AR&P sciendo Health Economics and Management Review, Issue 2, 2023
ISSN 2786-4634 (print) ISSN 2786-4626 (online)

THE USA MEDICAL INSURANCE AS A STIMULATING FACTOR TO INCREASE LABOUR EFFICIENCY

Vitaliia Koibichuk, ORCID: http://orcid.org/ 0000-0002-3540-7922
Ph.D., Associate professor, Sumy State University, Ukraine
Bhola Khan, ORCID: https://orcid.org/0000-0002-4479-2762
Professor, MJP Rohilkhand University, India
Serhii Drozd, ORCID: http://orcid.org/ 0000-0002-0716-3078
Sumy State University, Ukraine
Corresponding author: Vitaliia Koibichuk, v.koibichuk@biem.sumdu.edu.ua
Type of manuscript: research paper

Abstract: *Medical insurance is critical for state labour efficiency. In many countries (including in the United States of America), it is tightly connected to labour, which makes workers have valid insurance policies for free and constant access to medical aid. That strongly secures workers' health and their high performance. In state-supporting insurance cases, citizens have a common access to medical services (regardless of their employment type). Here, people can be provided with medical aid without worrying about any prices, which keeps their strong health and high productivity skills. Within employment-related medical insurance, it is employers who are fully responsible for their employees' insurance. As a tangible financial business burden, it may keep workers close to their employment place itself: if resigned, they can lose good medical insurance at all. The medical insurance system is a key and decisive factor to raise labour efficiency. To achieve and secure it, governments should permanently develop affordable and reliable insurance systems. In our research, we chose the following indexes: coverage of state and private insurances, labour efficiency levels, national employment levels, life expectancy, healthcare costs (% of gross domestic product), healthcare costs by volume. We conducted the given study via data normalisation and regression modelling (backward data selection). We applied Multivariate Adaptive Regression Splines (MARS) as a regression-based method to describe non-linear variable relations. Among our engaged methods, there were also bibliography analysis, data processing, systematisation, comparison and logical generalisation. The current research results are relevant for politics and business. Politicians may use them in developing social-economic principles to improve medical insurance and labour efficiency. Enterprises can involve such information to define medical insurance payments for the health and labour efficiency increase among all types of employees in any countries.*

Keywords: insurance; health; national employment; life expectancy; MARS; regression.

JEL Classification: G22; I10; I12

Received: 25 April 2023 **Accepted:** 9 June 2023 **Published:** 30 June 2023

Funding: There is no funding for this research

Publisher: Academic Research and Publishing UG, Germany
Founder: Academic Research and Publishing UG, Germany; Sumy State University, Ukraine

Cite as: Koibichuk, V., Khan, B., & Drozd, S. (2023). The USA medical insurance as a stimulating factor to increase labour efficiency. *Health Economics and Management Review*, 2, 40-50. https://doi.org/10.21272/hem.2023.2-04

Introduction. Medical insurance and labour efficiency are tightly interconnected. With an access to medical insurance and disease prevention, employees have more chances to remain healthy and productive. Besides, insurance decreases nervous tension and makes employees feel more confident of their work. That is also important for high performance.

Having reviewed different sources, we defined some main medical insurance systems which are used worldwide. They cover medical aid costs: hospitalisation, doctor's consultations, prescribed drugs, etc (Smirnova et al., 2020).

1. The private medical insurance. Separate persons or households buy it from private insurance companies.
2. The social medical insurance. Governments or healthcare authorities provide citizens with it. Such a tax-funded insurance is available for everybody.
3. The employment medical insurance. Employers provide their employees with it.
4. The state medical insurance. Country authorities provide a tax-funded medical aid for everybody.
5. The high-risk medical insurance. You can get such a government-supported aid if private insurance companies refuse to pay because of previous diseases.

Moreover, there are several medical insurance models: The Bismarck and Beveridge models, the national insurance model, the direct medical service payment model.

The Bismarck model is based on paying medical services via a tax-filled state budget. Additionally, contributions are taken from salaries of employers and employees. This model is valid in Germany, Austria, Belgium, France, the Netherlands, Japan, Switzerland.

The advantages are decentralised insurance management and funding, high quality of medical aid, branch competition, many selectable insurance companies and medical services.

The disadvantages are increasingly non-affordable medical prices for some people. Besides, you must be employed and pay contributions to obtain such an insurance (Nguyen, 2017).

The Beveridge model is also based on paying medical services via a tax-filled state budget. Most medical institutions are state-owned; each citizen may get access to medical aid. There are no patient's bills to pay for provided medical services. This model is valid in the United Kingdom, Ireland, Sweden, New Zealand, Cuba, Greece, Portugal.

The advantages are affordable prices. Service quality is controlled by state authorities.

The disadvantages are centralised insurance management and funding, lack of stimuli to raise medical aid efficiency, few innovative approaches to treatment (Physicians for a National Health Program, 2010).

The national insurance model is based on private hospitals healthcare, which is funded via citizens contributing to state insurance programs. This model is valid in Canada, South Korea, Taiwan.

The advantages are affordable prices for most people. Service quality is controlled by state authorities.

The disadvantages are centralised healthcare management. Also, medical services may be provided for patients not always in time (Ajay,2021).

The direct medical service payment model is based on citizens themselves paying for medical aid. Usually, this model is valid in most developing countries.

The advantages are in-time aid provision and branch competition, quick selection of medical services and their high quality.

The disadvantages are non-affordable prices for needy people, low quality control of medical services (Nguyen, 2016).

There are a lot of factors that can influence labour efficiency. Among such variables, we can single out employees' qualification, workplace quality, advanced training, access to modern technologies and new equipment. Besides, relevant factors are staff management styles; standard or non-standard labour; employees' morality, motivation and satisfaction. To make a strong performance, you should satisfy employees' needs, predict all production processes, keep environment clean, observe occupational safety, regulate salaries properly.

Literature Review. Among sources, Ho (2015) reviews functionality of medical insurances in different countries. Erlangga et al. (2019) assess influence of different medical insurances on aid provision. Sommers (2017) explains how medical insurance can affect patients with certain consequences. Rochet (1991) proves that social insurance is an effective and reasonable tool for modern people. Dizioli and Pinheiro (2016) developed a model of insurance influencing labour efficiency. Ben Halima et al. (2014) analyse salary difference for human health. Shen et al. (2017) defined that the state medical insurance policy can affect labour force on the market.

Within the Scopus base, we found 1,525 sources (1963-2023) via the queries «health insurance» and «productivity». The Bibliometrix and RStudio software (Aria and Cuccurullo 2017) defined top 20 keywords, countries and marking words in headlines of labour efficiency papers (Figure 1).

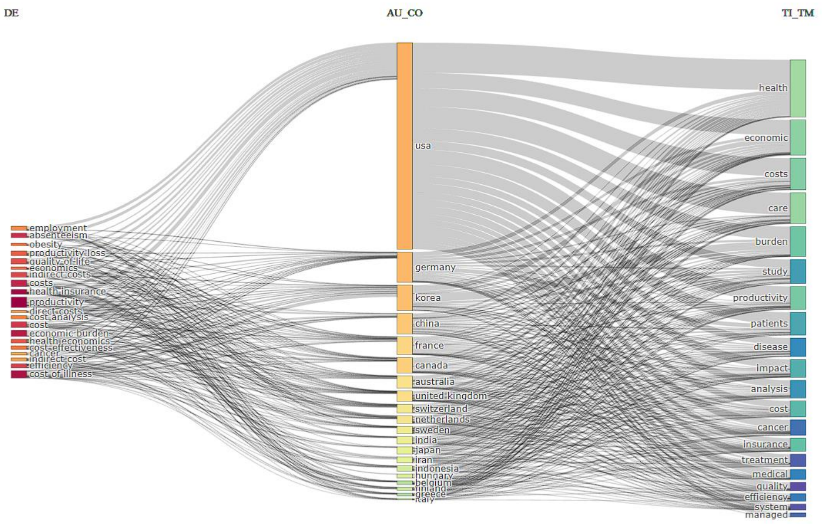

Figure 1. Interrelation within «keyword (DE) – country (AU_CO) – headline (TI_TM)»

Sources: developed by the authors based on the RStudio and Bibliometrix software.

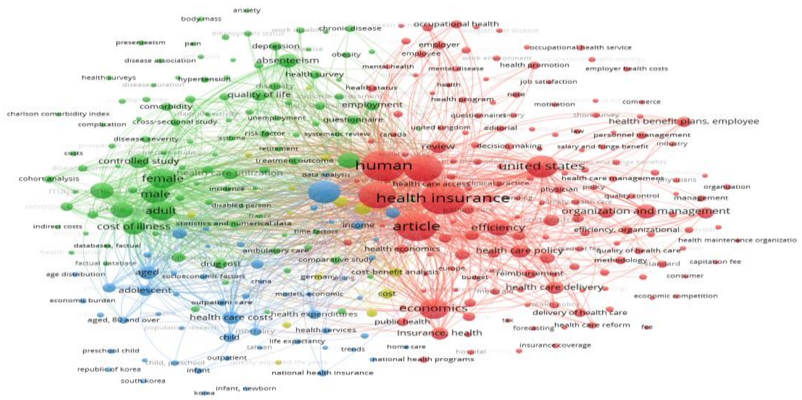

Figure 2. Keyword link map

Sources: developed by the authors based on the VOSviewer software.

The most used keywords within all 1,525 Scopus articles were: human, medical insurance, article, people, labour efficiency, medical service costs, female, adult, male, the USA, economics. detailed clinical research, middle age, disease costs, priority register, elderly, control study, absenteeism, healthcare costs.

Methodology and research methods. The informational research base is the 1987-2021 USA statistics for state and private insurance (USA Facts, 2023), labour productivity (OECD, 2023), national employment (USA Bureau of Labour Statistics, 2023), life expectancy (Macrotrends, 2023), healthcare costs within gross domestic product (USA Facts, 2023), healthcare costs by volume (USA Facts, 2023).

To produce all calculations, we should define normalized values of input data. It is necessary for adequate information processing in analyses and forecasts. Consequently, you can easily understand and properly index the arranged data. The obtained results may be used for further analysis.

The data normalisation is conducted via Formula (1):

$$K = \frac{1}{(1+e^{-3\frac{(x_i - md)}{(mx - md)}})} \quad (1)$$

where K – normalised value of input variables; x_i – input index value (i= ‾(1,…,35)); m_d – input index median; m_x – maximal input index value.

To confirm the model statistical quality, we realised the Statgraphics descriptive analysis. This application defines main numeric characteristics, regularities and conclusions.

To develop the regression model, we involve the Multiple Regression Backward Selection (MRBS) as an iteration algorithm to select variables from forecast sets within the regression model. It defines how labour efficiency is affected by state and private insurances, national employment, life expectancy, healthcare costs within gross domestic product and medical expenditures.

You start from all model predictors. Gradually, the highest p-variable is removed till you reach the pre-defined or minimal values for other predictors. This model is often used to assess relative significance of predictors by analysis of their influence on the common model. Finally, you get a simplified and accurate model with less variables.

In statistics, the MRBS is applied to produce a regression model via gradually added and subtracted predictors till you find a statistical reason for the model use or refusal. The process seeks for the model with all predictor variables that have a statistically significant relation to the solution variable itself.

The MRBS defines the most significant variables within the regression model. It is engaged to increase the regression model efficiency, decrease sample size and simplify its analysis. The Backward Stepwise Selection (BSS) is realised via separate steps. Firstly, you define an assessing criterion for model coefficients, select predictors and target variables. Secondly, you examine correlations to establish how dense ties are between predictors. Thirdly, you select those predictors whose correlation with target variables is the densest. Fourthly, you produce a regression model through the selected predictors, check statistical quality and significance (Effroymson, 1960).

The model statistical quality is checked via the Fisher and Student criteria, p-value, R2 and MAE (mean absolute error).

The next research stage is creation of multivariate adaptive regression splines by the most significant indexes. They affect the resulting variables (labour efficiency) after selection procedures.

The Multivariate Adaptive Regression Splines (MARS) is a statistical technique to analyse relations between dependent and independent variables that can be linear and non-linear.

The MARS is based on regression models where spline functions describe non-linear relations between variables. These are piecewise linear functions: there are junctions in break points with various slopes and shifts. The MARS technique uses splines to produce basis functions that can describe non-linear data.

The MARS algorithm looks for optimal break points for each variable and their optimal combinations as well. They may also reflect non-linear relations between variables. Via basis functions and data adaptation, the MARS generates very flexible models where intervariable relations are explained accurately.

An important MARS advantage is its applicability for absent data values and intervariable relations, which is difficult to trace through conventional linear models. Moreover, the MARS provides an automatic variable selection and reduces model sizes. Therefore, you do not need for retraining.

The MARS is a powerful machine method to generate non-linear regression models. It is reasonable for many predicted variables and their interrelations in data sets. The MARS piecewise linear functions reflect common data sections. They combine to produce a model with fixed data variations. In contrast to the simple linear regression, this technique gives more accurate results (Friedman, 1991).

The MARS model is defined as a weighted sum of basis functions $B_i(x)$:

$$\hat{f}(x) = \sum_{i=1}^{k} c_i B_i(x) \tag{2}$$

where c_i – stable coefficient; k – amount of basis functions.

The basis function (hinge function) is applied for machine learning, support vector machines, classifiers. Hinge losses measure the stock between classifier predictions and actual feature values. If the stock is low, penalties are imposed. The hinge function domain is max $(0, x - constant)$ or max $(0, constant - x)$. Thus, the MARS model automatically selects the hinge function shape, variables and their values. Also, it may define interaction between two or more variables as a product of hinge functions.

Therefore, the MRBS presupposes generalised check of basis function overload via the Generalised Cross-Validation (GCV) criterion. To select the best model subset, the following rule is observed: lower GCV values mean better results. In other words, GCV is a regulating method that includes contrast between model simplicity and efficiency (Craven and Wahba, 1978):

$$GCV = \frac{RSS}{(N \cdot (1 - ENP)/N)^2)} \tag{3}$$

where RSS – residual sum of squares (sum of difference squares between actual and planned model variables); N – observation amount.

Effective Number of Parameters (ENP) is measured as (Friedman, 1991):

$$ENP = (NMT) + (penalty) \cdot ((NMT) - 1)/2 \quad (4)$$

where NMT – number of MARS terms; $penalty$ is 2 or 4; $(NMT - 1)/2$ is number of hinge function knots.

The GCV criterion (3) adjusts the RSS considering the model flexibility. Therefore, the imposed flexibility penalty is necessary. Too flexible models produce data noise rather than their systematic structure (Bottegal and Pillonetto, 2018).

Results. Statistical input data are indicated in Table 1.

Table 1. Input data for analysing and modelling

Year	K1	K2	K3	K4	K5	K6
2021	298 686 000.00	1.0938	152586.25	78.99	0.1830	4255127
2020	297 680 000.00	1.0636	147812.75	78.93	0.1970	4144077
2019	298 438 000.00	1.0324	157534.3333	78.87	0.1760	3757382
2018	296 206 000.00	1.0209	155763	78.81	0.1761	3604428
2017	296 890 000.00	1.0073	153334.5833	78.84	0.1773	3446454
2016	292 320 000.00	0.9994	151436.4167	78.86	0.1774	3307404
2015	289 903 000.00	1.0000	148844.6667	78.89	0.1742	3165394
2014	283 200 000.00	0.9940	146318.6667	78.91	0.1716	3002623
2013	271 606 000.00	0.9893	143940.6667	78.94	0.1704	2856626
2012	263 165 000.00	0.9848	142474.5833	78.79	0.1718	2783252
2011	260 214 000.00	0.9785	139885.1667	78.64	0.1722	2676540
2010	256 603 000.00	0.9728	139077.1667	78.49	0.1727	2589637
2009	255 295 000.00	0.9408	139893.9167	78.34	0.1725	2492739
2008	256 702 415.60	0.9253	145373.25	78.19	0.1633	2402354
2007	255 017 524.00	0.9177	146050.1667	77.99	0.1595	2305514
2006	251 609 644.70	0.9077	144417.5833	77.79	0.1568	2165093
2005	250 799 438.40	0.8988	141710.0833	77.58	0.1557	2026567
2004	249 414 000.00	0.8823	139239.75	77.38	0.1555	1894672
2003	246 331 727.20	0.8586	137729.25	77.18	0.1545	1770371
2002	246 157 492.10	0.8357	136480.9167	77.04	0.1491	1631013
2001	244 058 640.00	0.8124	136939.3333	76.9	0.1401	1483415
2000	242 931 546.10	0.8036	136900.6667	76.75	0.1332	1365999
1999	239 101 984.00	0.7853	133500.9167	76.61	0.1322	1273213
1998	227 462 000.00	0.7646	131475.9167	76.47	0.1323	1198443
1997	225 646 000.00	0.7484	129572.3333	76.31	0.1321	1132947
1996	225 077 000.00	0.7335	126720.1667	76.14	0.1330	1073558
1995	223 733 000.00	0.7204	124908.25	75.98	0.1336	1020280
1994	222 387 000.00	0.7178	123071.1667	75.81	0.1326	966369
1993	220 040 000.00	0.7107	120258.6667	75.65	0.1334	914871
1992	218 189 000.00	0.7052	118487.9167	75.5	0.1307	852199
1991	216 003 000.00	0.6815	117712.5833	75.35	0.1277	785965
1990	214 167 000.00	0.6769	118795.6667	75.19	0.1205	718730
1989	212 807 000.00	0.6732	117327	75.04	0.1138	642177
1988	211 005 000.00	0.6648	114974.0833	74.89	0.1101	576649
1987	210 161 000.00	0.6580	112439.3333	74.79	0.1060	514473

Note: K1 – coverage of state and private insurances; K2 – labour efficiency levels K3 – national employment levels; K4 – life expectancy in years; K5 – healthcare costs as a GDP percentage; K6 – healthcare costs by volume in dollars.

Sources: developed by the authors based on (USA Facts, 2023a,b,c; OECD, 2023; USA Bureau of Labour Statistics, 2023; Macrotrends, 2023).

Since each variable is measured differently, we should convert input indexes into the comparable format to keep modelling. Here, the normalisation procedure is necessary according to Formula (1). The data normalisation results are given in Table 2.

Table 2. Normalised research indexes

Year	K1	K2	K3	K4	K5	K6
2021	0.9526	0.9526	0.8999	0.9526	0.8795	0.9526
2020	0.9497	0.9290	0.8053	0.9473	0.9526	0.9458
2019	0.9519	0.8937	0.9526	0.9414	0.8147	0.9143
2018	0.9453	0.8771	0.9377	0.9349	0.8158	0.8978
2017	0.9474	0.8548	0.9103	0.9382	0.8290	0.8778
2016	0.9317	0.8403	0.8817	0.9404	0.8291	0.8576
2015	0.9217	0.8415	0.8303	0.9434	0.7947	0.8341
2014	0.8867	0.8298	0.7644	0.9454	0.7623	0.8035
2013	0.7943	0.8202	0.6879	0.9482	0.7452	0.7725
2012	0.6979	0.8105	0.6346	0.9326	0.7648	0.7557
2011	0.6587	0.7963	0.5328	0.9128	0.7696	0.7298
2010	0.6077	0.7831	0.5000	0.8878	0.7765	0.7075
2009	0.5886	0.6962	0.5331	0.8568	0.7738	0.6814
2008	0.6092	0.6477	0.7356	0.8190	0.6369	0.6559
2007	0.5845	0.6228	0.7565	0.7571	0.5715	0.6277
2006	0.5334	0.5890	0.7043	0.6822	0.5233	0.5851
2005	0.5211	0.5582	0.6054	0.5921	0.5028	0.5418
2004	0.5000	0.5000	0.5066	0.5000	0.5000	0.5000
2003	0.4532	0.4165	0.4454	0.4079	0.4808	0.4606
2002	0.4506	0.3404	0.3960	0.3467	0.3859	0.4170
2001	0.4192	0.2704	0.4140	0.2902	0.2474	0.3722
2000	0.4026	0.2467	0.4125	0.2361	0.1660	0.3381
1999	0.3480	0.2016	0.2877	0.1924	0.1566	0.3122
1998	0.2081	0.1584	0.2252	0.1550	0.1570	0.2922
1997	0.1904	0.1300	0.1758	0.1199	0.1556	0.2753
1996	0.1852	0.1080	0.1183	0.0903	0.1643	0.2605
1995	0.1731	0.0914	0.0909	0.0686	0.1699	0.2476
1994	0.1617	0.0884	0.0690	0.0509	0.1597	0.2351
1993	0.1433	0.0805	0.0448	0.0383	0.1682	0.2235
1992	0.1300	0.0750	0.0340	0.0292	0.1427	0.2100
1991	0.1157	0.0547	0.0301	0.0223	0.1176	0.1964
1990	0.1047	0.0514	0.0357	0.0166	0.0738	0.1832
1989	0.0972	0.0489	0.0283	0.0126	0.0467	0.1691
1988	0.0880	0.0437	0.0195	0.0096	0.0362	0.1577
1987	0.0839	0.0398	0.0130	0.0080	0.0270	0.1475

Sources: developed by the authors.

Table 3 shows numerical characteristics of normalised values from the Statgraphics software. They confirm the statistical significance of features. Variance is over 5%. Standard Kurtosis and Standard Skewness range from -2 to 2.

Table 3. Descriptive analysis

Numerical characteristics/Index	K1	K2	K3	K4	K5	K6
Count	35	35	35	35	35	35
Average	0.4953	0.4654	0.4577	0.5008	0.4599	0.5183
Median	0.5	0.5	0.5	0.5	0.5	0.5
Geometric mean	0.3772	0.2965	0.2659	0.2402	0.3174	0.4390
Harmonic mean	0.2666	0.1593	0.0974	0.0609	0.1762	0.3645
5% Trimmed mean	0.4927	0.4624	0.4551	0.5031	0.4581	0.5146
5% Winsorised mean	0.4954	0.4648	0.4575	0.5007	0.4581	0.5184
Variance	0.0980	0.1155	0.1063	0.1556	0.0980	0.0762
Standard deviation	0.3131	0.3398	0.3260	0.3945	0.3131	0.2760
Variation coefficient	63.20%	73.02%	71.23%	78.78%	68.07%	53.25%
Gini coefficient	0.3649	0.4176	0.4134	0.4437	0.3879	0.3086

Continued Table 3

Numerical characteristics/Index	K1	K2	K3	K4	K5	K6
Standard error	0.0529	0.0574	0.0551	0.0667	0.0529	0.0466
Geometric standard deviation	2.2789	3.0442	3.7861	4.9266	2.7570	1.8479
5% Winsorized sigma	0.3321	0.3596	0.3451	0.4186	0.3286	0.2922
Mean absolute deviation	0.7147	0.9915	1.1261	1.3557	0.8818	0.5468
MAD	0.3148	0.3415	0.3053	0.4326	0.3290	0.2557
Sbi	0.3191	0.3508	0.3371	0.4031	0.3227	0.2844
Minimum	0.0839	0.0398	0.0130	0.0080	0.0270	0.1475
Maximum	0.9526	0.9526	0.9526	0.9526	0.9526	0.9526
Range	0.8686	0.9127	0.9396	0.9446	0.9256	0.8051
Lower quartile	0.1731	0.0914	0.0909	0.0686	0.1570	0.2476
Upper quartile	0.7943	0.8202	0.7565	0.9349	0.7738	0.7725
Interquartile range	0.6212	0.7288	0.6656	0.8663	0.6167	0.5249
1/6 sextile	0.1300	0.0750	0.0357	0.0292	0.1427	0.2100
5/6 sextile	0.9317	0.8415	0.8303	0.9414	0.8147	0.8576
Intersextile range	0.8017	0.7665	0.7946	0.9122	0.6721	0.6476
Skewness	0.1710	0.0148	-0.0392	-0.0455	0.0217	0.1500
Stnd. skewness	0.4130	0.0357	-0.0947	-0.1099	0.0524	0.3622
Kurtosis	-1.3727	-1.7412	-1.4489	-1.8402	-1.7030	-1.5350
Stnd. kurtosis	-1.6577	-2.1027	-1.7497	-2.2223	-2.0566	-1.8540
Sum	17.3369	16.2886	16.0196	17.5264	16.0973	18.1390
Sum of squares	11.9198	11.5068	10.9465	14.0680	10.7362	11.9900

Sources: developed by the authors based on the Statgraphics software.

The next modelling step is applying the Multiple Regression Backward Selection (MRBS) to sift out variables. K1 is a dependent variable. K2, K3, K4, K5, K6 are independent variables. The model is developed via the Statgraphics Centurion 19 software and comprises 3 years.

Firstly, we analyse input data and define the R-square value as 98.4915%. The adjusted R-square is 98.3455%; MSE is 0.0300303.

Secondly, we remove K2 with the P-to-remove index as 0.507289. Now, 4 variables remain in the model. R-square is 98.55%. The adjusted R-square is 98.35%; MSE is 0.001614.

Thirdly, we remove K5. 3 variables (K3, K4, K6) remain in the model. R-square is 98.49%. The adjusted R-square is 98.35%; MSE is 0.0016215.

The regression model equation has the following form:

$$K1 = -0.0982849 + 0.311939 \cdot K3 - 0.238534 \cdot K4 + 1.10041 \cdot K6 \quad (5)$$

In case of unchanging K3, K4 and K6, K1 (coverage of state and private insurances) falls by -0.0982849 in each extra period. K3 (national employment) affects K1: the number of insured persons rises by 0.311939 annually. K4 (life expectancy) has an inverse influence on insurance. It is deceased by -0.238534: the older you are, the more money you should pay for insurance. K6 (national costs for medical goods and services) makes K1 increase by 1,10041 within each possible historical date.

Table 4 shows values of obtained regression model coefficients and their statistical significance check via the Student criterion, standard error and p-value. The Student criterion values for 35 variables as 2.030 with p=0.05 reflect contrast between samples. In case of p<0.05, all variables are significant within the regression equation.

Table 4. The Student test

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-0.0982849	0.0266862	-3.68298	0.0009
k3	0.311939	0.0754862	4.1324	0.0003
k4	-0.238534	0.0910946	-2.61853	0.0135
k6	1.10041	0.151703	7.2537	0.0000

Sources: developed by the authors based on the Statgraphics software.

Table 5. Analysis of variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	3.28185	3	1.09395	674.65	0.0000
Residual	0.0502666	31	0.0016215		
Total (Corr.)	3.33212	34			

Sources: developed by the authors based on the Statgraphics software.

R-square as 98,4915% shows that the variables are highly interrelated. According to this value, practically all dependent variables between independent ones can be covered via the linear regression. Such a model is very statistically significant.

R-square (adjusted for d.f.) as 98.3455% reflects dispersion via dependent variables, which is explained by the model. This assessment is extremely high.

Standard error as 0.0402679 demonstrates low deviation of our data. Thus, they are reliable.

Within the model, the mean absolute error is usually 0.0300303.

Lag 1 residual autocorrelation as 0.68334 reflects interrelations between previous and current values. It means that the model can evolve further.

The next stage is the MARS model generation via the Salford Predictive Modeler 8 software. Here, we engage K1, K3, K4, K6. To accomplish the task, we chose such settings: target and predicted variables, regression, the MARS analysing engine, basis function search with their limit of 40 units. Predicted variables are related in pairs.

The MARS model generation results in 11 basis functions. Among them, the 5th basis function is optimal (according to the minimal GCV – Figure 3).

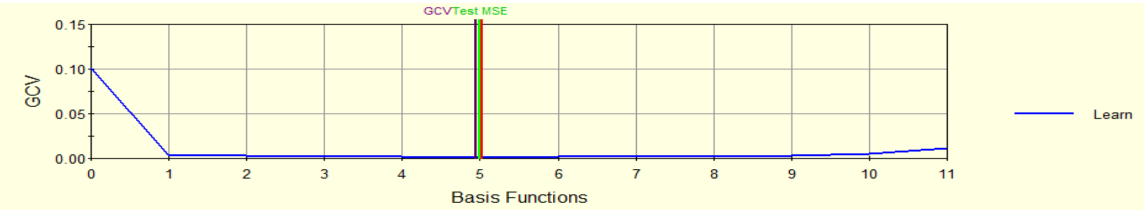

Figure 3. The MARS visualisation

Sources: developed by the authors based on the Salford Predictive Modeler 8 software.

Detailed statistical information about all produced basis functions is given in Table 6.

Table 6. Description of all basis functions

Basis Functions	N Predictors	N Inputs	Effective Parameters	GCV	GCV R-Sq	Learn MSE	Learn R-Sq
11	3	3	30.00	0.01190	0.88203	0.00024	0.99745
10	3	3	27.36	0.00514	0.94909	0.00024	0.99743
9	3	3	24.73	0.00287	0.97153	0.00025	0.99740
8	3	3	22.09	0.00187	0.98144	0.00025	0.99732
7	3	3	19.45	0.00188	0.98133	0.00037	0.99610
6	3	3	16.82	0.00159	0.98428	0.00043	0.99551
5	3	3	14.18	0.00129	0.98721	0.00046	0.99521
4	3	3	11.55	0.00152	0.98494	0.00068	0.99283
3	3	3	8.91	0.00200	0.98021	0.00111	0.98835
2	2	2	6.27	0.00329	0.96738	0.00222	0.97672
1	1	1	3.64	0.00310	0.96926	0.00249	0.97384
0	0	0	1.00	0.10089		0.09520	

Sources: developed by the authors based on the Salford Predictive Modeler 8 software.

Below we are going to explain the optimal basis function №5. The MARS model consists of five basis functions. The optimal MARS model using four basis functions is shown in Formula (6) and Table 7.

$$K1 = 0.304392 + 0.488719 \cdot BF1 + 2.79333 \cdot BF2 - 0.598569 \cdot BF5 - 1.76137 \cdot BF6 + 1.03716 \cdot BF10 \tag{6}$$

The MARS basis functions (6) are expressed as:

$$BF1 = \max(0, K6 - 0.147524) \quad (7)$$

Table 7. Basis functions

Basis Function	Coefficients	Variable	Sign	Parent Sign	Parent	Knot
0	0.30439					
1	0.48872	K6	+			0.14752
2	2.79333	K4	+	+	K6	0.85678
5	-0.59857	K3	-			0.39604
6	-1.76137	K6	+	+	K3	0.77252
10	1.03716	K3	+	+	K6	0.63465

Sources: developed by the authors based on the Salford Predictive Modeler 8 software.

Statistical characteristics of the optimal MARS model (6) are indicated in Table 8.

Table 8. Statistical characteristics of the optimal MARS model

Name	Learn
RMSE	0.02136
MSE	0.00046
GCV	0.00129
MAD	0.01644
MAPE	0.04893
SSY	3.33212
SSE	0.01597
R-Sq	0.99521
R-Sq Norm	0.99521
GCV R-Sq	0.98721
MSE Adjusted	0.00038
R-Sq Adjusted	0.99438

Sources: developed by the authors.

The research results in comparing Table 9. Here, such data are included as input coverage of state and private insurance (USA Facts, 2023) and calculated regression model values (5). The latter describes how state and private insurances (K1) depend on national employment (K3), life expectancy (K4), public costs for medical goods and services (K6). Besides, Table 9 represents the defined MARS model values (6) within six basis functions.

Table 9. Comparison of actual labour efficiency values and predicted ones via regression (5) and MARS (6)

Year	K1	K1 MARS	K1 reg	$\|\Delta\|MARS$	$\|\Delta\|reg$
2021	0.952574	0.974898	1.003419	0.022323	0.050845
2020	0.949729	0.912644	0.967712	-0.03708	0.017983
2019	0.951887	0.974182	0.98042	0.022295	0.028533
2018	0.945267	0.951095	0.959171	0.005828	0.013904
2017	0.947382	0.940841	0.927871	-0.00654	-0.01951
2016	0.931655	0.92634	0.896165	-0.00532	-0.03549
2015	0.921668	0.898265	0.853535	-0.0234	-0.06813
2014	0.886664	0.855464	0.798812	-0.0312	-0.08785
2013	0.794325	0.803941	0.740224	0.009616	-0.0541
2012	0.697886	0.730421	0.708827	0.032535	0.010941
2011	0.658715	0.680012	0.653288	0.021297	-0.00543
2010	0.607714	0.626547	0.624455	0.018833	0.016741
2009	0.588574	0.565296	0.613445	-0.02328	0.024871
2008	0.609156	0.606108	0.657634	-0.00305	0.048478
2007	0.584477	0.599699	0.647774	0.015222	0.063297
2006	0.533372	0.549862	0.602528	0.016491	0.069156
2005	0.521076	0.497087	0.545534	-0.02399	0.024458

Continued Table 9

Year	K1	K1 MARS	K1 reg	$\|\Delta\|MARS$	$\|\Delta\|reg$
2004	0.5	0.476654	0.490683	-0.02335	-0.00932
2003	0.45322	0.457392	0.450205	0.004172	-0.00302
2002	0.450592	0.436091	0.401426	-0.0145	-0.04917
2001	0.419197	0.414208	0.371233	-0.00499	-0.04796
2000	0.402588	0.397515	0.346065	-0.00507	-0.05652
1999	0.347995	0.449696	0.289143	0.101701	-0.05885
1998	0.208073	0.477331	0.256497	0.269258	0.048424
1997	0.19044	0.498633	0.230868	0.308193	0.040428
1996	0.185156	0.525824	0.203705	0.340667	0.018548
1995	0.173127	0.535978	0.186197	0.362851	0.01307
1994	0.161707	0.54292	0.169796	0.381212	0.008089
1993	0.143259	0.55175	0.152534	0.40849	0.009275
1992	0.129975	0.551627	0.136439	0.421652	0.006465
1991	0.115651	0.547306	0.12189	0.431655	0.00624
1990	0.104699	0.537541	0.110524	0.432842	0.005824
1989	0.097188	0.535051	0.093648	0.437864	-0.00354
1988	0.087977	0.534772	0.079096	0.446795	-0.00888
1987	0.08394	0.533668	0.06621	0.449728	-0.01773

Sources: developed by the authors.

Due to Table 9, the MARS model is more accurate for 2000-2021 rather than for 1987-1999. Obtained via the regression analysis, the values produce more accurate predictions for 1987-1999 rather than 2000-2021.

Conclusions. To understand how labour efficiency is affected by national employment, life expectancy, healthcare costs as a GDP percentage and public costs for medical goods and services within the USA medical insurance, we conducted a three-stage study.

Firstly, we normalised and described the research data.

Secondly, the Statgraphics software was applied to process values of national employment, life expectancy, healthcare costs as a GDP percentage and public costs for medical goods and services.

Thirdly, we generated the MARS model via the software Salford Predictive Modeler 8. Within the obtained basis model, hinge points were detected with assessing relations between independent values (K3, K4, K6) and dependent ones (K1).

Finally, we compared initial values and those calculated through the regression and MARS models. The acquired MARS model values are more similar to initial ones for 2000-2021. Also, calculated values and actual normalised indexes differ in 1987-1999. Simultaneously, the regression model has more accurate indexes in 1987-1999 rather than in 2000-2021. Use of these two methods opens search for other significant variables.

Author Contributions: conceptualization, S.D.; methodology, V.K.; software, S.D.; validation, V.K. and S.D.; formal analysis, B.K.; investigation, V.K.; resources, B.K. and S.D.; data curation, B.K. and S.D.; writing-original draft preparation, V.K.; writing-review and editing, V.K.; visualization, S.D.; supervision V.K.

References

Ajay, G. (2021). *Module 4 Practice of Life Insurance*, 75-94. Retrieved from [Link]

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. [Google Scholar] [CrossRef]

Ben Halima, M. A., & Rococo, E. (2014). Wage differences according to health status in France. *Social Science and Medicine*, 120, 260-268. [Google Scholar] [CrossRef]

Bottegal, G., & Pillonetto, G. (2018). The generalised cross validation filter. *Automatica*, 90, 130-137. [Google Scholar] [CrossRef]

Craven, P., & Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische Mathematik*, 31, 377-403. [Google Scholar] [CrossRef]

Dizioli, A., & Pinheiro, R. (2016). Health insurance as a productive factor. *Labour Economics*, 40, 1-24. [Google Scholar] [CrossRef]

Effroymson, M. A. (1960) Multiple regression analysis. In Ralston, A., & Wilf, H.S. (Eds.). *Mathematical Methods for Digital Computers* (pp. 191-203). Wiley, New York. [Google Scholar]

Erlangga, D., Suhrcke, M., Ali, S., & Bloor K. (2019). The impact of public health insurance on healthcare utilisation, financial protection and health status in low- and middle-income countries: A systematic review. *PLoS One*, 14(8). [Google Scholar] [CrossRef]

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1), 1-67. [Google Scholar] [CrossRef]

Ho, A. (2015). Health insurance. In: ten Have, H. (Eds.), *Encyclopaedia of Global Bioethics* (pp. 1-9). Springer, Cham. [Google Scholar] [CrossRef]

Macrotrends (2023). USA life expectancy 1950-2023. Retrieved from [Link]

Nguyen, A. (2016). International healthcare systems. Part 4: The out-of-pocket model. *Morning Sign Out*. Retrieved from [Link]

Nguyen, A. (2017). International healthcare systems. Part 3: The Bismarck model. *Morning Sign Out*. Retrieved from [Link]

OECD (2023). Labour productivity forecast. Retrieved from [Link]

PNHP (2010). *Healthcare Systems: Four Basic Models*. Retrieved from [Link]

Rochet, J. (1991). Incentives, redistribution and social insurance. *The Geneva Papers on Risk and Insurance Theory*, 16(2), 143-165. [Google Scholar]

Shen, Z., Parker, M., Brown, D., & Fang, X. (2017). Effects of public health insurance on labour supply in rural China. *China Agricultural Economic Review*, 9(4), 623-642. [Google Scholar] [CrossRef]

Smirnova, V., Klymuk, N., & Vakulenko, D. (2020). Analysis of medical insurance models. *Bulletin of Social Hygiene and Healthcare in Ukraine*, 3, 103-105. [CrossRef]

Sommers, B. (2017). Why health insurance matters and why research evidence should too. *Academic Medicine*, 92(9), 1228-1230. [Google Scholar] [CrossRef]

USA Bureau of Labour Statistics (2023). Labor force statistics from the current population survey. Retrieved from [Link]

USA Facts (2023). Healthcare expenditures as a percent of GDP. Retrieved from [Link]

USA Facts (2023). National spending on healthcare goods and services. Retrieved from [Link]

USA Facts (2023). People covered by public or private health insurance. Retrieved from [Link]

Віталія Койбічук, к.е.н., доцент, Сумський державний університет, Україна
Бхола Хан, професор, Університет Махатми Джіотіби Пхуле Рохілханд, Індія
Сергій Дрозд, Сумський державний університет, Україна

Особливості медичного страхування США як фактор-стимулятор до підвищення ефективності праці

Система медичного страхування є критично важливим чинником продуктивності праці в країні. У багатьох країнах медичне страхування пов'язане з роботою, а це означає, що працівники повинні мати активне страхування, щоб отримати доступ до необхідних їм медичних послуг. Це гарантує, що працівники здорові та здатні витримувати фізичні та розумові навантаження, пов'язані з роботою. У країнах, де медичне обслуговування забезпечує держава, громадяни мають загальний доступ до медичних послуг, незалежно від статусу зайнятості. Це означає, що люди можуть звертатися за медичною допомогою, коли вона їм потрібна, не турбуючись про вартість, що допоможе зберегти їхнє здоров'я та продуктивність. У країнах із медичним страхуванням на основі зайнятості роботодавці несуть відповідальність за медичне страхування своїх працівників. Це є значною вартістю для бізнесу, а також призведе до того, що працівники почуватимуться менш мобільними, оскільки вони можуть неохоче залишати роботу, якщо це означає втрату медичної страховки. Система медичного страхування в країні є основним фактором, що визначає ефективність праці. Для забезпечення високого рівня продуктивності праці країни повинні створювати системи, які є доступними, недорогими та забезпечують надійне покриття. Для проведення дослідження було обрано показники: охоплення людей державним або приватним страхуванням, рівень продуктивності праці, рівень зайнятості населення, середня тривалість життя, витрати на систему охорони здоров'я у відсотках від ВВП, витрати на систему охорони здоров'я в натуральному вираженні. Було проведено дослідницьку роботу з використанням методів нормалізації даних, пошуку та побудови регресійної моделі на основі методо зворотного відбору даних, застосовано Multivariate Adaptive Regression Splines або MARS метод заснований на регресійній моделі, що використовує сплайн-функції для опису нелінійних відносин між змінними. В дослідженні використано методи бібліографічного аналізу, обробки даних, систематизації, порівняння, логічного узагальнення. Результати цього дослідження мають важливе значення для політиків і бізнесу. Політики можуть використовувати цю інформацію для розробки соціально-економічних принципів, які покращать систему медичного страхування та підвищить продуктивність праці. Підприємства можуть використовувати цю інформацію для прийняття рішень щодо виплат медичного страхування, які можуть покращити здоров'я та продуктивність працівників.

Ключові слова: страхування; здоров'я; зайнятість населення; середня тривалість життя; MARS; регресія.