24. SCIENTIFIC STATISTICAL SEMINAR
"MARBURG−WROCŁAW"
KOWARY, 25-28 SEPTEMBER 2017
EXTENDED ABSTRACTS

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY
Nr 16(22)

ISSN 1644-6739
e-ISSN 2449-9765

DOI: 10.15611/sps.2018.16.09

MARRIAGE REVERSE ANNUITY CONTRACT WITH DEPENDENCE BETWEEN SPOUSES[1]

Joanna Dębicka, Stanisław Heilpern, Agnieszka Marciniuk
(Wrocław University of Economics)

According to the reverse annuity contract, an owner receives benefits in exchange for renunciation of his/her rights to their real estate to a company (a mortgage fund in Poland), created especially for this purpose. This contract is offered to elderly people. The owner is guaranteed by a notarial act, the right to stay in the property until his/her death, however he/she formally is not the owner of the property. In Poland, this contract is not the same as reverse mortgages (cf. [Dębicka, Marciniuk 2014]). Our research concerns the marriage reverse annuity contract, which is a new product and is not currently offered in Poland.

In our research we concentrate on the marriage reverse annuity contract, which is a variation of the individual reverse annuity contract. Under these contracts, annuity benefits are payable when both spouses are alive and sometimes after the death of whichever spouse. Thus we distinguish between two types of such contracts: a *joint-life status contract* (JLS), when the benefit is paid only until the death of the first spouse and a *last surviving status contract* (LSS) by which the benefit is paid until the death of the other spouse.

We focus on a discrete-time model, where an annuity is paid at the beginning of particular time units. We assume that the evolution of the contracted risk is described by time-nonhomogeneous Markov chain. Moreover, actuarial values are considered under the assumption of stochastic interest rate.

[1] The research developed within the grant scheme Non-Standard Multilife Insurance Products with Dependence between Insured 2013/ 9/B/HS4/00490.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

The aim of this research is to model the probabilistic structure and cash flows arising from marriage reverse annuity contracts. Note that, the analysis of multi-life contracts is usually based on the assumption of the independence between the insured. In contrast to this classical approach, dependence of future lifetimes between the spouses is assumed. This is more realistic assumption than independence, because the husband and the wife are exposed to the same risks, which cause the dependence of their future lifetimes. We may observe the influence of the death of a spouse on the future lifetime of the second one. The "broken heart syndrome" occurs in such a situation. Moreover, a common external event called "shock", e.g. a car or airplane crash which causes the death of both spouses, may occur. We applied copulas to model the dependence structure of the length of the spouses' lives. The modelling of spouses' lifelengths taking into account the dependence by the use of copula was considered for example in [Carriere 2000; Gourieroux, Lu 2015; Luciano et al. 2008; Luciano et al. 2016; Spreeuw 2006]. Moreover in the article [Gourieroux, Lu 2015; Luciano et al. 2016; Spreeuw 2006] the authors have applied these copulas to determine the actuarial value of the different kinds of life annuities. However, according to our considerations, no application of these copulas to the calculation of the value of the marriage reverse annuity contract has been found in the literature. We hope that this article will bridge this gap in the research of marriage insurance.

Multiple state modelling is a stochastic tool to design and implement an insurance contract (c.f. [Haberman, Pitacco 1999; Pitacco 2014]). Then we apply the multiple state model for marriage insurances to model the marriage reverse annuity contract. This allows us to determine the benefit of the marriage reverse annuity contract based on matrix formulas for the first moments of cash flows arising from the contract. Matrix notation allows for the efficient analysis of the stochastic structure of the model and cash flows resulting from the realization of the contracts. Moreover, this tool facilitates the analysis of the impact of the probabilistic structure of the model (with dependence and independence) of a marriage reverse annuity contract on the actuarial value of annual benefit for both the joint-life status and the last surviving status.

In order to apply a matrix notation to the analysis of annual benefits for the marriage reverse annuity contract, we have to follow two steps. Firstly, we have to derive a matrix which describes the probabilistic structure of the model during the whole contract period under the assumption that dependence between the future lifetimes of the spouses is modelled by the copula. Secondly, we derive a general matrix formula for annuity benefits paid in more than one state, which can be applied to any type of contract being modelled by the multiple state model. In the case of a joint-life status contract, when the annuity benefit is paid in advance under condition that both spouses are alive, the valuation of the benefits may be done analogously to the results derived in [Dębicka 2013] for the valuation of period premiums. Last surviving status contract is more advanced, allowing annuity benefits to be paid to both living spouses and when one of them is alive. This situation goes beyond the scope of the models analysed in [Dębicka 2013] and needs a different approach. We suggest solving this problem by the use of the graph optimization methodology to find the shortest path between particular states of the model.

The main result of our research is a theorem, where formulas for elements of the matrix describing the probabilistic structure of the model in the whole contract under the assumption that the future lifetimes of spouses are dependent and modelled by copulas are derived. In the next theorem we derive general matrix expressions for the annuity due benefits paid in more than one state. As a corollary the matrix formula for the benefit of the marriage reverse annuity contract is given.

We present some numerical examples. The type of copulas are determined on the basis of actual data for the Lower Silesia Voivodship and data which comes from Wroclaw cemeteries. To model the spot interest rate we based on actual Polish market data related to the yield to maturity on zero-coupon and fixed interest bonds. For our calculations, we apply the Svensson model of spot interest rate, whose parameters are estimated by using the least-squares method. The calculations are made by the use of own programs written in MATLAB. We compare the results obtained for the dependence and independence of future lifetimes of spouses. Summing up, generally the benefit is much higher for JLS than for LSS. The function of the benefit is convex for JLS contrary to LSS. The benefit is higher for independent than for the dependent future lifetimes of spouses in cases of JLS, which is more profitable for customers. In the case of LSS, the benefits for younger people are not higher in cases of independence. However, for elderly people the result is similar. The relative increase between the benefits for independent and dependent is considerable.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

References

Carriere J.F., 2000, *Bivariate survival models for coupled lives*, Scandinavian Actuarial Journal, vol. 1, pp. 17-31.

Dębicka J., 2013, *An approach to the study of multistate insurance contracts*, Appl. Stochastic Models Bus. Ind., vol. 29, iss. 3, pp. 224-240.

Dębicka J., Marciniuk A., 2014, *Comparison of reverse annuity contract and reverse mortgage on the Polish market*, 17-th AMSE, Applications of Mathematics in Economics. International Scientific Conference: Poland, 27-31 August, 2014, Conference Proceedings. Wroclaw University Press, 2014, pp. 55-64, ISBN 978-83-7695-421-9.

Gourieroux C., Lu Y., 2015, *Love and death: A Freund model with frailty*, Insurance: Mathematics and Economics, vol. 63, pp. 191-203.

Haberman S., Pitacco E., 1999, *Actuarial Models for Disability Insurance*, Chapman & Hall/CRC.

Heilpern S., 2015, *Dependent structure induced by Markov chain in the multiple life insurance*, 18-th AMSE. Applications of Mathematics in Economics. International Scientific Conference: Jindřichův Hradec 2-6 September, 2015, Conference Proceedings, http://amse2015.cz/proceedings.

Luciano E., Spreeuw J., Vigna E., 2008, *Modeling stochastic mortality for dependent lives*, Insurance: Mathematics and Economics, vol. 43, pp. 234-244.

Luciano E., Spreeuw J., Vigna E., 2016, *Spouses' Dependence across Generations and Pricing Impact on Reversionary Annuities*, Risks, vol. 4(2), pp. 1-18.

Pitacco E., 2014, *Health Insurance. Basic Actuarial Models*, EAA Series, Springer.

Spreeuw J., 2006, *Types of dependence and time-dependent association between two times in single parameter copula models*, Scandinavian Actuarial Journal, vol. 5, pp. 286-309.

PRICING AND PROFITABILITY OF MARRIAGE INSURANCE CONTRACT WITH MORTALITY DEPENDENCE[2]

Joanna Dębicka, Stanisław Heilpern, Agnieszka Marciniuk
(Wrocław University of Economics)

An important variation of individual life insurance is the so called joint life insurance where multiple lives are involved. Under these contracts, death benefits are payable on whichever death within a group. A special case of multi-life insurance contract are marriage insurance contracts which protect against the serious financial impact that results from a spouse's death. We distinguish between two types

[2] The research developed within the grant scheme Non-Standard Multilife Insurance Products with Dependence between Insured 2013/09/B/HS4/00490.

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY

Nr 16(22)

of such contracts: a *joint-life status insurance* issued to a married couple where the death benefit is paid if the husband or wife dies and a *last surviving status insurance* by which the lump sum is paid after the death of each of the spouses.

The analysis of marriage insurance contracts is usually based on the assumption of the independence between the insured. In contrast to this classical approach, dependence of future lifetimes between the spouses is assumed. This is more realistic assumption than independence, because the husband and the wife are exposed to the same risks, which cause the dependence of their future lifetimes. We applied copulas to model the dependence structure of the length of the spouses' lives and apply these copulas to the valuation of the marriage life insurance contract consisting of different types of benefits, such as lump sum and annuity.

We focus on a discrete-time model, where insurance payments are made at the ends of time units. We assume that the evolution of the insured risk is described by time-nonhomogeneous Markov chain. Moreover, actuarial values are considered under the assumption of stochastic interest rate.

We applied and developed a matrix representation to the analysis of all types of future cash flows arising from the marriage insurance contract. For this purpose we derive transition probabilities under the assumption that the future lifetimes of spouses are dependent and such dependence is modelled by a copula. The appropriate accommodation of the modified multiple state model enables us to obtain matrix formulas for actuarial values of multi-life insurance contract under assumptions on dependence between the insured. Matrix notation provides a compact form for both the joint-life status and the last surviving status benefits [Dębicka et al. 2016b].

Apart from actuarial values related to the pricing insurance contract (such as premiums and reserves), the relevant elements of the set of the contract's terms are the expected profit and the expected cash flows for each year of the insurance policy. These values allow to determine the range of profit, which is taken into account in the process of adjusting the features of a contract, which is called profit testing. The new solvency regime of the European Union (Solvency II) uses *worst-case scenarios* (also called the *first-order basis*) for the calculation of solvency capital requirements for the life insurance business. In parallel, the premiums and reserves are calculated (securely with some surplus) under the assumptions that:

I-1. the rate of interest equals the long-term rate, which is determined on the basis of real Polish market data by the use of the Svensson model of yield curve (cf. [Dębicka et al. 2016a]),

I-2. the future lifetime of spouses are independent random variables and to count the distribution of their future lifetime we use Polish Life Tables for the Lower Silesia Voivodeship in Poland.

The *second-order basis* is the determination of the insurer's profit and loss on the basis of realistic parameters. We take into account that

II-1. the interest rate has a greater value of than in the first-order basis,

II-2. the future lifetimes of spouses are dependent random variables and the distribution of their future lifetime is modeled by the Gumbel copula (cf. [Heilpern 2015]).

In the numerical examples, we analyze the impact of the probabilistic structure of the model and the value of the interest rate on the particular elements of the valuation and the profit testing of marriage life insurance. Let pair (x, y) denote the age at entry of husband and wife. We made calculations for a husband and wife aged 60 or 65 i.e. $x, y \in \{60,65\}$.

Under assumption I-1, it appears that the premiums are lower on condition that the future lifetimes of spouses are dependent. In the age of asymmetric case, the net single premium is higher when the wife is older and the net period premium is higher when the husband is older. Moreover net premium prospective reserves (reserves in cases when both spouses are alive) are lower on condition that the future lifetimes of spouses are dependent (II-2). On the same condition, net prospective reserves for cases that only one of the spouses is alive are higher.

In the second part of the numerical analysis, we would like to observe the influence of the probabilistic structure for the second-order basis on the expected cash flows and the expected profit to emerge in the whole insurance period depending on age at entry of the spouses [Dębicka et al. 2016a]. We would like to investigate only the impact of the probability structure, so we assume that the interest rates in assumptions I-1 and II-1 are equal. It appears that in the case of peers, expected profit is greater in the first half of the insurance period for the elderly spouses, later the situation is changing. Moreover, in the case of married couples where the husband is younger, expected profit is higher than for couples in which the man is older.

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY

Nr 16(22)

References

Dębicka J., Heilpern S., Marciniuk A., 2016a, *Profitability of marriage insurance contract,* 19-th AMSE, Applications of Mathematics in Economics, International Scientific Conference: Banská Štiavnica, Slovakia, 31 Aug.-4 Sep. 2016, Conference Proceedings.

Dębicka J., Heilpern S., Marciniuk A., 2016b, *The application of copulas to the modelling of the marriage reverse annuity contract,* manuscript https://www.actuaries.org/lisbon2016/papers/20.JoannaDebickaPaper.pdf.

Heilpern S., 2015, *Dependent structure induced by Markov chain in multiple life insurance*. 18-th AMSE, Applications of Mathematics in Economics. International Scientific Conference: Jindřichův Hradec 2-6 September, 2015, Conference Proceedings, http://amse2015.cz/proceedings.

ENABLING THE RELEASE OF CONFIDENTIAL INSTITUTIONAL DATA BY THE POST-RANDOMIZATION METHOD

Heiko Grönitz (Philipps-Universität Marburg)

1. Introduction

Let us assume that an institution possesses micro-data for a number of persons and several characteristics. These characteristics comprise identifying variables, background variables and confidential variables. By the identifying variables, we can unambiguously detect the person that belongs to a data row in the dataset. Examples for such variables might be name, address, social security number or tax number. Background variables are socio-demographic attributes such as gender, age, ethnicity, marital status, geographic region, education and religion. Confidential variables contain private information on the persons. One may think of income, property, medical records with information on illnesses or expenditures for medical bills, bank rating or socially undesired behavior such as low tax morality.

We further assume that this institution is in principle willing to share or publish the data. Before sharing or publishing the data, the institution has to take data protection into account. That is, the institution must ensure that it is not possible to find out the confidential information of individuals with the help of the released data. This duty of the institution is regulated in the laws of data protection.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

To fulfill this obligation of non-identifiability, a statistical disclosure control (SDC) must be operated. Several SDC methods exist. The first step in a SDC is normally to delete the identifying variables. However, after removing the direct identifiers, persons might still be identifiable. On the one hand, rare combinations of the background characteristics or, one the other hand, values of the confidential variables itself (e.g. very high income) can disclose the person that belongs to a data row. For this reason, further SDC methods are usually used before releasing the data.

In this paper we consider one such technique in detail, namely the post-randomization method (PRAM). Here, the values of one or more variables are altered according to a certain probability mechanism. This alteration is conducted post hoc, i.e. after the original data have been collected. By applying the PRAM to the original data, a new dataset with indirect data arises. This new dataset can be shared with other institutions or published for free disposal. In addition to the indirect data, the probability scheme that is applied for the transformation of the original data must be revealed by the institution that provides the data. In this article, we present some inference techniques for datasets with indirect data generated by the PRAM.

2. Post-randomization method applied to categorical confidential variables

Let us say the direct identifiers were already deleted. Then, a data set with p background characteristics $X_1,...,X_p$ and v confidential attributes $Y_1,...,Y_v$ remains. Regarding the confidential attributes, we assume throughout that they are categorical and that Y_i has the categories $1,...,k_i$. We consider a situation that the background characteristics are left unchanged and that each confidential characteristic is transformed with the help of the PRAM. In other words, the confidential variables become misclassified. The particular procedure for the transformation is as follows: for each value of Y_i, a random number is generated according to a given probability distribution. With this random number, the value of Y_i is changed to an indirect value according to a given scheme. Let Z_i be the variable with the indirect data corresponding to Y_i.

The distribution of Z_i and the distribution of Y_i are coupled by a transition matrix, which contains the chances that Y_i value j is changed to the value 1 $\left(j = 1,...,k_i\right)$.

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY

Nr 16(22)

3. Estimation of the distribution of one confidential variable

Let us say, $v = 1$ holds, i.e., we have a single confidential variable Y_1 in the dataset. Based on the values of Z_1, we can set up the log-likelihood function for the distribution of Y_1. In the log-likelihood, the transition matrix T_1 appears. Regarding T_1 some regularity conditions are needed. The maximization of the log-likelihood can be conducted, for example, by Fisher scoring, the EM algorithm, or the simplex algorithm for nonlinear problems. By differentiating the log-likelihood twice, we obtain the Fisher matrix, which yields an asymptotic variance estimation for the maximum likelihood (ML) estimator. Another variance estimation can be obtained via the bootstrap approach.

4. Joint distribution of multiple confidential variables

The joint distribution of $Y_1,...,Y_v$ and the joint distribution of the variables with the indirect data $Z_1,...,Z_v$ are connected by the Kronecker product of the v transition matrices $T_1,...,T_v$. With this fact we can set up the log-likelihood function for the joint distribution of $Y_1,...,Y_v$ based on the indirect data, i.e. the data for $Z_1,...,Z_v$. For the maximization of this log-likelihood, we can apply Fisher scoring, the EM algorithm, or the simplex algorithm again. To estimate the estimation variance, we present an asymptotic approach and a bootstrap approach. From the estimate for the joint distribution, we obtain estimates for the marginal distributions by summation.

5. Joint distribution of categorical background variables and confidential variables

The derivations of the previous section can be transferred to a situation in which we also have categorical background variables and the joint distribution of the background variables and the confidential variables is of interest. As mentioned before, each background characteristic X_i $(i = 1,..., p)$ is left as it is. However, this can be interpreted as transformation with the identity matrix as the transition matrix. Hence, the joint distribution of $X_1,..., X_p$, $Z_1,...,Z_v$ and the joint distribution of $X_1,..., X_p$, $Y_1,...,Y_v$ are coupled by the Kronecker

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

product of p identity matrices and the matrices $T_1,...,T_v$. With this, we can accomplish inference analog to the previous section.

6. Logistic regression inference

In this section we consider arbitrary background characteristics $X_1,...,X_p$ (categorical or continuous) and one confidential variable Y_1. Moreover, we assume that the influence of the background characteristics on the confidential attribute follows a logistic regression model. To estimate the parameters of the logistic regression model we consider the log-likelihood function based on the values of the covariates $X_1,...,X_p$ and the indirect data Z_1. We present the maximization of the log-likelihood via the EM algorithm where other techniques are also possible. For the estimation of the estimation variance we present a bootstrap procedure. Once the parameters of the logistic regression model have been obtained, we can estimate the conditional proportions of persons having Y_1 value y given the exogenous variables by inserting in the model equation.

7. Ordinal regressions

Often the categories of Y_1 can be ordered (e.g. low expenditures for medical bills, medium expenditures, high expenditures). For this reason, we consider some ordinal regression models for the dependence of Y_1 on the covariates $X_1,...,X_p$. These models possess less parameters compared with the logistic regression model. For each considered ordinal model, the ML inference is discussed.

8. Concluding remarks

The PRAM offers the possibility to release confidential micro-data without hurting data protection. This article showed for several situations how inference methods for direct data can be adjusted to indirect data resulting from the application of the PRAM. The price for the data protection by the PRAM is that the estimators possess a larger variance than in the case of direct data. Since the specifications of the applied PRAM (i.e. distributions of the random numbers and the transformation scheme) are chosen by the data proving institution, this institution can control the increase of the estimators' variance.

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY

Nr 16(22)

MODELLING OF DEPENDENT STRUCTURE. APPLICATION IN INSURANCE

Stanisław Heilpern (Wrocław University of Economics)

We will use the copulas and Markov chains to the modelling of a dependent structure. The copula is the link between the joint cumulative distribution function (c.d.f.) F and the marginal c.d.f. F_i [Nelsen 1999; Heilpern 2007] i.e.

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$

In the applications in insurance we often use the survival copula C^*:

$$S(x_1, \ldots, x_n) = C^*(S_1(x_1), \ldots, S_n(x_n)),$$

where $S(x_1, \ldots, x_n) = P(X_1 > x_1, \ldots, X_n > x_n)$ and $S_i(x_i) = P(X_i > x_i)$. For $n = 2$ we obtain the following relation between these copulas: $C^*(u, v) = u + v - 1 + C(1 - u, 1 - v)$. Knowing copulas we can determine the rank coefficients of correlations: Kendall τ and Spearman ρ

$$\tau = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1, \qquad \rho = 12\int_0^1\int_0^1 C(u,v)dudv - 3.$$

For independent random variables (r.v.) we have the copula $\Pi(u, v) = uv$, for strict, positive dependence we obtain $M(u, v) = \min\{u, v\}$ and for strict, negative dependence $W(u, v) = \max\{u + v - 1, 0\}$.

Copula C is Archimedean if it takes the form:

$$\varphi(C(u, v)) = \varphi(u) + \varphi(v),$$

where the generator φ: $[0, 1] \to [0, \infty]$ satisfies the conditions $\varphi(1) = 0$ and it is decreasing and it is a convex function. In practice we often use the families of copulas indexes with some coefficients, e.g. the Frank, Clayton, Gumbel and Ali-Mikhail-Morgenstern families. These are the Archimedean copulas. We can define the multidimensional Archimedean copulas by the formula: $\varphi(C(u_1, \ldots, u_n)) = \varphi(u_1) + \ldots + \varphi(u_n)$, but every pair X_i, X_j must be identically, positive correlated in this case. This significantly narrows the applicability of such copulas. For the Archimedean copula there is the latent r.v. Θ (fraitly), such that the moment generating function $M_\Theta(s) = \varphi^{-1}(-s)$ and r.v. X_i are conditionally dependent for fixed θ, i.e.

$$P(X_1 > x_1, \ldots, X_n > x_n \mid \Theta = \theta) = S_{1|\theta}(x_1) \ldots S_{n|\theta}(x_n),$$

where $S_{i|\theta}(x_i) = P(X_i > x_i \mid \Theta = \theta) = \exp(\theta M_\Theta^{-1}(S_i(x_i)))$.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

The elliptical copulas are another family of copulas. They are generated by multidimensional elliptical distributions Φ, e.g. Gauss, Student-t, logistic and Laplace:

$$C(u_1,...,u_n)=\Phi_n(\Phi_1^{-1}(u_1),...,\Phi_1^{-1}(u_n)).$$

We will also use the Spearman copula $Sp(u, v) = (1 - \rho)\Pi(u, v) + \rho M(u, v)$, where ρ is the Spearman coefficient of correlation.

Marriage insurance

Let T_x^M be the remaining lifetimes of an x-year old man and T_y^W such lifetimes of a y-year old woman. Unlike the classical approach, we will assume that the lifetimes of spouses may be dependent. Let the joint distribution of them be characterized by the copula C_{xy}. Based on the empirical data from Brussels, Denuit et al. [2001] selected the Gumbel copula $C(u, v) = \exp(-((-\ln u)^{\alpha} + (-\ln v)^{\alpha})^{1/\alpha})$, where $\alpha = 1{,}1015$. Using the data from Wroclaw [Heilpern 2011], the corresponding copula is the AMH copula $C(u, v) = \dfrac{uv}{1-\alpha(1-u)(1-v)}$ with $\alpha = 0{,}5879$.

We can describe the joint-life status by the formula

$$
\begin{aligned}
{}_t p_{xy} &= P(T_x^M > t, T_y^W > t) = C_{xy}^*({}_t p_x^M, {}_t p_y^W)\\
&= P(T_0^M > x+t, T_0^W > y+t \mid T_0^M > x, T_0^W > y) = \frac{C_{00}^*({}_{x+t}p_0^M, {}_{y+t}p_0^W)}{C_{00}^*({}_x p_0^M, {}_y p_0^W)},
\end{aligned}
$$

where ${}_t p_x^M = P(T_x^M > t) = \exp\left(-\int_0^t \mu_{x+s}^M ds\right)$, μ_x^M is force of mortality, ${}_t p_y^W = P(T_y^W > t)$ and we present the last-survival status by the formula

$$
\begin{aligned}
{}_t p_{\overline{xy}} &= P(\max\{T_x^M, T_y^W\} > t) = 1 - C_{xy}(1-{}_t p_x^M, 1-{}_t p_y^W)\\
&= \frac{C_{00}^*({}_{x+t}p_0^M, {}_y p_0^W) + C_{00}^*({}_x p_0^M, {}_{y+t}p_0^W) - C_{00}^*({}_{x+t}p_0^M, {}_{y+t}p_0^W)}{C_{00}^*({}_x p_0^M, {}_y p_0^W)}
\end{aligned}
$$

and ${}_t p_{x|y} = P(T_x^M \le t, T_y^W > t) = \dfrac{C_{00}^*({}_x p_0^M, {}_{y+t}p_0^W) - C_{00}^*({}_{x+t}p_0^M, {}_{y+t}p_0^W)}{C_{00}^*({}_x p_0^M, {}_y p_0^W)}$. These probabilities let us derive the annuities: the joint-life $a_{xy:\overline{n}|} = \sum_{k=1}^{n} v^k\, {}_k p_{xy}$, the last-survival $a_{\overline{xy}:\overline{n}|} = \sum_{k=1}^{n} v^k\, {}_k p_{\overline{xy}}$ and the widow's pension $a_{x|y} = \sum_{k=1}^{w_y^W} v^k\, {}_k p_{x|y}$, where v is the discounting factor and w_y^K is the difference between the age limit of woman and age y.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

If we have the aggregate data, e.g. from the statistical office (GUS), we can use the Markov model. We have 4 states: 0 – both spouses alive, 1 – husband dead, 2 – wife dead and 3 – both spouses dead and the transition of probability done by formulas:

$$p_{00}(t,s) = \exp\left(-\int_t^s (\mu_{01}(u) + \mu_{02}(u))du\right),$$

$$p_{ii}(t,s) = \exp\left(-\int_t^s \mu_{i3}(u)du\right),$$

$$p_{0i}(t,s) = \int_t^s p_{00}(t,u)\mu_{0i}(u)p_{ii}(u,s)du .$$

Denuit et al. [2001] assumed the following relation between the forces of mortalities, Markovian μ_{ij} and the standard μ_x^M , μ_y^K :

$$\mu_{01}(t) = (1 + a_{01})\mu_{x+t}^M, \qquad \mu_{23}(t) = (1 + a_{23})\mu_{x+t}^M, \qquad \mu_{02}(t) = (1 + a_{02})\mu_{y+t}^W,$$

$$\mu_{13}(t) = (1 + a_{13})\mu_{y+t}^W.$$

The joint-life status is done by formula [Heilpern 2011]

$${}_k p_{xy}(t) = \left({}_t p_x^M\right)^{1-a01} \left({}_t p_y^W\right)^{1-a02} .$$

So we can derive the annuities. We estimate the parameters a_{ij} using the Nelson-Aalen estimator.

Risk processes

Let W_i be the identically distributed the interclaim times with $E(W_i) = 1/\lambda$, X_i be the identically distributed the claim amounts with $E(X_i) = 1/\beta$ and $N(t) = \sum_{i=1}^{\infty} 1_{T_k \le t}$, where $T_n = \sum_{i=1}^{n} W_i$, is the counting claims process. The risk process is defining by formula

$$U(t) = u + ct - \sum_{i=1}^{N(t)} X_i ,$$

where u is initial capital and c premium rate. We will investigate the probability of ruin $\psi(u) = P(T < \infty | U(0) = u)$, where $T = \inf\{t: U(t) < 0\}$ is the time of ruin, and we assume that $c\beta > \lambda$.

In the classical approach [Rolski et al. 1999] we assume that all r.v. are independent and $N(t)$ is the Poisson process, i.e. W_i have the exponential distribution. Then we obtain $\psi(0) = \lambda/(c\beta)$ and $\psi(\infty) = 0$. In the Sparre Andersen model the W_i can have any distribution.

Next we assume that the claims X_i may be dependent and the dependent structure is described by the Archimedean copula C. We

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

obtain for fixed value θ of the latent variable Θ the risk process U_θ with the independent claims $X_{i|\theta}$. The conditional probability of ruin is equal to [Heilpern 2010]

$$\psi(u)=\int_{\theta_0}^{\infty}\psi_\theta(u)dF_\Theta(\theta)+F_\Theta(\theta_0)$$

in this case, where $\psi_\theta(u)$ is the conditional probability of ruin and θ_0 is the solution of equation $m(\theta) = c/\lambda$, $m(\theta) = E(X_{i|\theta})$. Then we obtain that $\psi(0) \le \psi_I(0)$ and $\psi(\infty) \ge \psi_I(\infty)$, where ψ_I is probability of ruin for the independent case. When $F_\Theta(\theta_0) > 0$, then we have $\psi(0) < \psi_I(0)$ and $\psi(\infty) > \psi_I(\infty)$.

We can also examine a situation when the interclaim times and the amount of claims are dependent. We assume that (W_i, X_i) are independent random vectors, W_i have exponential distribution $F_W(t) = 1 - e^{-\lambda t}$, but the dependent structure between W_i and X_i is describe by the Spearman copula C [Heilpern 2014]. Then the Laplace transform of probability of ruin satisfies the following relation:

$$\psi^*(s)=\frac{\psi(0)-\frac{1-\alpha}{s+\beta}b+\rho q\frac{q-b}{s-q}\psi^*(q)}{s-b+\rho q+b\beta\frac{1-\rho}{s+\beta}+\rho q\frac{q-b}{s-q}},$$

where $a = -c^2\beta^2 + 3c\beta\lambda - c\rho\beta\lambda - \lambda^2$, $b = 4c\beta\lambda(c\beta - \lambda)^2$ and $q = \lambda/a$.

When the claims X_i have the exponential distribution, we obtain the explicit formula on the probability of ruin:

$$\psi_\rho(u)=\psi_\rho(0)e^{s_1(\rho)u},$$

where $s_1(\rho)=\frac{-g(\rho)-\sqrt{g^2(\rho)+4c\beta\lambda a_1^2}}{2ca_1}$, $a_1 = (c\beta - \lambda)$, $g(\rho) = c^2\beta^2 - 3c\beta\lambda + c\beta\lambda\rho + \lambda^2$, $a_2 = c^2\beta^2 - \lambda^2$,

$$\psi_\rho(0)=\frac{2h(\rho)}{a_2+h(\rho)+\sqrt{4c\beta\lambda a_1^2+(h(\rho)-a_1^2)^2}}$$ and $h(\rho) = c\beta\lambda(1 - \rho)$.

If $\tau = 0$, the r.v. are independent, then we have the known formula $\psi_0(u)=\frac{\lambda}{c\beta}e^{-(\beta-\lambda/c)u}$ and for $\tau = 1$, strict dependence, $\psi_1(u) = 0$ for any u. If the degree of dependence α increases, then the probability of ruin $\psi_\rho(u)$ decreases.

Simulations

Knowledge of the copulas makes it easy to simulate the random marginal values of the multidimensional distributions. In the case of two r.v. X_1, X_2, the simulation is based on the conditional copula

$$G(t; u) = P(U_2 \leq t | U_1 = u) = \frac{\partial}{\partial u} C(u,t),$$

where U_1, U_2 is the r.v. uniformly marginal distributed on [0, 1] with joint c.d.f. C. The simulation procedure is as follows [Nelsen 1999; Romano 2002; Heilpern 2007]

i) Generate independent uniform on [0, 1] variates v_1, v_2.

ii) Set $u_1 = v_1$ and $u_2 = G^{-1}(v_2; u_1)$.

iii) Set $x_1 = F_1^{-1}(u_1)$, $x_2 = F_2^{-1}(u_2)$.

The points x_1, x_2 are the random value of the variables X_1, X_2.

For the Archimedean copula we have the explicit formula for $v_2 = \frac{(\varphi^{-1})'(\varphi(u_1) + \varphi(u_2))}{(\varphi^{-1})'(\varphi(u_1))}$ and the simulations for an elliptical copula are based on the Cholesky decomposition of the correlation matrix.

References

Denuit M., Dhaene J., Le Bailly de Tilleghem C., Teughem S., 2001, *Measuring the impact of a dependence among insured lifelengths*, Belgian Actuarial Bulletin, no. 1, pp. 18-39.

Heilpern S., 2007, *Funkcje łączące*, AE, Wrocław.

Heilpern S., 2010, *Wyznaczanie prawdopodobieństwa ruiny gdy struktura zależności wypłat opisana jest Archimedesową funkcją łączącą*, [w:] *Zagadnienia aktuarialne. Teoria i praktyka*, red. W. Otto, Wydawnictwo Uniwersytetu Warszawskiego, Warszawa.

Heilpern S., 2011, *Wyznaczanie wielkości renty w zależnych grupowych ubezpieczeniach na życie*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, vol. 230, pp. 30-48.

Heilpern S., 2014, *Ruin measures for a compound Poisson risk model with dependence based on the Spearman copula and the exponential claim sizes*, Insurance: Mathematics and Economics, vol. 59, pp. 251-257.

Nelsen R.B., 1999, *An Introduction to Copulas*, Springer, New York.

Rolski T., Schmidli H., Schmit V., Teugels J., 1999, *Stochastic Process Insurance and Finance*, Willey & Sons, Chichester.

Romano C., 2002, *Applying copula to risk management*, University of Rome, working paper.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

A PROOF OF A CONJECTURE IN BERNOULLI

Volker Mammitzsch (Philipps-Universität Marburg)

In their paper in Bernoulli [2016], M. Drton and H. Xiao [Drton, Xiao 2016] considered a certain random variable which appeared in the course of studying a large sample distribution of Wald statistics. They also gave a long list of references where the random variable occurs. Unfortunately, they did not succeed in finding the distribution of that variable. Instead, they stated a conjecture in a special case. We will prove that conjecture here.

To be precise, let $f : R^m \to R$ be continuously differentiable, homogeneous of degree $\lambda \neq 0$, and X an m-dimensional normal random vector with expectation 0 and covariance matrix Σ.

We will study the random variable

$$W_{f,\Sigma} = \frac{f(X)^2}{(\nabla f(X))' \Sigma \nabla f(X)} . \tag{1}$$

Here, ∇f means the gradient of f and a prime indicates transposition.

In [Drton, Xiao 2016], conjecture 1.1, Drton and Xiao stated the following.

Conjecture: Let Σ be any positive semidefinite $m \times m$ matrix with positive diagonal entries.

If $f(x_1,...,x_m) = x_1^{\alpha_1} \cdot ... \cdot x_m^{\alpha_m}$ with nonnegative integer exponents $\alpha_1,...,\alpha_m$ that are not all equal to zero, then for $W_{f,\Sigma}$ according (1) it holds

$$W_{f,\Sigma} \sim \frac{1}{(\alpha_1 + ... + \alpha_m)^2} \chi_1^2 .$$

Here, "~" means "distributed as".

Moreover, they proved, applying first passage times of a Brownian motion.

Theorem 1: If X has independent components, i.e. Σ is diagonal, then the Conjecture is true.

Here we will show

Theorem 2: The Conjecture is true if $X \sim N(0, \Sigma)$, Σ any positive semidefinite matrix with positive diagonal entries.

Proof: As f is homogeneous of degree $\lambda \neq 0$, it holds

$$X'\nabla f(x) = \lambda f(x) \text{ for all } x \in R^m, \tag{2}$$

see e.g. [Hazewinkel (ed.) 1989].

Remark 1: (2) is the well-known Euler formula. It stays true if f is defined on an open set $B \subset R^m$ only.

Remark 2: The differential equation (2) is also sufficient for f to be homogeneous of degree λ, c.f. [Hazewinkel (ed.) 1989].

Because Σ is symmetric it may be written as

$$\Sigma = U \operatorname{diag}(\lambda_1,\ldots,\lambda_m)U',$$

where U is an orthogonal matrix, i.e. $UU' = U'U = I_m$, I_m denoting the *m*-dimensional unit matrix and $\operatorname{diag}(\lambda_1,\ldots,\lambda_m)$ the matrix with diagonal entries $\lambda_1,\ldots,\lambda_m$ and zero otherwise, see e.g. [Richter, Mammitzsch 1973]. As Σ is nonnegative definite, all of the λ's are nonnegative, i.e. $\lambda_j = d_j^2$, $j = 1,\ldots,m.$ From this we conclude

$$\Sigma = WW' \text{ with } W = U \operatorname{diag}(d_1,\ldots,d_m). \tag{3}$$

As X is a normal random vector with $E(X) = 0$, $\operatorname{cov}(X) = \Sigma$, it may be written as

$$X = AY \tag{4}$$

where Y is a standard normal vector of dimension n and A a real $m \times n$ matrix. Without loss of generality we may assume $n = m$ and choose $A = W = U \operatorname{diag}(\lambda_1,\ldots,\lambda_n)$ as in (3).

In what follows, $f \circ U$ means the composition of the linear function $R^m \to R$ given by $x \to Ux$ for all $x \in R^m$ with f, thus

$$f \circ U(x) = f(Ux) \text{ for all } x \in R^m.$$

From (2) and (3) it follows after some algebraic calculation

$$\begin{aligned} W_{f,\Sigma} &= \frac{1}{\lambda^2} \cdot \frac{(\nabla f(X))' X X' \nabla f(X)}{(\nabla f(X))' \Sigma \nabla f(X)} \\ &= \frac{1}{\lambda^2} \cdot \frac{(\nabla (f \circ U)(Z))' Z Z' \nabla (f \circ U)(Z)}{(\nabla (f \circ U)(Z))' \operatorname{diag}(d_1^2,\ldots,d_m^2) \nabla (f \circ U)(Z)} \end{aligned}$$

ŚLĄSKI PRZEGLĄD STATYSTYCZNY
Nr 16(22)

where $Z = \text{diag}(d_1,...,d_n)Y$ is a normal vector with $E(Z) = 0$, $\text{cov(Z)} = \text{diag}(d_1,...,d_n)Y$.

Because $f \circ U$ is a continuously differentiable function, again homogeneous of degree λ, we come up with

$$W_{f,\Sigma} = W_{f \circ U, \text{diag}\left(d_1^2,...,d_m^2\right)}.$$

Moreover, as Z has independent components, by Theorem 1 we find

$$W_{f,\Sigma} = \frac{1}{(\alpha_1 + ... + \alpha_m)^2} \chi_1^2.$$

Remark 3: $W_{f,\Sigma}$ in general has not a multiple of χ_1^2 distribution.

For example, if $f(x_1,...,x_m) = x_1^2 + ... + x_m^2$, $m > 1$, X standard normal, then

$$W_{f,I} = \frac{\left(X_1^2 + ... + X_m^2\right)^2}{4X_1^2 + ... + 4X_m^2} \sim \frac{1}{4} \chi_m^2, \text{ see also [Drton, Xiao 2016].}$$

Remark 4: Generally, $W_{f,\Sigma}$ is stochastically smaller than $\frac{1}{\lambda^2} \chi_m^2$, because

$$W_{f,\Sigma} = \frac{1}{\lambda^2} \frac{\left(\nabla f\left(WY\right)'WY\right)^2}{\left|W'\nabla f\left(WY\right)\right|^2} \le \frac{1}{\lambda^2} \frac{\left|\nabla f\left(WY\right)'WY\right|^2 \cdot |Y|^2}{\left|\nabla f\left(WY\right)'WY\right|^2} = \frac{1}{\lambda^2} |Y|^2 \sim \frac{1}{\lambda^2} \chi_m^2.$$

References

Drton M., Xiao H., 2016, *Wald tests of singular hypotheses*, Bernoulli, 22 (1), pp. 38-59.
Hazewinkel M. (ed.), 1989, *Encyclopedia of Mathematics 4*, Kluwer.
Richter H., Mammitzsch V., 1973, *Methode der kleinsten Quadrate*, Verlag Berliner Union.

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY

Nr 16(22)

HOW HUGE DOES STORAGE NEED TO BE IN A FULLY RENEWABLE POWER SYSTEM

Mänz Posselt (Philipps-Universität Marburg)

While fossil fuels are limited resources, renewable energy is an almost infinite resource. Partially for this reason, our energy system transforms into a mostly regenerative supply system. Nevertheless there are some difficulties to deal with on the way to change. The biggest share of renewable energy sources are photovoltaic and wind power. Both have extensive stochastic behaviour. Moreover power demand is not constant. To satisfy the demand, storages have to buffer power supply. Looking at Germany the question arises how large does the storage in a fully renewable energy system has to be at the least? And furthermore, which of the possibilities minimizes costs? To answer these questions data from 2012 to 2016 on a daily basis is used. Therefore it is necessary to calculate the degree of capacity utilization, which is then fed with energy divided by the installed capacity. To model the installed capacity of photovoltaic monthly data is interpolated. The installed capacity of wind power is modeled via a regression model with a third order polynom on a yearly data base. Power demand (respectively total load) is normed to its mean, so that the arithmetic mean equals one. Secondly the "Power-to-Gas" technology is presumed for storage. This technology converts electricity into gas (hydrogen and/or methane), which is easily stored. Later electricity is produced out of this gas. Assuming the following model, where storage S_t at time t is define as

$$S_t = \begin{cases} \min(E_t \cdot \epsilon_{PtG} + S_{t-1}, S_0) \text{ if } E_t \geq 0 \\ S_{t-1} + E_t \cdot (\epsilon_{GtP})^{-1} \text{ if } E_t < 0 \end{cases}$$

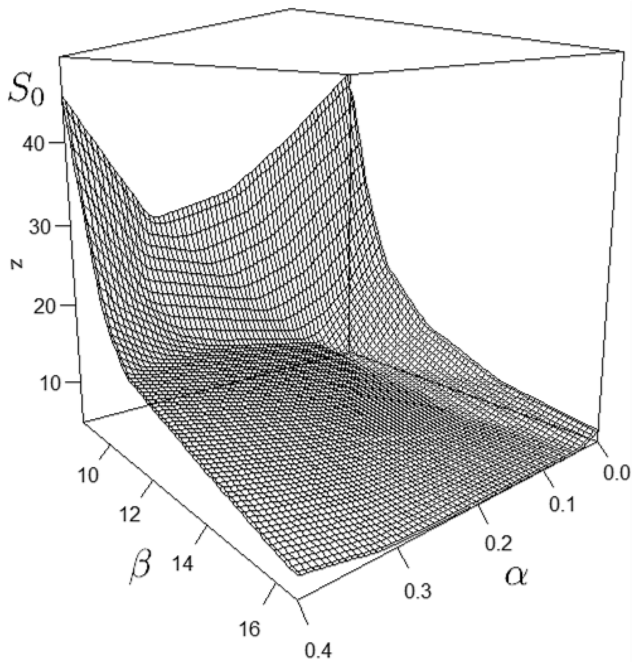
with

$$E_t = (\alpha \cdot p_t + (1-\alpha) \cdot w_t) \cdot \beta - l_t .$$

S_0 is the initialized storage respectively the maximal storage capacity. The parameter $\alpha \in [0, 1]$ describes the percentage of installed photovoltaic capacity and more over p_t is the degree of utilized photovoltaic capacity at time t. Further w_t characterizes the degree of utilized wind power capacity and l_t the power demand. Parameter β describes the surplus of capacity to cover the demanded load. The efficiency factor for converting power to gas is described by ϵ_{PtG} and from gas to power by ϵ_{GtP}. If there is an excess of supplied power (E_t is positive), the storage would be charged with $E_t \cdot \epsilon_{PtG}$. Otherwise the storage decreases about $E_t \cdot (\epsilon_{GtP})^{-1}$ to cover the demand. Therefore, the minimization problem is finding the smallest S_0 for the given time series $t = 1, \dots, T$ where $S_t \geq 0$. Therefore it is described as follows:

$$\min(S_0) \mid S_t \geq 0 \; \forall \; t = 1, \dots, T.$$

If one does so for a set of parameters α and β and assumes $\epsilon_{PtG} = \epsilon_{GtP} = 0.58$ one gets the following figure:



As can be seen there is a parabolic trend among α. This is based on two effects. First, with α decreasing, the mean degree of capacity utilization is slightly rising and tends to decrease the need for storage. Second, diversification occurred because of the negative empirical correlation between photovoltaic and wind power ($\hat{\rho} = -0.36$). Furthermore there is a decreasing need for storage if β increases. If one is now minimizing S_0 it is clear that β tends to go to infinity. Therefore it is more likely to involve costs. From the point of view of a benevolent planner it is clear to minimize the overall costs. If one does so, one gets the cost minimal point at around $\alpha = 0.424$ and

$\beta = 7.852$ with costs of about 889 billion euro and storage (in MWh) of about 99 days of average demand, which equals (assuming a heating value of 11 $\frac{kWh}{m^3}$ and an efficiency factor of 0.58) 23.6 billion cubic meters. Germany already stores about 23.8 billion cubic meters. Therefore it is already enough for a fully renewable electrical power system. For further details, like extended assumptions, cost structures and so forth, one will find it as part of the Ph. D. thesis of Posselt.

References

Data sources: Agora Energiewende, Bundesverband WindEnergie, Bundesnetzagentur.

Lehner M., Tichler R., Steinmueller H., Koppe M., 2014, *Power-to-Gas: Technology and Business Models*, 1st ed., Springer.

AN ARMA-GARCH MODEL FOR SPOT PRICES OF ELECTRICITY

Karl-Heinz Schild (Philipps-Universität Marburg)

The author suggests an ARMA-GARCH model for daily spot prices of electricity based on weekly difference of the spot price (so that in Box-Jenkins terminology it could be called an SARIMA-GARCH-model). The model can be interpreted as describing a reversion of the spot price to an adaptive target, obtained as an exponentially smoothed version of the real price series. The exponential smoothing depends on the MA-component only, showing that the MA-component captures the long-term dynamics of the price. The AR-component, in contrast, serves to capture the short-term dynamics, i.e. the reversion to the adaptive mean given by the exponentially smoothed price. Applying an ARCH, GARCH- or EGARCH-scheme to explain the variance of the ARMA-innovations, this variance can be directly interpreted as a measure for the uncertainty about tomorrow's spot price given its long and short term dynamics. The author defines the volatility of the spot price to be (the square root of) this quantity.

The author fits the model to the German EPEX spot price in the years 2009 to 2016 to assess the transmission of volatility from wind power feed-in to the German spot price. The results depend on what is understood to be the volatility of wind power feed-in. Measuring this volatility by the speed of absolute changes of wind feed-in, transmission elasticities of 0.2 to 0.3 are found quite robustly in all

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 16(22)

subperiods. (That is, a 1-percent-increase of this volatility is accompanied by a 0.25-percent increase of spot price volatility on average). Measuring the volatility of wind feed-in by the speed of relative changes (i.e. in the same way as it is usually done with financial price series), the estimated transmission elasticities are usually larger, but depend much more on the subperiod considered. On some subperiods we even find elasticities above one (i.e. a one-percent increase of the volatility of wind feed-in measured in relative terms can increase the spot price volatility by more than one percent).

The above ARMA-GARCH-model also provides a straightforward way to decouple spikes in the spot price from its diffusional behavior (using the long-term component to identify spikes). An interesting question (beyond the scope of this paper) would be whether the spikes can at least partially be attributed to extreme wind-power feed-in, either in terms of level or volatility.

COST SHARING MODEL BASED ON AN EXAMPLE OF ONCOLOGICAL TREATMENT IN POLAND

Beata Zmyślona (Wrocław University of Economics)

Health care is a socially sensitive area of a great importance for the economy in Poland and also in other countries. Financing health care has become increasingly difficult due to limited resources compared to the demand for medical services. The growth of demand for medical services is observed all over the world and is determined by at least four factors: demographic, cultural, income and technological. The demographic factor is associated with the phenomenon of the ageing of society which causes an increase of the incidence of various diseases and of the demand for geriatric care. The cultural factor determines the eating habits and lifestyle which affect the spread of civilization diseases. The income factor is related to the increase in wealth of society. All over the world, the phenomenon of increased health care expenditure caused by raised income is observed. In addition, these expenses are rising faster than income, which indicates that health is perceived as a luxury good. The technological factor is related to the development of modern medical technologies and knowledge which makes treatment more expensive.

Thus, the above mentioned factors point to the fact that health systems in modern economies are insolvent. Rationing of resources is a common practice, which adversely affects a patient's health. Many tools to deal with the deficit should be considered within the framework of social and health politics. The increase of a basic health premium size in the public system is one of them. The insufficiency of the public health insurance system entails the necessity of increasing the patient's contribution to the cost of care by introducing a copayment system.

The cost-sharing concept is based on an agreement between a patient and a provider of medical services, which defines the scope and financial framework of the participation of a medical service beneficiary in the expenses of the payer. Many studies and analyses are conducted on aggregated data sets which makes the estimation of an average cost depending on a patient's health stage, age or sex impossible. A precise valuation of a unit cost of treatment for a disease is crucial for defining cost-sharing rules.