

ECONOMETRIC MODEL OF THE CZECH LIFE INSURANCE MARKET

Radek Hendrych, Tomáš Cipra*

Abstract:

The aim of the article is to introduce a complex econometric model of cash-flows for the Czech life insurance market. Namely, technical-actuarial links among insurance variables observed in annually published summary balance sheets of life insurers are described by means of an econometric system of linear simultaneous equations. The suggested model is statistically verified and thus it can provide useful economic interpretations. Further, adjusted residual bootstrapping is introduced in this context as a straightforward alternative which can solve possible problems with questionable asymptotic distribution properties of residuals. This technique can be applied *e.g.* for significance testing purposes. Finally, an important practical illustration of scenario analysis is considered. Such an analysis might be really useful, *e.g.* for internal calculations of the Czech life insurers, financial planning or stress testing in the framework of Solvency II. Two general approaches are presented: deterministic and stochastic. The second one is capable of delivering various empirical probabilities concerning possible future developments.

Keywords: econometric model, econometric system of simultaneous equations, insurance market, life insurance, scenario analysis, residual bootstrap, Solvency II.

JEL Classification: C30, C32, C39, G22

1. Introduction

In this article, a complex econometric model of cash-flows for the Czech life insurance market is introduced. In particular, a dynamic econometric system of linear simultaneous equations is applied to capture different technical-actuarial links among observed insurance variables (*e.g.* the indemnity, the acquisition and administrative expenses, the insurance premium, *etc.*). These variables are compulsorily reported in annual balance sheets of life insurers operating in the Czech Republic and thus they can be simply used for this modelling aim. Generally, a statistically verified econometric model concerning this phenomenon is useful also from both the economic and the actuarial point of view. The issues relative to the discussed topic, *i.e.* to econometric modelling of cash-flows in life insurance, are handled in several academically or practically oriented works, *e.g.* by Schüller (1988), Cipra (1998) or Baranoff (2007).

* Radek Hendrych, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Charles University in Prague, Prague, Czech Republic (hendrych@karlin.mff.cuni.cz); Tomáš Cipra, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Charles University in Prague, Prague, Czech Republic (cipra@karlin.mff.cuni.cz). This work was supported by the Czech Science Foundation (the Grant No. GA P402/12/G097).

Section 2 introduces a general framework based on a simultaneous equation model (SEM) together with assumed insurance variables and their elementary statistical characteristics. Section 3 presents estimates of the proposed econometric model. Moreover, statistical validity is verified and economic interpretations are given. Section 4 shows capabilities of adjusted residual bootstrapping in the connection with the considered econometric model. This technique can solve possible inaccuracies caused by using of theoretical asymptotic distribution of residuals and thus it can be applied in a broad statistical context, *e.g.* for significance testing purposes. Section 5 deals with prognoses and scenario generations (optimistic, pessimistic or randomly generated anticipations), which should be taken into account by life insurance companies, *e.g.* in their internal calculations or for stress testing presented by Solvency II. Section 6 contains conclusions.

2. Model Framework

An approach, which is applied here for econometric modelling of cash-flows for the Czech life insurance market, is based on the dynamic econometric system of linear simultaneous equations, *i.e.* a special case of the general class of multivariate econometric models. Such a system allows capturing simultaneously more than one dependent variable and thus one can model analysed economic phenomena more precisely. Thus, it has an extensive range of practical applications, *e.g.* predictions or scenario analysis of possible future developments.

Suppose a dynamic econometric system of simultaneous equations in a general structural form

$$\mathbf{y}_t^T \boldsymbol{\Gamma} + \sum_{\ell=1}^p \mathbf{y}_{t-\ell}^T \boldsymbol{\Phi}_\ell + \mathbf{x}_t^T \mathbf{B} + \boldsymbol{\varepsilon}_t^T = 0, \quad t = p+1, \dots, T, p \in \mathbb{N} \quad (1)$$

where \mathbf{y}_t denotes the vector of m endogenous variables, $\mathbf{y}_{t-\ell}$ is the vector of m lagged endogenous variables, *i.e.* the predetermined variables which are uncorrelated with current and future values of residual components, and \mathbf{x}_t is the vector of k strictly exogenous variables (uncorrelated with all residuals components) with first element normalized to 1 (everything for a given time t). The matrices $\boldsymbol{\Gamma} (m \times m)$, $\mathbf{B} (k \times m)$ and $\boldsymbol{\Phi}_\ell (m \times m)$ contain real parameters of the model and $\boldsymbol{\varepsilon}_t$ stands for the vector of m structural residuals (*i.e.* structural disturbances including error terms).

In the framework of the model (1), assume:

- (A1) $\mathbb{E}(\boldsymbol{\varepsilon}_t) = 0$, $\text{cov}(\boldsymbol{\varepsilon}_t) = \mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T) = \boldsymbol{\Sigma}$ for a given positive definite matrix ($m \times m$) and $\text{cov}(\boldsymbol{\varepsilon}_s, \boldsymbol{\varepsilon}_t) = \mathbb{E}(\boldsymbol{\varepsilon}_s \boldsymbol{\varepsilon}_t^T) = 0$, $s \neq t$,
- (A2) $\mathbb{E}(\mathbf{x}_t \boldsymbol{\varepsilon}_t^T) = \mathbb{E}(\mathbf{x}_t) \mathbb{E}(\boldsymbol{\varepsilon}_t^T) = 0$,
- (A3) $\mathbb{E}(\mathbf{x}_t \mathbf{x}_t^T) = \boldsymbol{Q}$, \boldsymbol{Q} is a finite matrix ($k \times k$),
- (A4) $\boldsymbol{\Gamma}$ is an invertible matrix,
- (A5) $\boldsymbol{\Gamma}$ has elements -1 on its diagonal.

The assumption (A5) is not a serious restriction. It can be viewed as a suitable scaling of equations that consists of allocating one variable on the right-hand side of the model (1). Multiplying the equation (1) from the right by $\boldsymbol{\Gamma}^{-1}$ transfers the model to the

so-called reduced form. This transformed formulation of SEM together with the assumptions (A1)-(A5) is simply a vector autoregressive process with added exogenous elements and residuals $v_t^T = -\varepsilon_t^T \Gamma^{-1}$ (Lütkepohl, 2005). Hence, the reduced form can be consistently estimated via the standard least squares method (unlike the structural version). On the other hand, it suffers from indistinct interpretations of modelling links among variables.

Generally, the structural version of the dynamic system of linear simultaneous equations (1) is in literature connected with two main specifics: (i) necessity of special estimation techniques (limited information *vs.* full information methods) and (ii) a problem of identification, *i.e.* how to (uniquely) determine relationships among parameters of the structural and reduced version of the system. See *e.g.* Dhrymes (1994) or Greene (2003) for deeper discussion given therein.

2.1 Econometric model of the Czech life insurance market

For simplicity, we consider only relationships arising from the annual technical report of life insurance (*i.e.* the compulsory part of the balance sheet). Links among these economic data can be investigated through econometric modelling based on actuarial theory (see *e.g.* Cipra, 2010). On the other hand, one could possibly extend the introduced model into a more general form including other insurance or economic variables.

In particular, we assume these life insurance variables: CS_t - the indemnity (including claims and surrenders of policies) in year t (in thousands of CZK), DV_t - the contribution to the technical reserves in year t (in thousands of CZK), EAA_t - the acquisition and administrative expenses in year t (in thousands of CZK), EX_t - the number of policies terminated in year t (in pieces), N_t - the number of the new life policies (the newcomers) in year t (in pieces), P_t - the insurance premium written in year t (in thousands of CZK), $PORT_t$ - the insurance portfolio (the number of active insurance policies) in year t (in pieces) defined by the identity $PORT_t = PORT_{t-1} + N_t - EX_t$, $PROF_t$ - the investment income in life insurance in year t (in thousands of CZK), $PROFS_t$ - the profit share in life insurance in year t (in thousands of CZK), V_t - the technical reserves in year t (in thousands of CZK) and R_t - the life insurance result in year t (in thousands of CZK), ($t = 1, \dots, 15$ ($t = 1$ refers to the year 1997, $T = 15$ to the year 2011)). The yearly based data from 1997 to 2011 were subtracted from the summary annual reports of the Czech Insurance Association (ČAP), *i.e.* the association of commercial insurance companies with 98% share in the total premium written in the Czech Republic.¹ In Table 1, basic statistical characteristics of the data are shown. The sample correlation matrix of the analysed data is computed in Table 2.

¹ http://www.cap.cz/default_en.aspx, 04/02/2013.

Table 1 | Basic Statistical Characteristics of the Analysed Data

	Mean	Median	Max	Min	Std. Dev.	Skew.	Kurt.	Jarque-Bera Stat.
CS	2.06E+7	1.89E+7	4.16E+7	8.18E+6	1.09E+7	0.591	2.052	1.436 (0.488)
DV	1.30E+7	1.39E+7	2.29E+7	4.80E+6	5.05E+6	0.127	2.355	0.301 (0.860)
EAA	9.36E+6	9.71E+6	1.53E+7	3.10E+6	3.88E+6	-0.125	1.885	0.816 (0.665)
EX	9.42E+5	7.61E+5	2.71E+6	2.35E+5	5.92E+5	1.832	6.357	15.436 (0.000)
N	9.83E+5	9.35E+5	1.59E+6	7.03E+5	2.60E+5	0.888	2.993	1.970 (0.373)
P	4.05E+7	4.20E+7	7.04E+7	1.23E+7	1.88E+7	0.053	1.894	0.771 (0.680)
PORT	6.87E+6	6.64E+6	8.76E+6	6.02E+6	8.62E+5	0.919	2.728	2.159 (0.340)
PROF	7.70E+6	5.87E+6	2.09E+7	-2.61E+6	5.76E+6	0.754	3.572	1.626 (0.443)
PROFS	1.27E+5	8.95E+4	6.89E+5	5.52E+3	1.64E+5	2.915	10.544	56.815 (0.000)
R	4.72E+6	4.25E+6	1.64E+7	-8.99E+5	4.55E+6	1.192	4.124	4.342 (0.114)
V	1.49E+8	1.39E+8	2.57E+8	6.38E+7	6.60E+7	0.234	1.689	1.212 (0.546)

Table 2 | The Sample Correlation Matrix of the Analysed Data

	CS	DV	EAA	EX	N	P	PORT	PROF	PROFS	R	V
CS	1.000										
DV	0.465	1.000									
EAA	0.948	0.622	1.000								
EX	0.405	0.176	0.404	1.000							
N	0.284	0.257	0.434	0.348	1.000						
P	0.958	0.654	0.991	0.372	0.404	1.000					
PORT	0.511	0.330	0.626	-0.145	0.259	0.599	1.000				
PROF	0.369	0.674	0.334	0.370	-0.229	0.385	-0.133	1.000			
PROFS	0.362	0.260	0.348	0.801	-0.058	0.320	0.043	0.602	1.000		
R	0.641	0.758	0.640	0.432	0.033	0.695	0.161	0.882	0.554	1.000	
V	0.964	0.599	0.978	0.358	0.290	0.986	0.625	0.372	0.344	0.678	1.000

Now, proceed to the formulation of the econometric model of simultaneous equations which describe the relationships among particular life insurance market variables. Suppose the following system:

$$CS_t = \beta_{11} + \beta_{21}EX_t + \beta_{31}N_t + \beta_{41}PORT_t + \beta_{61}PROF_{t-1} + \phi_{71}V_{t-1} + \varepsilon_t^{CS},$$

$$DV_t = \beta_{12} + \beta_{52}PROF_t + \gamma_{12}CS_t + \gamma_{32}EAA_t + \gamma_{42}P_t + \gamma_{52}PROFS_t + \varepsilon_t^{DV},$$

$$EAA_t = \beta_{13} + \beta_{33}N_t + \beta_{43}PORT_t + \gamma_{13}CS_t + \gamma_{43}P_t + \phi_{73}V_{t-1} + \varepsilon_t^{EAA},$$

$$\begin{aligned}
 P_t &= \beta_{14} + \beta_{24}EX_t + \beta_{34}N_t + \beta_{54}PROF_t + \phi_{74}V_{t-1} + \varepsilon_t^P, \\
 PROFS_t &= \beta_{15} + \beta_{25}EX_t + \beta_{35}N_t + \beta_{45}PORT_t + \beta_{55}PROF_t + \beta_{65}PROF_{t-1} + \varepsilon_t^{PROFS}, \\
 R_t &= \beta_{16} + \gamma_{96}TR_t - \gamma_{86}TC_t + \varepsilon_t^R, \\
 V_t &= DV_t + \beta_{17} + \phi_{77}V_{t-1} + \varepsilon_t^V, \\
 TC_t &= CS_t + DV_t + EAA_t + PROFS_t, \\
 TR_t &= P_t + PROF_t.
 \end{aligned}
 \tag{2}$$

To clarify the structure of the suggested model (2), one can take into consideration mainly these arguments: (i) it follows the typical econometric modelling framework (*i.e.* the dynamic system of linear simultaneous econometric equations), (ii) it includes all key variables, that are relevant from a general perspective of life insurance, and it captures various actuarial relationships among them, and (iii) achieved numerical results empirically confirm the adequate form of (2), see further.

The considered econometric system (2) originally includes seven stochastic equations (*i.e.* the equations with stochastic residuals). The stochastic equation for the insurance result in year t , R_t , is expressed as the difference between a part of the total revenues $TR_t = P_t + PROF_t$ in year t and a part of the total costs $TC_t = CS_t + DV_t + EAA_t + PROFS_t$ in year t in the presence of the intercept and the residual terms. Parts (apparently minority) of the total revenues and the total costs are transferred outside of the technical account of life insurers. The last stochastic equation describing evolution of the technical reserves in year t , V_t , respects the fact that the lagged technical reserves can be also adjusted, *e.g.* by the technical interest rate.

In the suggested model, the intercepts and variables $EX_t, N_t, PORT_t, PROF_t$, (and thus also $PROF_{t-1}$) are assumed to be strictly exogenous, *i.e.* these variables enter the system from outside. This choice seems to be reasonable due to the apparent external character of such variables. The lagged variable V_{t-1} is supposed to be predetermined, *i.e.* it is fully determined by the given system in time $t - 1$. To sum up, the dynamic system of linear simultaneous equations (2) follows the considered structural form (1) with $p = 1$, $k = 6$ and $m = 9$. It has 9 endogenous and 7 exogenous (6 strictly and 1 predetermined) variables. Note that the indices of the model parameters β , γ and ϕ refer to the corresponding components of the matrices \mathbf{B} , $\mathbf{\Gamma}$ and $\mathbf{\Phi}_1$, respectively. Moreover, each equation in the system (2) satisfies the necessary (order) condition of identification, see Greene (2003). Using the order condition, one can easily see that all equations in (2) are overidentified (*i.e.* the number of exogenous variables excluded from the given equation is greater than the number of endogenous variables included on the right-hand side of the equation). This is consistent with general econometric practice that most of simultaneous equations are overidentified. Furthermore, one can statistically test overidentifying restrictions, *e.g.* by Sargan tests (see further).

3. Model Estimation and Interpretation

To estimate the unknown parameters of the previous stochastic simultaneous econometric equations, the three-stage least squares method (3SLS) is considered (Greene, 2003).

This full information estimation technique exploits all information available in the system. In general, it is a special case of the more universal full information generalized method of moments (GMM). Thus, it offers a set of suitable properties (under general assumptions). Namely, such estimates are consistent, asymptotically normally distributed and asymptotically efficient (Dhrymes, 1994). From the application point of view, the 3SLS procedure is easily accessible in the econometric and statistical software, *e.g.* EViews or R, and therefore it can be handled in a comfortable way.

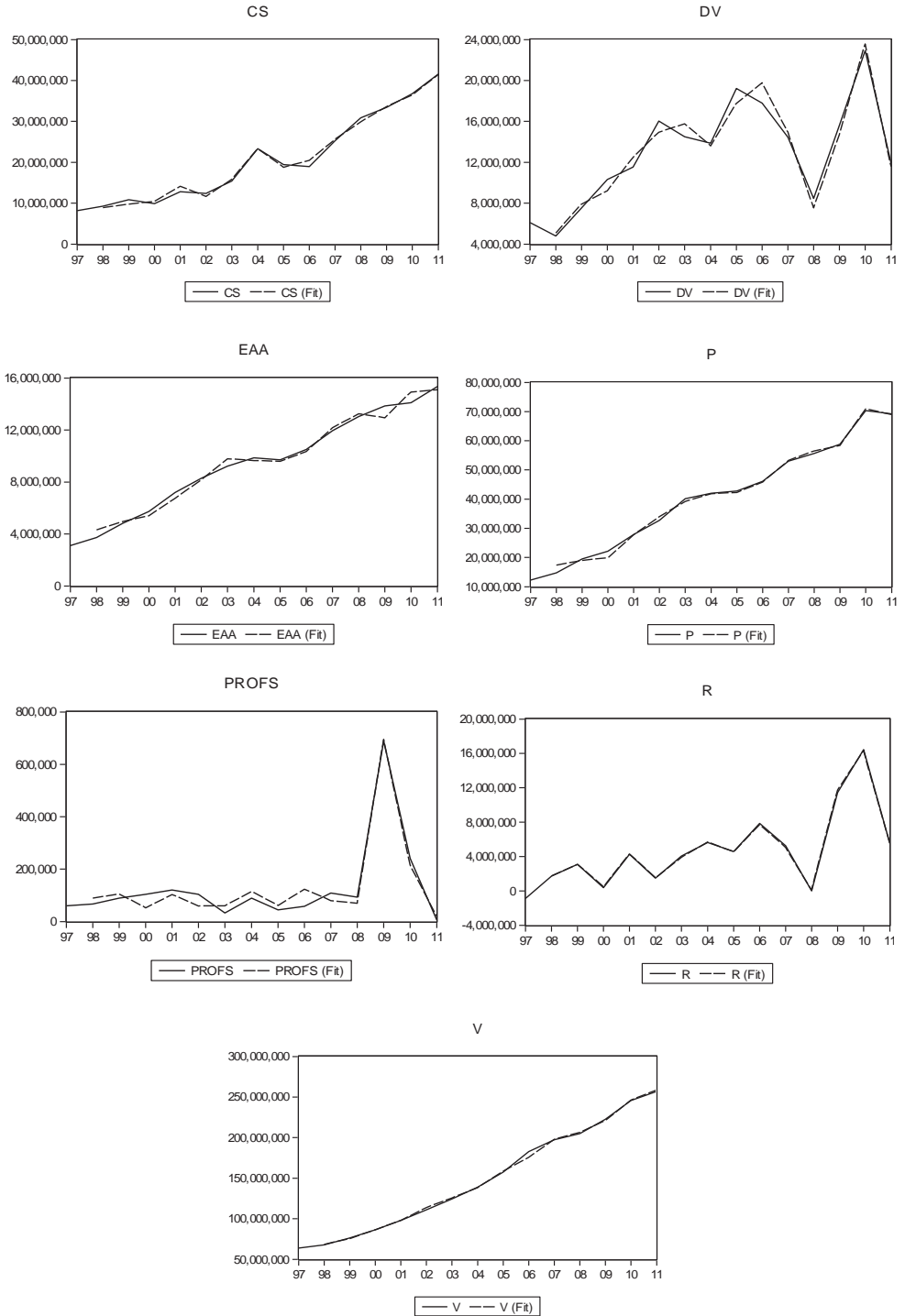
The 3SLS estimates of the system with their estimated standard errors delivered by EViews 7.0 are

$$\begin{aligned}
 CS_t &= 39773733.25 - 7.877EX_t + 4.206N_t - 7.153PORT_t - 0.791PROF_{t-1} + 0.289V_{t-1}, \\
 &\quad (3138486.46) (0.744) (0.800) (0.518) (0.068) (0.010) \\
 DV_t &= 1424964.85 + 0.577PROF_t - 0.987CS_t + 0.430EAA_t + 0.596P_t - 6.844PROFS_t, \\
 &\quad (3378588.65) (0.173) (0.068) (2.253) (0.454) (5.389) \\
 EAA_t &= -958930.77 + 1.475N_t + 0.193PORT_t + 0.015CS_t + 0.138P_t + 0.012V_{t-1}, \\
 &\quad (980630.79) (0.534) (0.158) (0.044) (0.040) (0.014) \\
 P_t &= -11608264.44 - 2.871EX_t + 15.020N_t + 0.551PROF_t + 0.265V_{t-1}, \\
 &\quad (1423929.55) (0.546) (1.1363) (0.060) (0.005) \\
 PROFS_t &= -274833.67 + 0.204EX_t - 0.213N_t + 0.060PORT_t + 0.010PROF_t - 0.009PROF_{t-1}, \\
 &\quad (73991.63) (0.017) (0.035) (0.009) (0.002) (0.001) \\
 R_t &= 73383.50 + 0.978TR_t - 0.987TC_t, \\
 &\quad (111342.22) (0.012) (0.015) \\
 V_t &= DV_t - 683878.04 + 1.007V_{t-1}, \\
 &\quad (1546512.47) (0.010) \\
 TC_t &= CS_t + DV_t + EAA_t + PROFS_t, \\
 TR_t &= P_t + PROF_t.
 \end{aligned} \tag{3}$$

Moreover, one can verify significance of the included parameters using the *z*-statistics (since the 3SLS procedure gives asymptotically normally distributed estimates, see above). In particular, the *z*-statistics is simply a ratio between the estimated parameter and its estimated standard deviation. For instance, considering the equation for V_t in (3), one obtains the following *z*-statistics: -0.442 and 99.859 (with the relevant 5% critical value 1.96), which strongly confirms the significance of the parameter estimate $\hat{\phi}_{77}$. On the other hand, it might be questionable to rely on the asymptotical normality with regard to the relatively low number of observations. This can strongly motivate residual bootstrapping, see Section 4.

Figure 1 shows the observed values of the endogenous variables acting in the stochastic equations in the model (2) with their fitted counterparts. One can see that the model fits the data adequately.

Figure 1 | The Observed Endogenous Variables with their Fitted Counterparts



Let us proceed to statistical verification of the considered model (2) which is mainly based on the calculated multivariate residuals. Firstly, Table 3 presents the sample correlation matrix of the estimated residuals. The several relatively high correlations point to the appropriateness of the introduced modelling framework of the econometric simultaneous equations, see the assumption (A1).

Table 3 | The Sample Correlation Matrix of the Estimated Residuals of the Model (2)

	CS	DV	EAA	P	PROFS	R	V
CS	1.000						
DV	0.600	1.000					
EAA	-0.314	0.360	1.000				
P	-0.356	0.058	0.433	1.000			
PROFS	0.224	0.493	0.001	0.021	1.000		
R	-0.204	-0.609	-0.666	-0.059	-0.156	1.000	
V	-0.596	-0.565	0.238	0.215	-0.630	0.134	1.000

Secondly, the Portmanteau test (the Ljung-Box version) for autocorrelations of the residuals gives $Q(2) = 106.246$ with p-value 0.267 and $Q(4) = 214.568$ with p-value, *i.e.* the null hypothesis of no residual autocorrelations up to lag 2 or 4 cannot be rejected in both cases (Greene, 2003). Thirdly, the joint Jarque-Bera statistics based on the Cholesky decomposition, which verifies the multivariate normality of the residuals, attains the value 4.955 with achieved p-value 0.986, *i.e.* the null hypothesis of multivariate normality cannot be rejected (Lütkepohl, 2005). Fourthly, the Hausman test of the model specification gives the statistics $W = 19.982$ with 34 degrees of freedom and achieved p-value 0.973 so that the null hypothesis cannot be rejected. In particular, this test verifies that all exogenous variables of the system (strictly exogenous and predetermined ones) are uncorrelated at any time with the residuals (Greene, 2003), compare with the assumption (A2). Finally, the Sargan statistics for testing overidentifying restrictions are calculated, see Table 4. This statistical criterion tests that in each equation all exogenous variables and the residual term satisfy the orthogonality condition. A large value of the test statistics is obviously taken as evidence that there are inappropriately omitted exogenous variables in the given equation (Greene, 2003). Using the obtained p-values (see Table 4), one cannot reject the correct specification of each equation in the system (2). To conclude, the suggested econometric system (2) with its estimation (3) seems to be correctly specified from the statistical point of view.

Table 4 | The Sargan Statistics Based on the 2SLS Estimates of the Model (2)

Equation	Sargan stat.	df	p-value
CS	0.179	1	0.673
DV	0.044	1	0.834
EAA	2.403	1	0.121
P	0.102	2	0.950
PROFS	0.020	1	0.887
R	9.375	4	0.052
V	5.474	4	0.242

To interpret roles of the selected parameters in the model (2), one can conveniently use the reduced form of the considered system, *i.e.* the special case of (1) with $p = 1$ multiplied from the right by the matrix Γ^{-1} :

$$\mathbf{y}_t^T = -\mathbf{y}_{t-1}^T \Phi_1 \Gamma^{-1} - \mathbf{x}_t^T \mathbf{B} \Gamma^{-1} - \boldsymbol{\varepsilon}_t^T \Gamma^{-1}, \quad t = 2, \dots, T. \quad (4)$$

Put $\Delta_1 = -\Phi_1 \Gamma^{-1}$, $\Pi = -\mathbf{B} \Gamma^{-1}$ and $\mathbf{v}_t^T = -\boldsymbol{\varepsilon}_t^T \Gamma^{-1}$ to simplify the notation, *i.e.*

$$\mathbf{y}_t^T = \mathbf{y}_{t-1}^T \Delta_1 + \mathbf{x}_t^T \Pi + \mathbf{v}_t^T, \quad t = 2, \dots, T \quad (5)$$

By backward substitution, one obtains

$$\mathbf{y}_t^T = \mathbf{y}_1^T \Delta_1^{t-1} + \sum_{s=0}^{t-2} \mathbf{x}_{t-s}^T \Pi \Delta_1^s + \sum_{s=0}^{t-2} \mathbf{v}_{t-s}^T \Delta_1^s, \quad t = 2, \dots, T. \quad (6)$$

Thus, it can be simply regarded how the variables on the right-hand side of the equation (6) contribute to change of the current values of the endogenous variables in \mathbf{y} . Generally, two different types of effects can be distinguished: (i) a short-run effect, (ii) a long-run effect. The short-run effect of change of any explanatory variable is defined by means of the matrix Π . This contains information concerning the immediate change due to the fact that

$$\frac{\partial y_{jt}}{\partial x_{it}} = (\Pi)_{ij}. \quad (7)$$

The long-run effect of change of any explanatory variable is defined in terms of the following infinite sum:

$$\sum_{s=0}^{\infty} \Pi \Delta_1^s \quad (8)$$

with regard to the fact that it holds

$$\frac{\partial y_{jt}}{\partial x_{i,t-s}} = (\Pi \Delta_1^s)_{ij}. \quad (9)$$

If $\lim_{s \rightarrow \infty} \Delta_1^s = \mathbf{0}$ by elements (or equivalently, if all eigenvalues of Δ_1 lie inside the unit circle) then the sum (8) can be simplified as

$$\sum_{s=0}^{\infty} \Pi \Delta_1^s = \Pi [\mathbf{I}_m + \Delta_1 + \Delta_1^2 + \dots] = \Pi [\mathbf{I}_m - \Delta_1]^{-1} \quad (10)$$

In the case of our econometric system (2) and its estimation (3), the only nonzero eigenvalue of the matrix Δ_1 is 0.903 (it lies obviously inside the unit circle), *i.e.* the estimated system is stable. Thus, both the short-run and long-run effects can be computed (see the results in Table 5 and Table 6). Hence, various interpretations can be considered. For instance, the short and long-run effects of the equation for *PROFS* are similar due to strictly exogenous character of all explanatory variables therein. If the value of *PROF* is increased by 1, *i.e.* 1,000 CZK, *ceteris paribus*, the expected change of *PROFS* is + 0.01, *i.e.* extra 10 CZK in the profit share. Further, if one assumes one new additional insurance contract in *N* (and consequently in *PORT*), *ceteris paribus*, the short-run effect for the explained variable *V* is simply 7.810 + 6.385, *i.e.* extra 14,495 CZK in the technical reserves. From the long-run view, it means 80.771 + 69.143, *i.e.* extra 14,9914 CZK in the technical reserves. A lot of similar conclusions can be made. Finally, note that there are evident relationships between the short-run and long-run effects in the cases of TC_t , TR_t and R_t , *i.e.* the effects for *R* are simply given by the following identity: $0.978 \times \text{effect for } TR - 0.987 \times \text{effect for } TC$ (compare with the estimated model (3)).

Table 5 | The Matrix $\hat{\Pi}$, *i.e.* the Short-run Effects

	CS	DV	EAA	P	PROFS	R	V	TC	TR
1	3.98E+7	-4.37E+7	-1.95E+6	-1.16E+7	-2.75E+5	-5.20E+6	-4.44E+7	-6.16E+6	-1.16E+7
EX	-7.876	4.445	-0.517	-2.872	0.204	0.887	4.445	-3.744	-2.872
N	4.206	7.810	3.612	15.020	-0.213	-0.526	7.810	15.415	15.020
PORT	-7.153	6.685	0.084	0.000	0.060	0.320	6.685	-0.324	0.000
PROF	0.000	0.868	0.076	0.551	0.010	0.575	0.868	0.954	1.551
PROF₋₁	-0.791	0.836	-0.012	0.000	-0.009	-0.024	0.836	0.025	0.000

Table 6 | The Matrix $\hat{\Pi}[\mathbf{I}_m - \hat{\Delta}_1]^{-1}$, *i.e.* the Long-run Effects

	CS	DV	EAA	P	PROFS	R	V	TC	TR
1	-9.27E+7	3.90E+6	-2.65E+7	-1.33E+8	-2.75E+5	-1.63E+7	-4.59E+8	-1.16E+8	-1.33E+8
EX	5.388	-0.322	1.940	9.318	0.204	1.998	45.972	7.209	9.318
N	27.511	-0.566	7.928	36.437	-0.213	1.425	80.771	34.660	36.437
PORT	12.797	-0.485	3.779	18.334	0.060	1.990	69.143	16.150	18.334
PROF	2.591	-0.063	0.556	2.932	0.010	0.791	8.980	3.094	3.932
PROF₋₁	1.705	-0.061	0.450	2.293	-0.009	0.185	8.649	2.085	2.293

4. Residual Bootstrapping

If distribution of residuals of a considered econometric system of simultaneous equations is unknown, the so-called bootstrap methods may be applied to investigate the distributions of functions of multiple time series or stochastic processes. See *e.g.* Freedman (1984) or Faïre (2004) for more details and the references given therein.

In this context, a residual based bootstrap is frequently used to find an estimator of a quantity of our interest, say $q: = q(\mathbf{\Gamma}, \mathbf{B}, \mathbf{\Phi}_1, \dots, \mathbf{\Phi}_p)$. Freedman (1984) presents several additive (stronger) assumptions: (i) $\{(\mathbf{x}_t, \boldsymbol{\varepsilon}_t)\}$ are i.i.d. random vectors with components having finite fourth moments, (ii) $\{(\mathbf{y}_t)\}_t$ is a weakly stationary stochastic process, (iii) the parameter matrices $\Delta_\ell = -\mathbf{\Phi}_\ell \mathbf{\Gamma}^{-1}$ are required to be stable, *i.e.* the eigenvalues of Δ_ℓ lie inside the unit circle. Compare with Faïre (2004).

Suppose that the values of strictly exogenous variables $\mathbf{x}_{p+1}, \dots, \mathbf{x}_T$ and endogenous variables $\mathbf{y}_1, \dots, \mathbf{y}_T$ are given. The residual bootstrap algorithm for an econometric system of simultaneous equations, *i.e.* an adaptation of a general bootstrap technique (Lütkepohl, 2005), proceeds as follows:

1. The residuals of the model are estimated as

$$\hat{\boldsymbol{\varepsilon}}_t^T = -\mathbf{y}_t^T \hat{\mathbf{\Gamma}} - \sum_{\ell=1}^p \mathbf{y}_{t-\ell}^T \hat{\mathbf{\Phi}}_\ell - \mathbf{x}_t^T \hat{\mathbf{B}}, \quad t = p+1, \dots, T,$$

where $\hat{\mathbf{\Gamma}}$, $\hat{\mathbf{B}}$ and $\hat{\mathbf{\Phi}}_1, \dots, \hat{\mathbf{\Phi}}_\ell$ are consistent estimates of parameter matrices (*e.g.* by the discussed 3SLS method). The calculated residuals are centred as $\hat{\boldsymbol{\varepsilon}}_{p+1} - \bar{\boldsymbol{\varepsilon}}, \dots, \hat{\boldsymbol{\varepsilon}}_T - \bar{\boldsymbol{\varepsilon}}$

where $\bar{\boldsymbol{\varepsilon}} = \frac{1}{T} \sum_{t=p+1}^T \hat{\boldsymbol{\varepsilon}}_t$.

2. The bootstrap residuals $\boldsymbol{\varepsilon}_{p+1}^*, \dots, \boldsymbol{\varepsilon}_T^*$ are obtained by the random drawing with replacement from the sequence of centred residuals constructed in Step 1. The bootstrap time series are computed recursively as

$$\mathbf{y}_t^{*T} = -\sum_{\ell=1}^p \mathbf{y}_{t-\ell}^{*T} \hat{\mathbf{\Phi}}_\ell \hat{\mathbf{\Gamma}}^{-1} - \mathbf{x}_t^T \hat{\mathbf{B}} \hat{\mathbf{\Gamma}}^{-1} - \hat{\boldsymbol{\varepsilon}}_t^{*T} \hat{\mathbf{\Gamma}}^{-1}, \quad t = p+1, \dots, T,$$

with $\mathbf{y}_1^* = \mathbf{y}_1, \dots, \mathbf{y}_p^* = \mathbf{y}_p$.

3. Based on the bootstrap time series, the parameter matrices $\mathbf{\Gamma}$, \mathbf{B} , and $\mathbf{\Phi}_1, \dots, \mathbf{\Phi}_p$ are consistently reestimated and the bootstrap version of the statistic \hat{q} *i.e.* \hat{q}^* , is calculated.
4. Step 2 and 3 of this algorithm are repeated N times, where $N \in \mathbb{N}$ is a large number.
5. In previous steps of this procedure, the values $\hat{q}_1^*, \dots, \hat{q}_N^*$ are obtained. Define $\bar{\hat{q}}_1^*$, the bootstrap estimator of q , simply as the mean value of $\hat{q}_1^*, \dots, \hat{q}_N^*$.

In this framework, one can simply construct confidence intervals suitable for testing or predicting purposes. In particular, assume the studentised statistics and its bootstrap counterpart as

$$R = \frac{q - \hat{q}}{\sqrt{\widehat{\text{var}}(\hat{q})}} \quad \text{and} \quad R^* = \frac{\hat{q} - \hat{q}^*}{\sqrt{\widehat{\text{var}}(\hat{q}^*)}}.$$

Note that the variances are estimated using common statistical techniques.

If H is the distribution function of the statistics R and γ_p is its p -quantile, the $(1 - \alpha)\%$ confidence interval for q , $\alpha \in (0,1)$, is defined as follows

$$CI_\alpha = \left(\hat{q} - \gamma_{1-\alpha/2} \sqrt{\widehat{\text{var}}(\hat{q})}, \hat{q} - \gamma_{\alpha/2} \sqrt{\widehat{\text{var}}(\hat{q})} \right).$$

Similarly, the $(1 - \alpha)\%$ studentised bootstrap confidence interval for q , $\alpha \in (0,1)$, is given by

$$CI_\alpha^* = \left(\hat{q} - \gamma_{1-\alpha/2}^* \sqrt{\widehat{\text{var}}(\hat{q})}, \hat{q} - \gamma_{\alpha/2}^* \sqrt{\widehat{\text{var}}(\hat{q})} \right),$$

where γ_p^* is the p -quantile of the distribution function H^* of the statistics R^* .

Note that $\sqrt{T}(\hat{q} - q)$ and $\sqrt{T}(\hat{q}^* - \hat{q})$ converge as $T \rightarrow \infty$ to the same limit distribution under suitable conditions. Therefore CI_α^* has the correct size asymptotically, *i.e.* $\mathbb{P}(q \in CI_\alpha^*) \rightarrow (1 - \alpha)$ as $T \rightarrow \infty$, see *e.g.* Lütkepohl (2005).

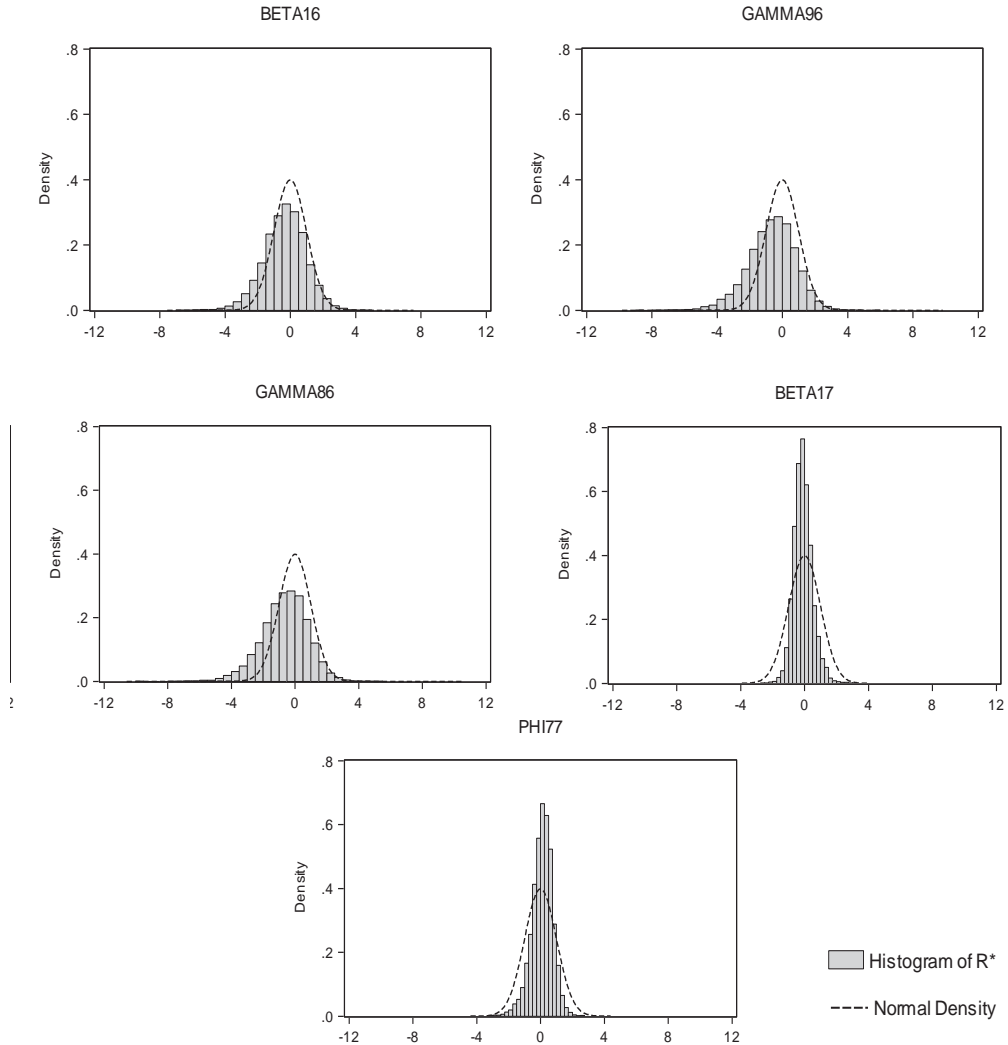
As mentioned previously, one can compare the introduced approaches with the construction of other confidence intervals, *e.g.* for significance testing purposes. Namely, for one model of cash-flows in the Czech life insurance market, only the relatively short data range is available. For this reason, one cannot fully rely on the classical approximations given by the normal distribution (which follow from the asymptotic properties of the 3SLS estimation technique). The bootstrap confidence intervals should be thus regarded as more reasonable.

Figure 2 compares the empirical distribution of the selected representative studentised statistics $\{R_i^*\}_{i=1}^N$ with the $N(0,1)$ distribution. It is clear that the normal approximation is not fully precise and that bootstrapping could give a better idea about distribution of the studentised statistics. Generally, the bootstrap intervals are usually asymmetric and in most cases wider than the normal ones. On the other hand, they can be more precise with respect to the theoretical results and previous discussion. Furthermore, one can compare the differences between the 3SLS 95% asymptotic normal and the 3SLS bootstrap confidence intervals for the selected unknown parameters in the model (2), *e.g.*

$$\begin{aligned} \beta_{16} : CI_{0.05} &= (-144843.25, 291610.24), CI_{0.05}^* = (-160358.10, 409830.49), \\ \gamma_{96} : CI_{0.05} &= (0.9547, 1.0017), CI_{0.05}^* = (0.9545, 1.0236), \\ \gamma_{86} : CI_{0.05} &= (0.9585, 1.0160), CI_{0.05}^* = (0.9590, 1.0428), \\ \beta_{17} : CI_{0.05} &= (-3714986.79, 2347230.71), CI_{0.05}^* = (-2544758.06, 1108155.60), \\ \phi_{77} : CI_{0.05} &= (0.9872, 1.0268), CI_{0.05}^* = (0.9940, 1.0213) \end{aligned} \tag{11}$$

Despite these differences, conclusions concerning significance of the parameters are *de facto* similar. Both methods cannot reject zero values of the following parameters: β_{12} , γ_{32} , γ_{42} , γ_{52} , β_{13} , β_{43} , γ_{13} , ϕ_{73} , β_{16} and β_{17} (using 5% significance level). Moreover, the bootstrap method moreover does not identify β_{33} as significant (using 5% significance level). Note that all practical computational solutions have been achieved by an own programme procedure in the software EViews (Version 7.0) with the choice $N = 10,000$.

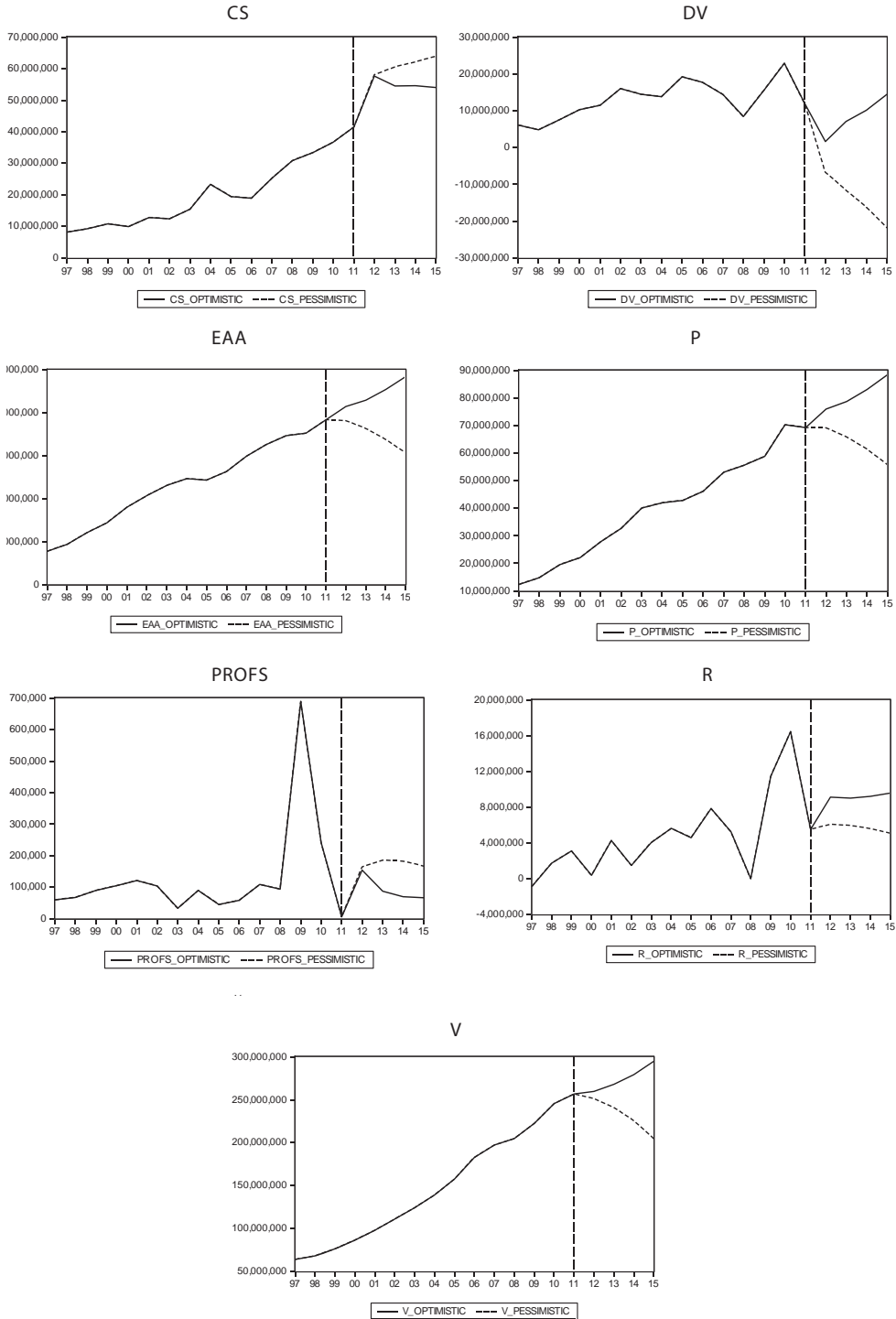
Figure 2 | The Histograms of $\{R_i^*\}_{i=1}^{10000}$ for β_{16} , γ_{96} , γ_{86} , β_{17} and ϕ_{77}



5. Scenario Analysis

From the practical point of view, anticipations of possible future developments of the considered (endogenous) insurance variables can be really useful, *e.g.* for purposes of internal insurers' calculations, stress testing or financial planning. Moreover, the suggested methodology could find a wide range of applications in future common practice of life insurance, especially in connection with planned Solvency II. In particular, Solvency II is a new regulatory schedule that can make use of insurer's internal models including stress tests, see *e.g.* Cipra (2010) or Sandström (2006). In the considered context, one can apply different general approaches to scenario analysis: (i) analysis of deterministic scenarios simply based on given expert (optimistic or pessimistic) values of strictly exogenous variables, (ii) analysis of random scenarios generated via some econometric models or (iii) their combination.

Figure 3 | The Results of Optimistic/Pessimistic Scenario Prognoses



5.1 Deterministic scenarios

To illustrate the main idea of deterministic scenario analysis, two different scenarios are supposed. First one is rather pessimistic: the number of new insurance policies N_t decreases by 10% each year, the number of terminated policies EX_t increases by 10% each year, the investment income $PROF_t$ remains on its first sample quartile, $t = 16, \dots, 19$, *i.e.* from 2012 to 2015. The second scenario follows optimistic expectations in the Czech life insurance market: the number of new insurance policies N_t increases by 10% each year, the number of terminated policies EX_t decreases by 10% each year, the investment income $PROF_t$ constantly retains its third sample quartile value.

The results of this analysis are summarized in Figure 3. One can see that the differences between two proposed scenarios are really significant. On the other hand, they respect logical expectations, *e.g.* for the pessimistic scenario the technical interest reserves decrease, the insurance premium decreases, the indemnity increases or the acquisition and administrative expenses decrease mainly due to the smaller insurance portfolio and low investment returns.

5.2 Randomly generated scenarios

Analysis of randomly generated scenarios might be more sophisticated in the sense that possibly uses more realistic scenario predictors of strictly exogenous variables, which respect evolution of the data generating process, *e.g.* by (linear) statistical time series models as ARMA and ARIMA (or their more complex multivariate counterparts). In particular, given original exogenous data can be fitted by such a model. Moreover, future anticipated values of strictly exogenous variables are obtained via stochastic predicting when predictions follow the estimated model with impact of randomly generated error innovations.

Namely, in the case of the econometric system (2), the exogenous variables EX_t , N_t and $PROF_t$ are captured by univariate linear ARMA models with normal residuals (the high values of the initially nonnegative integer variables EX_t and N_t allow using normal approximations with adjustments to integers). All models have been fully statistically verified. Essentially, they offer suitable fits. Finally, one obtains different random prognoses for all strictly exogenous variables. See Figure 4 with 100,000 generated scenarios. The obtained values have been used to predict all endogenous variables of the estimated econometric system (3). Figure 5 graphically presents the achieved results. From the economic point of view, the results offer several particular interpretations. For instance, in the case of scenarios for R_t , one can conclude that the sum of predicted returns over the prediction horizon is negative in 18.75% of scenarios, minimally one negative return in the four-year prediction horizon is in 38.84% of scenarios, only negative returns are in 0% of generated scenarios, new maximum is achieved in 8.77% and new minimum in 34.46% of scenarios.

Figure 4 | 100,000 Generated Scenarios of Exogenous Variables

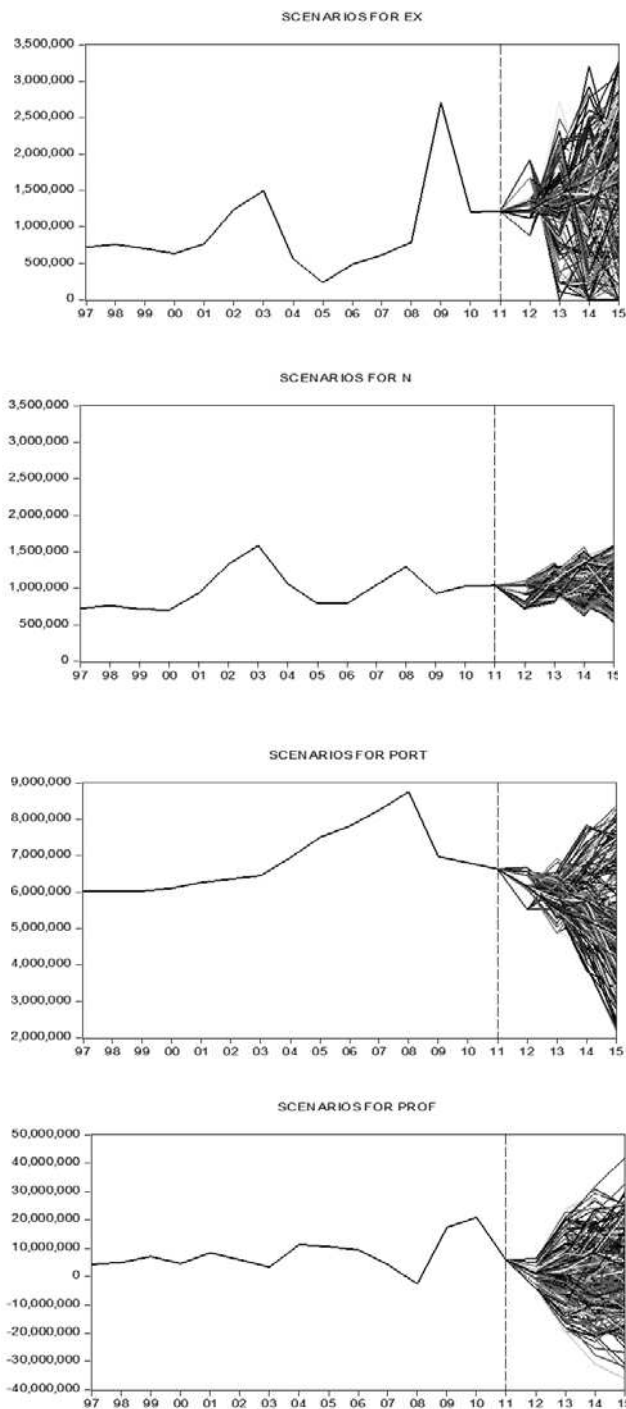
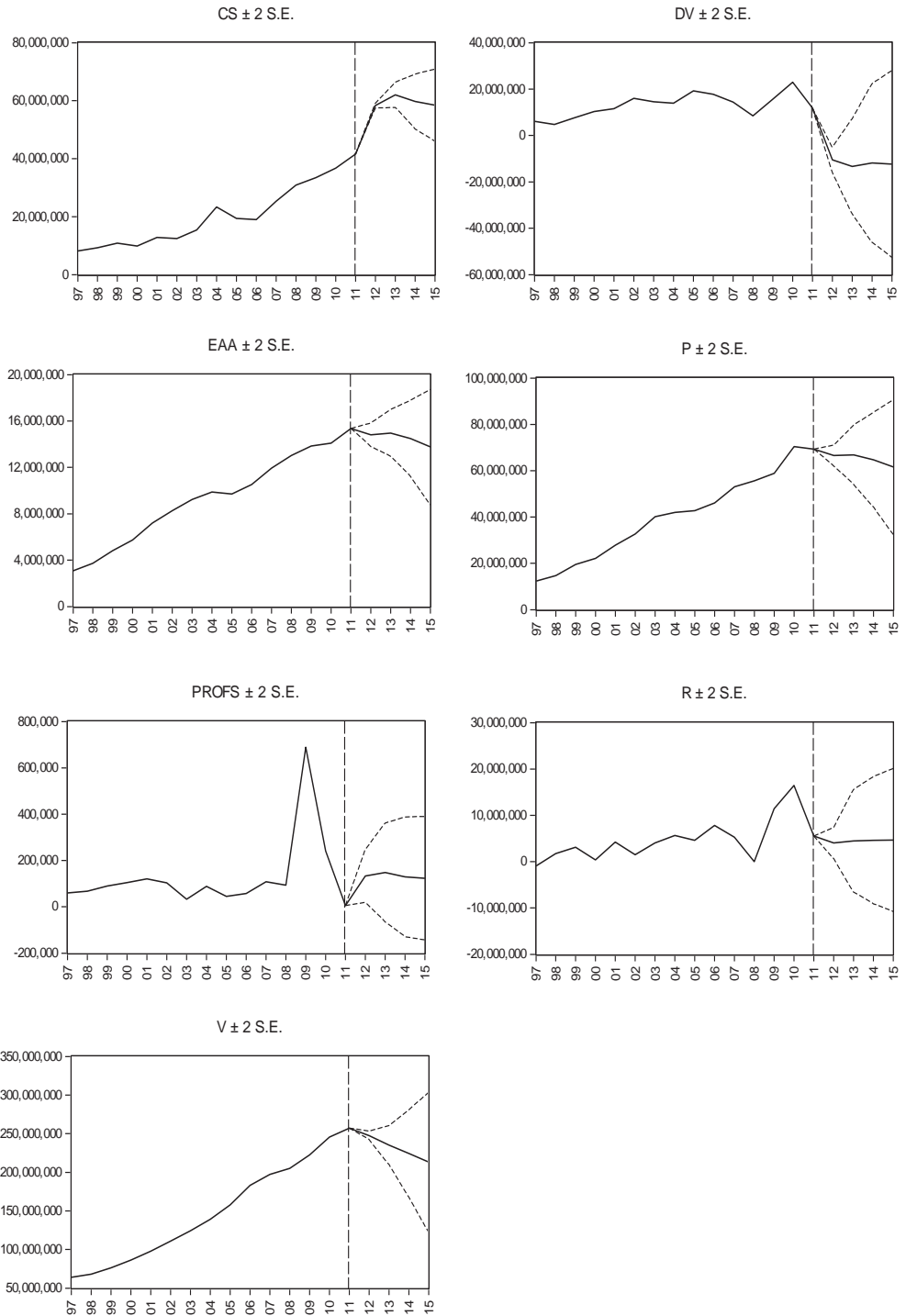


Figure 5 | The Mean and ± 2 s.d. Bounds of 100,000 Generated Scenario Predictions



6. Conclusion

The article deals with the complex econometric model of cash-flows for the Czech life insurance market. In particular, the system of linear simultaneous equations was considered to capture economic-actuarial relationships within the annually published summary technical reports of the Czech life insurers. Namely, the whole designed model was effectively determined by behaviour of three (exogenous) insurance variables: the number of new insurance policies, the number of terminated policies and the investment income of the life insurers (which is generally produced by various financial operations and allocations with financial assets). It was shown that the three-stage least squares estimation of the introduced model is valid from the statistical point of view which was verified by several criteria, *e.g.* the Hausman test or the Sargan test. Hence, it provides reasonable and useful interpretations also from the economic point of view, *e.g.* both the short and the long-run effect due to the increased investment income by 1,000 CZK (*ceteris paribus*) simply mean extra 10 CZK in the profit share, *etc.*

Further, a more sophisticated significance testing of model parameters was suggested. The adjusted residual bootstrapping is generally able to work with unknown distribution of residuals. In the framework of the proposed econometric model, this technique seems to be reasonable due to the relatively short range of the data which leads to doubts on normal approximations based on the asymptotic properties of the 3SLS estimation theory. The final differences in significance of the parameters were relatively small (with only one exception, namely the parameter β_{33}). On the other hand, it is clear that the theoretically achieved asymptotical normality is not fully reliable here, especially in comparison with the bootstrap counterpart.

Finally, scenario analysis offered effective anticipations of possible future developments which is useful from both the practical and the actuarial point of view. Two approaches were considered: the deterministic and the stochastic one. The first one could capture different expectations, *e.g.* optimistic and pessimistic scenarios. The obtained results respected natural expectations, *e.g.* in the sense of the increasing/decreasing technical reserves, the corresponding changes in the insurance premium, *etc.* The second one included a broader simulation framework and thus enabled to test deeply future possibilities, *e.g.* to calculate various empirical probabilities of specific phenomena such as the probabilities of newly achieved maximum and minimum in a prediction horizon (*e.g.* for the total return R it was 8.77% and 34.46 in the given four-year horizon, respectively).

References

- Baranoff, E. G., Papadopoulos, S., Sager, T. W.** (2007), "Capital and Risk Revisited: A Structural Equation Model Approach for Life Insurers." *Journal of Risk and Insurance*, Vol. 74, No. 3, pp. 653–681.
- Cipra, T.** (1998), "Econometric Analysis of Cash-flows in a Life Insurance Company." *Pojistné rozpravy*, Vol. 1998, No. 3, pp. 66–72.
- Cipra, T.** (2010), *Financial and Insurance Formulas*. New York: Physica-Verlag/Springer.
- Dhrymes, P. J.** (1994), *Topics in Advanced Econometrics: Linear and Nonlinear Simultaneous Equations*. New York: Springer.
- Faire, R. C.** (2004), *Estimating how the Macroeconomy Works*. Cambridge (MA): Harvard University Press.
- Freedman, D. A.** (1984), "On Bootstrapping Two-Stage Least-Squares Estimates in Stationary Linear Models." *Annals of Statistics*, Vol. 12, No. 3, pp. 827–842.
- Greene, W. H.** (2003), *Econometric Analysis*. New Jersey: Prentice Hall.
- Lütkepohl, H.** (2005), *New Introduction to Multiple Time Series Analysis*. New York: Springer.
- Sandström, A.** (2006), *Solvency: Models, Assessment and Regulation*. New York: Chapman & Hall/CRC.
- Schüler, W., Hüls, H.** (1988), "Ein Einfaches Modell des Lebensversicherungsunternehmens und seine Anwendung zu Prognose- und Kontrollzwecken." Paper presented at the 23rd International Congress of Actuaries Helsinki.