocation have been studied in Denuit (2019) and Denuit and Robert (2020) when the number of participants is fixed. The conditional mean risk sharing rule satisfies the risk exchange fairness condition and enjoys many attractive theoretical properties so that it can be considered as a reference risk sharing rule in P2P insurance applications.

Under the assumption of independent (but not too heterogeneous) risks, it is expected that diversification takes place within the pool as the number of participants increases, and that the financial contribution of each participant to the pool converges to the expectation of his or her respective loss. In this paper, we study the asymptotic behaviors of the linear and conditional mean risk sharing rules and extract a particular linear risk sharing rule that is asymptotically equivalent to the conditional mean risk sharing rule. This is particularly relevant for practice.

The remainder of this paper is organized as follows. In Section 2, we present the class of fair linear risk sharing rules as well as the conditional mean risk sharing rule. In Section 3, it is established that under mild technical conditions, the individual contributions converge to the mathematical expectation of the loss (or pure premium) with probability one. In Section 4, we establish central-limit theorems for individual contributions. This is especially useful to get an idea of the volatility of P2P contributions around the pure premium. The rate and radius of convergence to the pure premium is studied in Section 5. The final Section 6 briefly discusses the results. For convenience, the proofs are gathered in appendix.

The following notation is adopted throughout the text. For two positive functions g_1 and

g_2 defined in a neighborhood of infinity, we write $g_1 = o(g_2)$ provided $\lim_{x\to\infty} g_1(x)/g_2(x) = 0$, and $g_1 = O(g_2)$ provided $\lim_{x\to\infty} |g_1(x)/g_2(x)| < \infty$.

2 Risk sharing rules

2.1 Fair risk sharing rules

Consider n participants to an insurance pool, numbered $i = 1, 2, \ldots, n$. Each of them faces a risk X_i. By risk, we mean a non-negative random variable representing a monetary loss. Throughout the paper, we assume that X_1, X_2, X_3, ... are mutually independent. Let us denote

$$\mu_i = \mathrm{E}[X_i] > 0 \text{ and } \sigma_i^2 = \mathrm{Var}[X_i] > 0$$

the mean and the variance of X_i, respectively. Both μ_i and σ_i^2 are assumed to be finite throughout the paper. We voluntarily exclude the cases where no randomness is present, that is, $\mu_i = 0 \Leftrightarrow X_i = 0$ with probability 1, and $\sigma_i^2 = 0 \Leftrightarrow X_i = \mu_i$ with probability 1.

Often in the literature devoted to insurance, the random variables X_i are assumed to be identically distributed. In this paper, we depart from the homogeneous situation and explicitly allow for different distributions.

Example 2.1 (Explanatory variables). *A typical case is when predictive explanatory variables $\boldsymbol{Z}_i$ for (a priori identically distributed) losses Y_i are available. If the random vectors $(Y_1, \boldsymbol{Z}_1)$, $(Y_2, \boldsymbol{Z}_2)$, ... are independent and identically distributed then X_i is distributed as Y_i given $\boldsymbol{Z}_i = \boldsymbol{z}_i$. We therefore have $\mu_i = \mathrm{E}[Y_i|\boldsymbol{Z}_i = \boldsymbol{z}_i]$ and $\sigma_i^2 = \mathrm{Var}[Y_i|\boldsymbol{Z}_i = \boldsymbol{z}_i]$.*

Borch (1962) established that under mild assumptions, participants' optimal risk sharing depends only on aggregate loss $S_n = \sum_{i=1}^n X_i$. Therefore we only focus on risk sharing rules associated to comonotonic risk allocation schemes. We denote by $h_{i,n}(s)$ the amount participant i contributes to the pool, where $s = \sum_{i=1}^n x_i$ is the sum of the realizations $x_1, x_2, \ldots, x_n$ of $X_1, X_2, \ldots, X_n$.

Definition 2.2. *A fair risk sharing rule is an allocation scheme such that, for all $n = 1, 2, \ldots$, there exist (measurable) functions $h_{1,n}$, ..., $h_{n,n}$ satisfying*

$$\sum_{i=1}^{n} h_{i,n}(s) = s \text{ for all } s \geq 0 \text{ and } \mathrm{E}[h_{i,n}(S_n)] = \mathrm{E}[X_i] \text{ for } i = 1, \ldots, n.$$

2.2 Linear fair sharing rules

When the participants enter the pool, they are informed about the amount $h_{i,n}(S_n)$ they will have to contribute as a function of the total realized loss S_n. In the design of a recognized scheme, it is important that the sharing rule represented by the functions $h_{i,n}$ is both intuitively acceptable and transparent. In that respect, linear risk sharing schemes of the form

$$h_{i,n}^{\mathrm{lin}}(S_n) = \mathrm{E}[X_i] + a_{i,n}\left(S_n - \mathrm{E}[S_n]\right), \quad i = 1, 2, \ldots, n,$$

where $\sum_{i=1}^{n} a_{i,n} = 1$ are especially appealing. Clearly, such linear risk sharing scheme allocate the full risk S_n and satisfy the fairness constraint $\mathrm{E}[h_{i,n}(S_n)] = \mathrm{E}[X_i]$ for $i = 1, \ldots, n$ so that linear rules $h_{i,n}^{\mathrm{lin}}$ qualify as fair risk sharing rules according to Definition 2.2.

The rule described by the functions $\{h_{i,n}^{\mathrm{lin}},\ i = 1, 2, \ldots, n\}$ can be understood as an agreement between participants to pay the pure premium $\mathrm{E}[X_i]$ and to divide deviations of S_n from the total pure premium $\mathrm{E}[S_n]$ (positive or negative) in proportion to the coefficients $a_{i,n}$. The numbers $a_{i,n}$ are called participation coefficients by Schumacher (2018). The design of the allocation scheme then amounts to select an appropriate set of participation coefficients $a_{i,n}$. Several choices of participation coefficients are possible, as shown in the next examples that will be used throughout this paper.

Example 2.3 (Proportional rule). *Participants may agree to take a fixed percentage of the total loss S_n, in accordance with the expected values of the risks they bring to the pool compared to the total expected loss, that is,*

$$a_{i,n}^{\mathrm{prop}} = \frac{\mathrm{E}[X_i]}{\mathrm{E}[S_n]}.$$

This is a financially fair rule, which is referred to as the proportional rule by Schumacher (2018). The amount to be paid by participant i is

$$h_{i,n}^{\mathrm{prop}}(S_n) = \mathrm{E}[X_i] + a_{i,n}^{\mathrm{prop}} \left(S_n - \mathrm{E}[S_n]\right) = \frac{\mathrm{E}[X_i]}{\mathrm{E}[S_n]} S_n.$$

This rule has been applied by Donnelly and Young (2017), for instance. With $h_{i,n}^{\mathrm{prop}}$, volatility is not accounted for because participants i_1 and i_2 with $\mu_{i_1} = \mu_{i_2}$ contribute equally to the total loss even if the respective variances $\sigma_{i_1}^2$ and $\sigma_{i_2}^2$ strongly differ. Also, the proportional rule is unique since it is completely determined by the constraint of financial fairness.

Example 2.4 (Linear regression rule). *Participants may also agree about a somewhat more elaborate scheme. Since $\mathrm{Var}[S_n] < \infty$, we can also propose to share the total risk according to the relative volatility of the risks brought to the pool, that is, to adopt participation coefficients of the form*

$$a_{i,n}^{\mathrm{reg}} = \frac{\sigma_i^2}{\sum_{j=1}^{n} \sigma_j^2}.$$

Now, participants allocate the deviation from the pure premium according to the relative volatility of the risk they bring to the pool, compared to the volatility of the total loss. The corresponding risk sharing rule is the one that minimizes among all linear rules $h_{i,n}^{\mathrm{lin}}$ the expected squared difference between the risk X_i brought to the pool and the individual contribution $h_{i,n}(S_n)$, that is,

$$\mathrm{E}\Big[\big(X_i - h_{i,n}^{\mathrm{reg}}(S_n)\big)^2\Big] = \min_{a \in \mathbb{R}} \mathrm{E}\Big[\big(X_i - \mathrm{E}[X_i] - a\left(S_n - \mathrm{E}[S_n]\right)\big)^2\Big].$$

The solution a to this minimization problem is known to be

$$a = \frac{\mathrm{Cov}[X_i, S_n]}{\mathrm{Var}[S_n]} = \frac{\sigma_i^2}{\sum_{j=1}^{n} \sigma_j^2} = a_{i,n}^{\mathrm{reg}}.$$

The corresponding risk sharing rule

$$h_{i,n}^{\text{reg}}(S_n) = \mathrm{E}[X_i] + \frac{\sigma_i^2}{\sum_{j=1}^n \sigma_j^2}\left(S_n - \mathrm{E}[S_n]\right), \ \ i = 1, 2, \ldots, n,$$

is henceforth referred to as the linear regression rule. Here, deviations of the total loss S_n from the pure premium $\mathrm{E}[S_n]$ are allocated among participants according to the volatility of the risks brought to the pool.

2.3 Conditional mean risk sharing rule

Of course, there is no reason to restrict the study to linear rules, only. Participants may also agree about a nonlinear risk sharing mechanism in which the amounts allocated to the participants are determined as general (i.e., not necessarily linear) functions of the realized total loss. Denuit and Dhaene (2012) introduced a nonlinear risk sharing scheme that is particularly attractive in the context of P2P insurance: the conditional mean risk sharing defined as

$$h_{i,n}^{\star}(S_n) = \mathrm{E}[X_i|S_n], \ \ i = 1, 2, \ldots, n. \tag{2.1}$$

Since $\mathrm{Var}[S_n] < \infty$, we have

$$\mathrm{E}\big[\big(X_i - h_{i,n}^{\star}(S_n)\big)^2\big] = \min_{h(\cdot):\mathrm{Var}[h(S_n)]<\infty} \mathrm{E}\big[\big(X_i - h\left(S_n\right)\big)^2\big].$$

In words, the contribution $h_{i,n}^{\star}(S_n)$ paid by participant i is the closest to the loss X_i brought to the pool, in the sense that it minimizes the expected squared difference of the risk X_i and any measurable function $h(S_n)$ of the total loss S_n. The difference between $h_{i,n}^{\text{reg}}$ and $h_{i,n}^{\star}$ thus corresponds to the class of risk sharing rules under consideration: with $h_{i,n}^{\text{reg}}$, participants restrict the risk sharing rule to be linear whereas with $h_{i,n}^{\star}$ they also allow for nonlinear risk sharing rules. In both cases, the goal is to minimize the expected squared difference between the risk brought to the pool by each participant and his or her contribution to the realized total loss.

With $h_{i,n}^{\star}$, participant i must contribute the expected value of the risk X_i brought to the pool, given the total loss S_n. In the expected utility setting, every risk-averse decision-maker prefers $h_{i,n}^{\star}(S_n)$ over the initial risk X_i so that the conditional mean risk sharing rule appears to be beneficial to all participants (as an application of Jensen's inequality). Denuit and Dhaene (2012) established that the conditional mean risk sharing rule is Pareto-optimal for all risk-averse economic agents behaving according to the expected utility paradigm, as long as every function $h_{i,n}^{\star}$ is non-decreasing. The properties of this scheme have been studied in Denuit (2019) and Denuit and Robert (2020) when the number of participants is fixed. Notice that the conditional mean risk sharing rule is not based on individual preferences beyond risk aversion. This is important for the applications to P2P insurance where individual preferences cannot easily be elicited.

2.4 Relationships between sharing rules

In general, the conditional mean risk sharing $h_{i,n}^{\star}$ is not linear. Furman et al. (2018) studied the case where $h_{i,n}^{\star}(S_n) = \beta_i S_n$ for some β_i depending on the means of the risks under consideration (see Theorem 3.2 in that paper). Thus, $h^{\star}_{i,n}$ and $h^{\text{prop}}_{i,n}$ coincide in these cases. Denuit and Robert (2020) have also studied the asymptotic linearity of the conditional mean risk sharing rule, for sufficiently large realizations of the total loss. It is shown there that the rules $h^{\star}_{i,n}$ can be markedly nonlinear, depending on the respective characteristics of the risks brought to the pool.

When X_i are identically distributed, $a^{\text{prop}}_{i,n} = a^{\text{reg}}_{i,n} = 1/n$ and the three risk sharing rules considered so far coincide:

$$X_1, \ldots, X_n \text{ identically distributed} \Rightarrow h^{\text{prop}}_{i,n}(S_n) = h^{\text{ref}}_{i,n}(S_n) = h^{\star}_{i,n}(S_n) = \frac{1}{n}S_n.$$

This particular case has been extensively studied in the literature. The homogeneity assumption is however very restrictive for applications because pooling must also apply to heterogeneous risks, especially in the context of P2P insurance.

3 Almost sure behavior of participants' contributions

The results derived in this section show that insurance at pure premium can be obtained from the previous risk sharing rules when the number of participants tends to infinity. Without loss of generality, we provide asymptotic results for participant $i = 1$.

3.1 Linear fair sharing rules

We begin with the fair linear risk sharing rules. The following result shows that the contribution for each individual tends to the corresponding pure premium when the number of participants increases, provided some mild technical conditions are fulfilled.

Proposition 3.1. *If*

$$\sum_{i=1}^{\infty} \sigma_i^2 a_{1,i}^2 < \infty, \tag{3.1}$$

then

$$\lim_{n\to\infty} h^{\text{lin}}_{1,n}(S_n) = \mathrm{E}[X_1] \textit{ with probability } 1.$$

The proof of Proposition 3.1 is given in appendix. Condition (3.1) ensuring the convergence to the pure premium is essentially the one underlying Kolmogorov's strong law of large numbers. It is generally fulfilled, as shown in the next examples.

Example 3.2 (Proportional rule)**.** *Assume that there exist constants d_1 and d_2 such that the inequalities*

$$\mu_i \geq d_1 > 0 \textit{ and } \sigma_i^2 \leq d_2 < \infty \textit{ hold for all } i. \tag{3.2}$$

Condition (3.2) *is generally fulfilled in insurance applications, preventing pure premiums μ_i to become too small and variances σ_i^2 to become too large. Then,*

$$\sum_{i=1}^{\infty} \sigma_i^2 a_{1,i}^2 = \mu_1^2 \sum_{i=1}^{\infty} \frac{\sigma_i^2}{\left(\sum_{j=1}^{i} \mu_j\right)^2} \leq \mu_1^2 \frac{d_2}{d_1^2} \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty.$$

Therefore, Proposition 3.1 applies to $h^{\text{prop}}_{i,n}$ provided (3.2) *holds true.*

Example 3.3 (Regression rule). *Assume that there exists a constant d_3 such that*

$$d_3 < \sigma_i^2 \leq d_2 < \infty \textit{ hold for all } i. \tag{3.3}$$

Compared to (3.2), (3.3) *prevents variances σ_i^2 to become either too large or too small. Then,*

$$\sum_{i=1}^{\infty} \sigma_i^2 a_{1,i}^2 = \sigma_1^2 \sum_{i=1}^{\infty} \frac{\sigma_i^2}{\left(\sum_{j=1}^{i} \sigma_j^2\right)^2} \leq \sigma_1^2 \frac{d_2}{d_3^2} \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty.$$

Therefore, Proposition 3.1 applies to $h_{i,n}^{\text{reg}}$ provided (3.3) *holds true.*

Example 3.4 (Explanatory variables). *Assume that there exists $\varepsilon > 0$ such that*

$$\mathrm{E}[(\mathrm{Var}[Y_i|\boldsymbol{Z}_i])^{1+\varepsilon}] < \infty \textit{ and } \limsup_{n\to\infty} n a_{1,n} < \infty, \tag{3.4}$$

then

$$\sum_{i=1}^{\infty} \sigma_i^2 a_{1,i}^2 < \infty \textit{ with probability } 1.$$

The reasoning is the following one: the series $\sum_{i=1}^{n} \sigma_i^2 / i^2$ is almost surely convergent if for some $\alpha > 1$

$$\mathrm{P}\left[\frac{\sigma_i^2}{i^2} > \frac{1}{i^\alpha} \textit{ i.o.}\right] = 0.$$

But

$$\begin{aligned} \mathrm{P}\left[\frac{\sigma_i^2}{i^2} > \frac{1}{i^\alpha} \textit{ i.o.}\right] &= \lim_{n\to\infty} \mathrm{P}\left[\cup_{n=i}^{\infty} \left\{\mathrm{Var}[Y_i|\boldsymbol{Z}_i] > i^{2-\alpha}\right\}\right] \leq \lim_{n\to\infty} \sum_{n=i}^{\infty} \mathrm{P}\left[\mathrm{Var}[Y_i|\boldsymbol{Z}_i] > i^{2-\alpha}\right] \\ &\leq \lim_{n\to\infty} \sum_{n=i}^{\infty} \frac{\mathrm{E}[(\mathrm{Var}[Y_i|\boldsymbol{Z}_i])^{1+\varepsilon}]}{i^{(2-\alpha)(1+\varepsilon)}} = 0 \end{aligned}$$

if $\alpha < (1 + 2\varepsilon) / (1 + \varepsilon)$.

Proposition 3.1 shows that the respective participants' contributions tend to stabilize when n increases and that the limiting value is the pure premium for linear fair risk sharing rules.

3.2 Conditional mean risk sharing rule

Let us now consider the conditional mean risk sharing rule. Intuitively speaking, the impact of X_1 on S_n should vanish as n tends to ∞ as long as X_1 does not dominate the remaining $X_2, X_3, \ldots$. Therefore, X_1 and S_n become approximately independent for large pools and it seems reasonable to expect that $\mathrm{E}[X_1|S_n]$ tends to $\mathrm{E}[X_1]$ as it would be the case if X_1 and S_n were independent. This section formally establishes this result. Note however that we are not able to do it with conditions as general as for the linear risk sharing rules. In particular we will either assume that the random variables $X_1, X_2, \ldots$ are absolutely continuous or have Poisson compound distributions with absolutely continuous severities.

With such assumptions, it is actually possible to write the conditional mean $\mathrm{E}[X_1|S_n]$ as $\mathrm{E}[X_1]$ multiplied by the ratio of two density functions (see Denuit (2019)) and to study the asymptotic behaviors of these functions for independent and heteregeneous risks.

Henceforth, the variance of S_n is denoted by

$$s_n^2 = \mathrm{Var}[S_n] = \sum_{i=1}^{n} \sigma_i^2.$$

The result derived for the conditional mean risk sharing requires the following technical conditions:

Condition A: The random variables $X_1, X_2, \ldots$ are absolutely continuous with respective probability density functions $f_{X_1}, f_{X_2}, \ldots$ and have finite moments up to order 3 (and order 4 for X_1).

Condition A': The random variables $X_1, X_2, \ldots$ have Poisson compound distributions represented as

$$X_i = \sum_{k=1}^{N_i} C_{ik} \text{ with } N_i \sim \text{Poisson}(\lambda_i),\ \ i = 1, 2, \ldots, \tag{3.5}$$

where the claim severities C_{ik} are positive, absolutely continuous, distributed as C_i with probability density functions f_{C_i} and have finite moments up to order 3 (and order 4 for C_1), all these random variables being independent. There exist a positive constant η, such that, for all i, $\lambda_i > \eta$.

Condition B: We have

$$\limsup_{n\to\infty} \frac{s_{n+1}}{s_n} < \infty \text{ and } \sum_{n=1}^{\infty} \frac{\mathrm{E}\big[|X_n - \mu_n|^3\big]}{\big(s_n^2 2 \ln_2 s_n^2\big)^{3/2}} < \infty$$

where $\ln_2 t$ is defined for $t \geq 0$ as $\ln(\ln t)$ if $\ln t \geq \mathrm{e}$ and 1 otherwise.

Condition C: There exist positive constants g and G such that, for all n, we have

$$s_n^2 \geq ng \text{ and } \sum_{i=1}^{n} \mathrm{E}\big[|X_i - \mu_i|^3\big] \leq nG.$$

Moreover there exists a constant $\varepsilon \in (0, g/24G)$ such that the characteristic functions $t \mapsto \mathrm{E}[\mathrm{e}^{\mathrm{i}tX_j}]$ of $X_1, X_2, \ldots$ satisfy

$$\int_{|t|>\varepsilon} \prod_{j=1}^{n} \big|\mathrm{E}[\mathrm{e}^{\mathrm{i}tX_j}]\big| \,\mathrm{d}t = O\left(\frac{1}{n}\right).$$

Let us now briefly comment on these conditions. Condition A requires the finiteness of the first third moments (and also the fourth moment for X_1). Except in the Pareto case, this condition is generally fulfilled and appears to be very reasonable in the context of P2P insurance (which is often restricted to the lower risk layer). The assumptions in Condition

B come from Wittman (1985) who showed that they are sufficient for the law of iterated logarithm to hold for independent random variables. The assumptions in Condition C have been introduced in Petrov (1956, Theorem 2) to provide uniform approximations of the probability density functions of sums of independent random variables.

Example 3.5 (Explanatory variables). *Assume that* $\mathrm{E}[Y_i^3] < \infty$, *it follows by the law of large numbers that*

$$\lim_{n\to\infty} \frac{s_n^2}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 = \mathrm{E}[\mathrm{Var}[Y_i|\boldsymbol{Z}_i]] \textit{ with probability } 1,$$

and that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\big[|X_i - \mu_i|^3\big] = \mathrm{E}\big[|Y_i - \mathrm{E}[Y_i|\boldsymbol{Z}_i]|^3\big] \textit{ with probability } 1.$$

It is therefore easy to fix the values of the two positive constants g and G of Condition C.

We are now ready to state the main result of this section.

Proposition 3.6. *If Conditions A (or A'), B and C stated above are all valid, then*

$$\lim_{n\to\infty} h_{1,n}^{\star}(S_n) = \mathrm{E}[X_1] \textit{ with probability } 1.$$

The proof of Proposition 3.6 is given in appendix. When insurance losses are independent and identically distributed, Proposition 3.6 corresponds to Proposition 1.1 in Zabell (1980).

As established for linear fair risk sharing rules in Proposition 3.1, Proposition 3.6 shows that the respective participants' contributions tend to stabilize when the size n of the pool grows (i.e. by recruiting infinitely many participants) when the conditional mean risk sharing rule is adopted. Moreover, the limiting value is the pure premium.

These results appear to be very instructive as they link risk sharing to risk transfer through insurance contracts: if paying ex-post a random contribution, $h_{i,n}^{\mathrm{lin}}(S_n)$ or $h_{i,n}^{\star}(S_n)$, for some finite n is not considered as attractive, then commercial insurance can be considered as an alternative provided the participant is ready to pay more than $\mathrm{E}[X_i]$ to cover the random fluctuations to be absorbed by equity capital.

4 Central-limit theorems for participants' contributions

Let us investigate how the individual contribution fluctuates around the pure premium when the size n of the group is large enough.

4.1 Linear risk sharing rule

For a linear risk sharing rule, we have

$$h_{1,n}^{\mathrm{lin}}(S_n) - \mathrm{E}[X_1] = a_{1,n}\big(S_n - \mathrm{E}[S_n]\big).$$

Therefore, if $S_n - \mathrm{E}[S_n]$ obeys a central-limit theorem then $h_{1,n}^{\mathrm{lin}}(S_n) - \mathrm{E}[X_1]$ is asymptotically Normally distributed. This is formally stated in the next result, where the Lyapunov central-limit theorem is used (Theorem 4.9 in Petrov, 1995).

Proposition 4.1. *Assume that Lyapunov's condition holds, i.e., for some* $\delta > 0$,

$$\lim_{n\to\infty} \left(s_n^2\right)^{-1-\delta/2} \sum_{i=1}^{n} \mathrm{E}\left[|X_i - \mu_i|^{2+\delta}\right] = 0.$$

Then,

$$\frac{1}{a_{1,n} s_n} \left(h_{1,n}^{\mathrm{lin}}(S_n) - \mathrm{E}[X_1]\right) \xrightarrow{\mathcal{L}} \mathrm{Normal}\,(0,1)\,,$$

where $\xrightarrow{\mathcal{L}}$ *denotes the convergence in distribution.*

The same result holds under alternative assumptions ensuring that $(S_n - \mathrm{E}[S_n])/s_n$ converges to the standard Normal distribution. We refer the reader to Chapter 4 in Petrov (1995) for different statements.

Example 4.2 (Regression rule)**.** *When* $a_{1,n} = \sigma_1^2/s_n^2$, *we get*

$$\frac{s_n}{\sigma_1^2} \left(h_{1,n}^{\mathrm{reg}}(S_n) - \mathrm{E}[X_1]\right) \xrightarrow{\mathcal{L}} \mathrm{Normal}\,(0,1)\,.$$

Example 4.3 (Explanatory variables)**.** *Lyapunov's condition holds if* $\mathrm{E}[|Y_i - \mathrm{E}[Y_i|\boldsymbol{Z}_i]|^{2+\delta}] < \infty$ *for some* $\delta > 0$.

4.2 Conditional mean risk sharing rule

The following proposition shows that the regression sharing rule and the conditional mean sharing rule are asymptotically equivalent, i.e. the fluctuations of the individual contributions around the pure premium are roughly identical for $h_{1,n}^{\star}$ and $h_{1,n}^{\mathrm{reg}}$ when n becomes large enough. Given that both $h_{1,n}^{\star}(S_n)$ and $h_{1,n}^{\mathrm{reg}}(S_n)$ converge to $\mathrm{E}[X_1]$ with probability 1, we thus see that $h_{1,n}^{\mathrm{reg}}$ provides a reasonable approximation to $h_{1,n}^{\star}$ when the size of the pool becomes large.

Proposition 4.4. *If Conditions A (or A'), B and C stated above are all valid, then*

$$\frac{s_n}{\sigma_1^2} \begin{pmatrix} h_{1,n}^{\mathrm{reg}}(S_n) - \mathrm{E}[X_1] \\ h_{1,n}^{\star}(S_n) - \mathrm{E}[X_1] \end{pmatrix} \xrightarrow{\mathcal{L}} \mathrm{Normal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right).$$

The proof of Proposition 4.4 is given in appendix. The stated result appears to be useful to control the relative variations of individual contributions around the pure premium. Les z_ϵ be the ϵ-quantile of the standard Normal distribution, leaving a probability ϵ to its right. Proposition 4.4 then shows that

$$\mathrm{P}\left[\left|h_{1,n}^{\star}(S_n) - \mathrm{E}[X_1]\right| \le z_{\alpha/2}\frac{\sigma_1^2}{s_n}\right] \approx 1 - \alpha.$$

If we wish to limit the relative variations to $\beta\%$ of the pure premium, at confidence level $1-\alpha$, then we need a number n of participants such that

$$s_n \ge \frac{z_{\alpha/2}\sigma_1^2}{\beta \mathrm{E}[X_1]} \Leftrightarrow \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_1^2} \ge \left(\frac{z_{\alpha/2}}{\beta}\mathrm{CV}[X_1]\right)^2,$$

where $\mathrm{CV}[X_1]$ is the coefficient of variation of X_1. We thus see that the number n of participants does not matter in itself but the proper unit is σ_i^2/σ_1^2, that is, the relative variability of the risks brought to the pool by participants $i = 2, 3, 4, \ldots$, with respect to participant 1. Of course, there is no reason to privilege participant 1. The same constraint must thus hold for every participant.

5 Radius and rates of convergence

We know from Section 3 that $h_{1,n}^{\mathrm{lin}}(S_n)$ and $h_{1,n}^{\star}(S_n)$ both converge to $\mathrm{E}[X_1]$ as $n \to \infty$, with probability 1, provided some technical conditions are fulfilled. In this section, we are interested in the radius and the rates of convergence of $h_{1,n}^{\mathrm{lin}}(\sum_{i=1}^n \mu_i + c_n)$ and $h_{1,n}^{\star}(\sum_{i=1}^n \mu_i + c_n)$ to $\mathrm{E}[X_1]$ for different paths c_n such that $|c_n| \to \infty$ as $n \to \infty$.

5.1 Linear fair sharing rules

In the case of linear risk sharing rule, we obviously have

$$h_{1,n}^{\mathrm{lin}}\left(\sum_{i=1}^n \mu_i + c_n\right) = \mathrm{E}[X_1] + a_{1,n} c_n$$

and therefore $\lim_{n\to\infty} h_{1,n}^{\mathrm{lin}}\left(\sum_{i=1}^n \mu_i + c_n\right) = \mathrm{E}[X_1]$ if $c_n = o\left(1/a_{1,n}\right)$.

Example 5.1 (Regression rule). *When $a_{1,n} = \sigma_1^2/s_n^2$, we get $c_n = o\left(s_n^2\right)$.*

5.2 Conditional mean risk sharing rule

Zabell (1993) studied the behavior of random variables $\mathrm{E}[U|V_n+W_n]$ where the contribution of V_n to the sum $V_n + W_n$ is asymptotically negligible and both U and V_n are independent of W_n. Considering $U = V_n = X_1$ and $W_n = \sum_{i=2}^n X_i$ we find the problem investigated here. Zabell (1993) showed in his Theorem 4 that the radius of convergence c_n is smaller in general for the conditional mean risk sharing rule than for the regression risk sharing rule.

Proposition 5.2 (Zabell (1993)). *Assume that there exist $\alpha > 0$ and $0 < \beta < \infty$ such that for all $n \geq 1$, $s_n^2 \geq n\alpha$, $\mathrm{E}[|X_n - \mu_n|^{2+\delta}] \leq \beta$ for some $0 < \delta \leq 1$. Assume further that for every $T > 0$*

$$\int_{|t|>T} \prod_{j=1}^n \left|\mathrm{E}[\mathrm{e}^{\mathrm{i}tX_j}]\right| \mathrm{d}t = O\left(s_n^{-(2+\delta)}\right).$$

Then

$$h_{1,n}^{\star}\left(\sum_{i=1}^n \mu_i + c_n\right) = \mathrm{E}[X_1] + O\left(\frac{1}{s_n^{1+\delta}}\right) + O\left(\frac{|c_n|}{s_n^2}\right)$$

where $c_n = O\left(s_n\right)$.

Stronger results were derived by Zabell (1980) in the identically distributed case: under the additional assumption that $\mathrm{E}[\exp(\gamma X_1^\rho)] < \infty$ for some $\gamma > 0$ and $0 < \rho \leq 1$, it is shown there that $c_n = o(s_n^{2/(2-\rho)})$. Assuming that all moments of X_i exist, we prove that it is possible to have a radius of convergence such that $c_n = o(s_n^{5/3})$ in the heteregeneous case.

Let us now introduce the following technical conditions:

Condition D: $\sup_{x\geq 0} f_{X_i}(x) \leq C_i$ and $\inf_{i\geq 1} C_i > C > 0$.

Condition E: There exists a constant $K > 0$ such that

$$\mathrm{E}[|X_i - \mu_i|^k] \leq k! K^{k-2} \sigma_i^2, \qquad k = 3, 4, \ldots$$

Condition F: $\sup_{i\geq 1} \sigma_i^2 < \infty$.

Example 5.3 (Explanatory variables)**.** *Assume that Y_i given $\boldsymbol{Z}_i = \boldsymbol{z}_i$ has an Exponential distribution with parameter $\lambda_i = \lambda(\boldsymbol{z}_i)$ for some mesurable function λ. Assume further that there exists a positive constant K such that $\mu_i = \mathrm{E}[Y_i | \boldsymbol{Z}_i = \boldsymbol{z}_i] = 1/\lambda_i \leq K$, then Condition E holds, as well as Condition D with a constant $C < K^{-1}$.*

We are now in a position to state the following result for the conditional mean risk sharing rule.

Proposition 5.4. *Assume that Conditions A, D, E and F are fulfilled. If $c_n = o(s_n^{5/3})$ then*

$$h_{1,n}^{\star}\left(\sum_{i=1}^{n} \mu_i + c_n\right) = \mathrm{E}[X_1] + O\left(\frac{c_n^3}{s_n^5}\right).$$

The proof of Proposition 5.4 is given in appendix.

6 Conclusion

In this paper, we have studied the behavior of three risk sharing rules when the size of the pool becomes large: two linear fair sharing rules (proportional and linear regression ones) and the conditional mean risk sharing rule. Under mild technical conditions, we have established that (i) the individual contributions converge to the pure premium when the number of participants tends to infinity and (ii) the fluctuations of these contributions around the pure premium becomes ultimately Gaussian.

These results allowed us to identify the linear fair rule approximating the conditional mean risk sharing rule, providing practitioners with a useful simplification applicable within large pools. Also, the approximate number of participants needed to limit the variability of individual contributions to an acceptable level can be obtained from this central-limit theorem. This appears to be very useful for managing P2P insurance schemes since the volatility of participants' contributions is a crucial issue. The present paper offers practitioners an approximate number of participants required to reach this goal, that can be refined in a second stage using the actual distribution of participants' losses.

References

- Borch, K. (1962). Equilibrium in a reinsurance market. Econometrica 3, 424-444.

- Denuit, M. (2019). Size-biased transform and conditional mean risk sharing, with application to P2P insurance and tontines. ASTIN Bulletin 49, 591-617.

- Denuit, M., Dhaene, J. (2012). Convex order and comonotonic conditional mean risk sharing. Insurance: Mathematics and Economics 51, 265-270.

- Denuit, M., Robert, C.Y. (2020). Ultimate behavior of conditional mean risk sharing for independent losses. Submitted for publication.

- Donnelly, C. (2015). Actuarial fairness and solidarity in pooled annuity funds. ASTIN Bulletin 45, 49-74.

- Donnelly, C., Young, J. (2017). Product options for enhanced retirement income. British Actuarial Journal 22, 636-656.

- Furman, E., Zitikis, R. (2008). Weighted risk capital allocations. Insurance: Mathematics and Economics 43, 263-269.

- Petrov, V.V. (1956). A local theorem for densities of sums of independent random variables. Theory of Probability and Its Applications 1, 316-322.

- Petrov, V.V. (1995). Limit Theorems of Probability Theory: Sequences of Independent Random Variables. Oxford Studies in Probability.

- Saulis, L., Statulevicius, V. (1991). Limit Theorems for Large Deviations, Kluwer Academic Publishers, Dordrecht, The Netherlands.

- Schumacher, J.M. (2018). Linear versus nonlinear allocation rules in risk sharing under financial fairness. ASTIN Bulletin 48, 995-1024.

- Wittmann, R. (1985). A general law of iterated logarithm. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 68, 521–543.

- Zabell, S.L. (1980). Rates of convergence for conditional expectations. The Annals of Probability 8, 928-941.

- Zabell, S.L. (1993). A limit theorem for expectations conditional on a sum. Journal of Theoretical Probability 6, 267-283.

7 Appendix

7.1 Proof of Proposition 3.1

Kolmogorov's strong law of large numbers ensures that given a sequence $\{b_n,\ \ n=1,2,\ldots\}$ such that $b_n\to\infty$ as $n\to\infty$,

$$\sum_{i=1}^{\infty}\frac{\sigma_i^2}{b_i^2}<\infty\Rightarrow\frac{S_n-\mathrm{E}[S_n]}{b_n}\to 0 \text{ with probability 1.}$$

See, e.g., Theorem 6.7 in Petrov (1995). Now, since

$$h_{1,n}^{\mathrm{lin}}(S_n)=\mathrm{E}[X_1]+a_{1,n}\left(S_n-\mathrm{E}[S_n]\right),$$

the result then follows from Kolmogorov's strong law of large numbers with $b_n=1/a_{1,n}$. This ends the proof.

7.2 Proof of Proposition 3.6

Consider a non-negative random variable X with distribution function F_X and strictly positive expected value $\mathrm{E}[X]$. The size-biased transform of F_X is the distribution function $F_{\widetilde{X}}$ defined as

$$F_{\widetilde{X}}(x)=\frac{\mathrm{E}\big[X\mathrm{I}[X\leq x]\big]}{\mathrm{E}[X]},$$

where $\mathrm{I}[\cdot]$ denotes the indicator function (equal to 1 if the event appearing within the brackets is realized, and to 0 otherwise). In the case of absolutely continuity, when X has a positive probability density function f_X on $(0,\infty)$, $\widetilde{X}$ possesses the probability density function

$$f_{\widetilde{X}}(x)=\frac{xf_X(x)}{\mathrm{E}[X]}.$$

The moments of $\widetilde{X}$ are related to those of X by the relation

$$\mathrm{E}[\widetilde{X}^k]=\frac{\mathrm{E}[X^{k+1}]}{\mathrm{E}[X]}\text{ for }k=1,2,\ldots.$$

Let $\widetilde{X}_1$ be the size-biased versions of X_1, assumed to be independent and independent of $X_1,X_2,\ldots,X_n$.

i) Let us first assume that Condition A holds. It is proved in Denuit (2019a, Proposition 2.3), that, if $X_1,X_2,\ldots,X_n$ are absolutely continous random variables with respective probability density function $f_{X_1},f_{X_2},\ldots,f_{X_n}$, then for any $s>0$

$$h_{1,n}^{\star}(s)=\mathrm{E}[X_1|S_n=s]=\mathrm{E}[X_i]\frac{f_{S_n-X_1+\tilde{X}_1}(s)}{f_{S_n}(s)}. \tag{7.1}$$

The proof of Proposition 3.6 consists in approximating the density functions appearing in the numerator and denominator of (7.1) by standard Gaussian density function.

Define the sequences of constants

$$u_n = \sum_{i=1}^{n} \mu_i, \qquad u_{1,n} = \tilde{\mu}_1 + \sum_{j=2}^{n} \mu_j, \qquad s_{1,n}^2 = \tilde{\sigma}_1^2 + \sum_{j=2}^{n} \sigma_j^2,$$

and the sequences of random variables

$$Z_n = \frac{1}{s_n}\left(S_n - u_n\right), \qquad S_{1,n} = S_n - X_1 + \tilde{X}_1, \qquad Z_{1,n} = \frac{1}{s_{1,n}}\left(S_{1,n} - u_{1,n}\right).$$

With

$$x_n = u_n + s_n x \text{ and } x_{1,n} = u_{1,n} + s_{1,n} x,$$

we have

$$f_{Z_n}(x) = f_{S_n}(x_n)\, s_n \text{ and } f_{Z_{1,n}}(x) = f_{S_{1,n}}(x_{1,n})\, s_{1,n}.$$

By Theorem 2 in Petrov (1956) and Condition C, we know that there exists a positive constant C_0 such that, for all n sufficiently large, the inequalities

$$\left|f_{Z_n}(x) - \varphi(x)\right| \leq \frac{C_0}{\sqrt{n}} \text{ and } \left|f_{Z_{1,n}}(x) - \varphi(x)\right| \leq \frac{C_0}{\sqrt{n}}$$

are both valid, where the positive constant C_0 is independent of n and x, and where φ denotes the probability density function of the standard Gaussian distribution.

Since

$$x_n = u_{1,n} + s_{1,n}\left(\frac{s_n}{s_{1,n}}x - \frac{\widetilde{\mu}_1 - \mu_1}{s_{1,n}}\right),$$

we have

$$\begin{aligned} \frac{f_{S_{1,n}}(x_n)}{f_{S_n}(x_n)} &= \frac{f_{S_{1,n}}\left(u_{1,n} + s_{1,n}\left(\frac{s_n}{s_{1,n}}x - \frac{\widetilde{\mu}_1 - \mu_1}{s_{1,n}}\right)\right)}{f_{S_n}(x_n)} \\ &= \frac{s_n}{s_{1,n}} \frac{f_{Z_{1,n}}\left(\frac{s_n}{s_{1,n}}x - \frac{\widetilde{\mu}_1 - \mu_1}{s_{1,n}}\right)}{f_{Z_n}(x)} \\ &= \frac{s_n}{s_{1,n}} \frac{\varphi\left(\frac{s_n}{s_{1,n}}x - \frac{\widetilde{\mu}_1 - \mu_1}{s_{1,n}}\right) + O\left(1/n^{1/2}\right)}{\varphi(x) + O\left(1/n^{1/2}\right)}. \end{aligned}$$

Moreover

$$\begin{aligned} &\frac{\varphi\left(\frac{s_n}{s_{1,n}}x - \frac{\widetilde{\mu}_1 - \mu_1}{s_{1,n}}\right)}{\varphi(x)} \\ = \;& \exp\left(\frac{1}{s_{1,n}}\left(\widetilde{\mu}_1 - \mu_1\right)x + \frac{1}{2s_{1,n}^2}\left(x^2\left(\widetilde{\sigma}_1^2 - \sigma_1^2\right) - \left(\widetilde{\mu}_1 - \mu_1\right)\right)\right. \\ &\left. - \frac{1}{s_{1,n}^3}\left(\widetilde{\mu}_1 - \mu_1\right)\left(\widetilde{\sigma}_1^2 - \sigma_1^2\right)x\right). \qquad (7.2) \end{aligned}$$

Note also that

$$\frac{s_n}{s_{1,n}} = \sqrt{1 - \frac{\widetilde{\sigma}_1^2 - \sigma_1^2}{s_{1,n}^2}} = 1 + O\left(\frac{1}{s_{1,n}^2}\right).$$

Let

$$Y_n = \frac{Z_n}{\left(2\ln_2 s_n^2\right)^{1/2}} = \frac{\sum_{i=1}^n (X_i - \mu_i)}{\left(2s_n^2 \ln_2 s_n^2\right)^{1/2}}.$$

By Theorem 1.2 in Wittmann (1985), we know that Condition B ensures

$$\limsup_{n\to\infty} |Y_n| = 1 \text{ holds with probability 1.}$$

Since $\varphi(Z_n) = \varphi(Y_n (2\ln_2 s_n^2)^{1/2})$, we deduce that $1/\varphi(Z_n) = O(\ln s_n^2)$ with probability 1 and that

$$\begin{aligned} & \frac{\varphi\left(\frac{s_n}{s_{1,n}} Z_n - \frac{\widetilde{\mu}_1 - \mu_1}{s_{1,n}}\right) + O\left(1/n^{1/2}\right)}{\varphi(Z_n) + O(1/n^{1/2})} \\ = & \frac{\varphi\left(\frac{s_n}{s_{1,n}} Z_n - \frac{\widetilde{\mu}_1 - \mu_1}{s_{1,n}}\right) / \varphi(Z_n) + O\left(\ln s_n^2 / n^{1/2}\right)}{1 + O(\ln s_n^2 / n^{1/2})} \end{aligned}$$

with probability 1. Now, as $n \to \infty$,

$$\begin{aligned} & \frac{\varphi\left(\frac{s_n}{s_{1,n}} Z_n - \frac{\widetilde{\mu}_1 - \mu_1}{s_{1,n}}\right)}{\varphi(Z_n)} \\ = & \exp\left(\frac{\left(2\ln_2 s_n^2\right)^{1/2}}{s_{1,n}} (\widetilde{\mu}_1 - \mu_1) Y_n + \frac{1}{2s_{1,n}^2}\left(2\ln_2 s_n^2 Y_n^2 \left(\widetilde{\sigma}_1^2 - \sigma_1^2\right) - (\widetilde{\mu}_1 - \mu_1)\right)\right) \\ & \times \exp\left(-\frac{1}{s_{1,n}^3} (\widetilde{\mu}_1 - \mu_1)\left(\widetilde{\sigma}_1^2 - \sigma_1^2\right)\left(2\ln_2 s_n^2\right)^{1/2} Y_n\right) \\ \to & \ 1 \text{ with probability 1.} \end{aligned}$$

We finally deduce that

$$\lim_{n\to\infty} \mathrm{E}[X_1|S_n] = \mathrm{E}[X_1] \text{ with probability 1,}$$

as announced.

ii) Let us now assume that Condition A' holds. It is proved in Denuit and Robert (2020) that, for $s > 0$,

$$\mathrm{E}\left[X_1 | S_n = s\right] = \mathrm{E}\left[X_1\right] \frac{f_{S_n + \widetilde{C}_1}(s)}{\mathrm{P}\left[S_n > 0\right] f_{S_n | S_n > 0}(s)}. \tag{7.3}$$

Since for all i, $\lambda_i > \eta$, we have $\mathrm{P}\left[S_n = 0\right] \leq e^{-n\eta}$. Moreover

$$s_n f_{S_n|S_n>0}(x_n) = f_{Z_n|Z_n > -u_n/s_n}(x)$$

and it also holds that, for all n sufficiently large,

$$\left| f_{Z_n|Z_n>-u_n/s_n}(x) - \varphi(x)\right| \leq \frac{C_0}{\sqrt{n}}.$$

The same reasoning as previously can therefore be followed.

7.3 Proof of Proposition 4.4

We use the same notation as in the proof of Proposition 3.6. Note that by Condition C, Lyapunov's condition holds and therefore

$$Z_n \xrightarrow{\mathcal{L}} \text{Normal}\,(0,1)\,.$$

Moreover by

$$h_{1,n}^{\star}(S_n) = \mathrm{E}[X_1]\frac{s_n}{s_{1,n}}\frac{\varphi\left(\frac{s_n}{s_{1,n}}Z_n - \frac{\widetilde{\mu}_1-\mu_1}{s_{1,n}}\right)/\varphi\left(Z_n\right) + O\left(\ln s_n^2/n^{1/2}\right)}{1+O\left(\ln s_n^2/n^{1/2}\right)}$$

with probability 1. Equation (7.2) leads to

$$\begin{aligned} h_{1,n}^{\star}(S_n) - \mathrm{E}[X_1] &= \frac{\sigma_1^2}{s_{1,n}}Z_n + O\left(\frac{\ln_2 s_n^2}{s_{1,n}^2}\right) \\ &= \frac{\sigma_1^2}{s_n}Z_n + O\left(\frac{\ln_2 s_n^2}{s_{1,n}^2}\right) \end{aligned}$$

with probability 1. It follows that

$$\frac{s_n}{\sigma_1^2}\left(h_{1,n}^{\star}(S_n) - \mathrm{E}[X_1]\right) = Z_n + O\left(\frac{\ln_2 s_n^2}{s_{1,n}}\right).$$

This ends the proof.

7.4 Proof of Proposition 5.4

Let $Z_n = \left(S_n - u_n\right)/s_n$. By Theorem 6.1 in Saulis and Statulevicius (1991), we have, for x such that $|x| \leq c_n/s_n = o\left(s_n\right)$,

$$\frac{f_{Z_n}\left(x\right)}{\varphi\left(x\right)} = \exp\left(\frac{x^3}{s_n}\lambda_n\left(\frac{x}{s_n}\right)\right)\left(1+O\left(\frac{x}{s_n}\right)\right)$$

where

$$\lambda_n\left(x\right) = \sum_{k=0}^{\infty}\gamma_{k,n}x^k$$

is a power series which converges in a certain neighbourhood of the origin and $\gamma_{k,n}$ are expressed in terms of cumulants of Z_n. In particular

$$\gamma_{0,n} = \frac{1}{6s_n^2}\sum_{i=1}^{n}\Gamma_3\left(X_i - \mu_i\right)$$

where $\Gamma_3\left(X\right)$ is the third cumulant of the random variable X. Note that, by Condition E, $\sup_{n\geq 0}\left|\gamma_{0,n}\right| < \infty$.

Using the same notation as in the proof of Proposition 3.6, we have

$$\frac{f_{Z_{1,n}}(x)}{\varphi(x)} = \exp\left(\frac{x^3}{s_{1,n}}\lambda_{1,n}\left(\frac{x}{s_{1,n}}\right)\right)\left(1+O\left(\frac{x}{s_{1,n}}\right)\right)$$

with

$$\lambda_{1,n}(x) = \sum_{k=0}^{\infty}\gamma_{1,k,n}x^k$$

and

$$\gamma_{1,0,n} = \frac{1}{6s_{1,n}^2}\left(\Gamma_3\left(\widetilde{X}_1-\widetilde{\mu}_1\right)+\sum_{i=2}^{n}\Gamma_3(X_i-\mu_i)\right).$$

We also have

$$\mathrm{E}[X_1|S_n = u_n + xs_n] = \mathrm{E}[X_1]\frac{s_n}{s_{1,n}}\frac{f_{Z_{1,n}}(I_n(x))}{f_{Z_n}(x)}.$$

with

$$I_n(x) = \frac{s_n}{s_{1,n}}x - \frac{\widetilde{\mu}_1-\mu_1}{s_{1,n}}.$$

Let us recall that

$$s_{1,n}^2 = s_n^2 + \left(\widetilde{\sigma}_1^2-\sigma_1^2\right) = s_n^2\left(1+\frac{\widetilde{\sigma}_1^2-\sigma_1^2}{s_n^2}\right)$$

and therefore

$$s_{1,n} = s_n\left(1+O\left(s_n^{-2}\right)\right).$$

We have

$$\frac{f_{Z_{1,n}}(I_n(x))}{f_{Z_n}(x)} = \left(1+O\left(\frac{x}{s_n}\right)\right)\exp\left(\frac{I_n^3(x)}{s_{1,n}}\lambda_{1,n}\left(\frac{I_n(x)}{s_{1,n}}\right)-\frac{x^3}{s_n}\lambda_n\left(\frac{x}{s_n}\right)\right).$$

Note that

$$\lambda_n\left(\frac{x}{s_n}\right) = \gamma_{0,n}+O\left(\frac{x}{s_n}\right)$$

and

$$\begin{aligned}
\lambda_{1,n}\left(\frac{I_n(x)}{s_{1,n}}\right) &= \gamma_{1,0,n}+O\left(\frac{I_n(x)}{s_{1,n}}\right)\\
&= \gamma_{0,n}+\frac{1}{6s_{1,n}^2}\left(\Gamma_3\left(\widetilde{X}_1-\widetilde{\mu}_1\right)-\Gamma_3(X_1-\mu_1)\right)+O\left(\frac{x}{s_n}\right)\\
&= \gamma_{0,n}+O\left(\frac{x}{s_n}\right).
\end{aligned}$$

Moreover

$$I_n^3(x) = x^3\left(1+O\left(\frac{1}{s_n}\right)\right)$$

and it follows that

$$\begin{aligned} & \exp\left(\frac{I_n^3\left(x\right)}{s_{1,n}}\lambda_{1,n}\left(\frac{I_n\left(x\right)}{s_{1,n}}\right)-\frac{x^3}{s_n}\lambda_n\left(\frac{x}{s_n}\right)\right) \\ = & \exp\left(\frac{x^3}{s_n}\left(1+O\left(\frac{1}{s_n}\right)\right)\left(\gamma_{0,n}+O\left(\frac{x}{s_n}\right)\right)-\frac{x^3}{s_n}\left(\gamma_{0,n}+O\left(\frac{x}{s_n}\right)\right)\right) \\ = & \exp\left(O\left(\frac{x^3}{s_n^2}\right)\right). \end{aligned}$$

If $c_n = o(s_n^{5/3})$, we therefore have

$$h_{1,n}^*(s_n) = \mathrm{E}[X_1] + O\left(\frac{c_n^3}{s_n^5}\right),$$

as announced.