

Prognose der Zeitreihen „Bruttowertschöpfung in der Gesundheitswirtschaft in Sachsen“ und „Erwerbstätige in der Gesundheitswirtschaft in Sachsen“

Vorbemerkungen

Das Ziel der in Zusammenarbeit mit dem Statistischen Landesamt des Freistaates Sachsen erstellten Diplomarbeit mit dem Titel „Möglichkeiten und Grenzen der Anwendung von Methoden der Zeitreihenanalyse zur Vorabrechnung von Bruttowertschöpfung und Erwerbstätigenzahl der Gesundheitswirtschaft in Sachsen“ [1] war die Klärung, inwieweit die zeitnahe Bereitstellung von Rechenergebnissen zur Gesundheitswirtschaft in Sachsen – speziell zur Bruttowertschöpfung und zur Zahl der Erwerbstätigen – auf der Basis der zur Verfügung stehenden Informationen zur bisherigen Entwicklung dieser Indikatoren und unter Nutzung statistischer Methoden der Zeitreihenanalyse realisiert werden kann. In diesem Zusammenhang wurde ein Rechenwerk zur Durchführung von Prognosen für die nächsten Jahre entwickelt.¹⁾

Durch die Institutionalisierung des Wertschöpfungsansatzes²⁾ im Statistischen Landesamt können kontinuierlich Ergebnisse zur Bruttowertschöpfung sowie zur Erwerbstätigenzahl in der Gesundheitswirtschaft in Sachsen bereitgestellt werden. Die Berechnungen werden auf der Grundlage der Veröffentlichungen des Arbeitskreises Volkswirtschaftliche Gesamtrechnungen der Länder sowie unter Nutzung weiterer Statistiken im Rahmen der Gesundheitsökonomischen Gesamtrechnungen durchgeführt, die teilweise erst zwei Jahre nach dem Berichtsjahr zur Verfügung stehen. Dadurch ist eine zeitnahe Bereitstellung von Ergebnissen unter Nutzung dieses Ansatzes nicht möglich.

Im Rahmen der erstellten Diplomarbeit wurde die Frage geklärt, inwieweit trotz dieser Situation eine zeitnahe Bereitstellung von belastbaren Ergebnissen zur Bruttowertschöpfung und zur Zahl der Erwerbstätigen in der Gesundheitswirtschaft in Sachsen realisiert werden kann.

Dieser Artikel stellt eine Zusammenfassung der Diplomarbeit bezüglich der Vorgehensweise zur Lösung der Problemstellung dar. Weiterhin werden die Ergebnisse analysiert sowie das Verfahren der Arbeit bewertet.

Ausgangssituation

Auf der Grundlage der Daten zur Bruttowertschöpfung sowie zur Zahl der Erwerbstätigen in der Gesundheitswirtschaft in Sachsen für die Jahre 1996 bis 2008 sollte je eine Prognose für die nächsten drei Jahre ermittelt werden, die veröffentlicht werden kann (vgl. Tab. 1).

Bevor eine Prognose vorgenommen werden konnte, musste die optimale Trendfunktion der gegebenen Zeitreihe bestimmt und mit

dem Test auf Zufälligkeit geprüft werden, ob in der trendbereinigten Reihe noch Korrelationen vorhanden waren. War dies nicht der Fall, so konnte ein Signifikanztest angewendet werden. Falls ein signifikanter Trend vorhanden war, konnte eine qualifizierte Prognose durchgeführt werden.

- 1) Der hier vorgestellte Ansatz ist grundsätzlich auch für Schätzungen der Bruttowertschöpfung sowie Zahl der Erwerbstätigen der Gesundheitswirtschaft in der Abgrenzung nach WZ 2008 geeignet.[2]
- 2) Mit dem Wertschöpfungsansatz können die wirtschaftlichen Tätigkeiten aller in der Gesundheitswirtschaft aktiven Wirtschaftseinheiten in den Kategorien der Volkswirtschaftlichen Gesamtrechnungen abgebildet werden. Die Basis dafür ist eine wirtschaftsfachliche Abgrenzung der Gesundheitswirtschaft gemäß der Klassifikation der Wirtschaftszweige. Das Ziel des Wertschöpfungsansatzes ist es, die Bruttowertschöpfung und die Zahl der Erwerbstätigen in der Gesundheitswirtschaft zu ermitteln.

Tab. 1 Bruttowertschöpfung¹⁾ und Erwerbstätige in der Gesundheits-²⁾ und Gesamtwirtschaft 1996 bis 2008

Jahr	Bruttowertschöpfung		Erwerbstätige	
	Gesundheitswirtschaft	Gesamtwirtschaft	Gesundheitswirtschaft	Gesamtwirtschaft
	Millionen €		1 000 Personen	
1996	6 182	65 701	208	2 005
1997	6 362	66 089	209	1 977
1998	6 620	66 739	213	1 975
1999	6 831	67 926	217	1 982
2000	7 131	68 098	215	1 972
2001	7 211	70 386	220	1 940
2002	7 634	73 373	227	1 916
2003	7 861	74 926	233	1 907
2004	8 143	77 378	234	1 912
2005	8 430	77 411	235	1 895
2006	8 809	80 884	238	1 911
2007	8 970	83 969	242	1 941
2008	9 288	84 884	247	1 953

1) in jeweiligen Preisen

2) Schätzung Stand April 2011; alle Angaben vorläufig.

Datenquelle: AK VGRdL; Berechnungsstand August 2010/ Februar 2011.

Das Komponentenmodell

Für beide gegebenen Zeitreihen – mit je n Beobachtungswerten – wurde das additive Komponentenmodell ohne Saisonkomponente angesetzt, denn es sind nur Jahreswerte vorhanden:

$$x_t = m_t + u_t, \quad t = 1, \dots, n$$

Dabei beschreibt der Trend m_t die langfristige systematische Veränderung des mittleren Verlaufs der Zeitreihe. Die Restkomponente u_t beinhaltet die kurzfristigen, zufällig um Null schwankenden Einflüsse oder Störungen.

Bestimmung der optimalen Trendfunktion

Bei der Auswahl der optimalen Trendfunktion für die entsprechende Zeitreihe wurde sowohl die Differenzenmethode als auch das Kriterium der Restvarianz angewendet. Mit der Differenzenmethode kann ein polynomialer Trend vom Grad d ($d \geq 1$) durch die Bildung der Differenzen d -ter Ordnung beseitigt werden. Dieser Grad entspricht der Anzahl der Differenzenbildungen bis der variable Trend wegfällt. Differenzen 1. Ordnung:

$$\nabla x_t = x_t - x_{t-1} \quad \text{für } t = 2, 3, \dots, n$$

Differenzen d -ter Ordnung:

$$\nabla^d x_t = \nabla^{d-1} x_t - \nabla^{d-1} x_{t-1} \quad \text{für}$$

$$t = d + 1, \dots, n$$

Für das Kriterium der Restvarianz wurden den beiden gegebenen Zeitreihen folgende Trendfunktionen angepasst, die den Verlauf unterschiedlich gut wiedergeben:

Linearer Trend

$$m_t = a_0 + a_1 \cdot t$$

Quadratischer Trend

$$m_t = a_0 + a_1 \cdot t + a_2 \cdot t^2$$

Kubischer Trend

$$m_t = a_0 + a_1 \cdot t + a_2 \cdot t^2 + a_3 \cdot t^3$$

Exponentieller Trend

$$m_t = a_0 \cdot a_1^t$$

Logistischer Trend:

$$m_t = \frac{a_0}{1 + a_1 \cdot e^{-a_2 \cdot t}}$$

Zur Schätzung der unbekannt Parameter der entsprechenden Trendfunktion wurde

die Methode der kleinsten Quadrate genutzt. Durch das Einsetzen der geschätzten Parameter in die Regressionsfunktion war es möglich, den geschätzten Trend

$$\hat{m}_t = \hat{x}_t$$

zu erhalten. Die Ermittlung der Residuen erfolgte durch Subtraktion des geschätzten Trends von dem Beobachtungswert der Zeitreihe:

$$\hat{u}_t = x_t - \hat{x}_t$$

Anschließend konnte die jeweilige Restvarianz berechnet werden:

$$\hat{s}_l^2 = \frac{1}{n - m} \cdot \sum_{t=1}^n (x_t - \hat{x}_t)^2$$

Dabei ist l der Trendfunktionstyp – wie der lineare Trend – und m die Anzahl der freien Parameter – beispielsweise gilt für den linearen Trend $m=2$.

Zur Auswertung dieses Kriteriums lässt sich sagen, dass die Trendfunktion mit der kleinsten Restvarianz am besten an die Zeitreihe angepasst ist.

Durchführung des Tests auf Zufälligkeit

Die Signifikanztests der Regressionsanalyse dürfen nur genutzt werden, wenn die Residuen als Realisierungen von unabhängigen, normal verteilten Zufallsgrößen angesehen werden können. Sonst gelten diese Tests nur als grobe Näherung. Meist ist die genannte Voraussetzung für die Residuen von Zeitreihen nicht erfüllt, da in der Regel Abhängigkeiten zeitlich aufeinanderfolgender Werte – sogenannte Autokorrelationen – auftreten. Mit dem Test auf Zufälligkeit konnte geprüft werden, ob die Restkomponente als zufällig angesehen werden konnte – sofern die Annahme gerechtfertigt war, dass die Residuen Realisierungen von Zufallsgrößen waren. Vor der Durchführung des Tests war es notwendig, die empirische Autokorrelationsfunktion zu bestimmen:

$$r_k = \frac{c_k}{c_0}$$

$$\text{mit } c_k = \frac{1}{n} \cdot \sum_{t=1}^{n-k} (u_t - \bar{u}) \cdot (u_{t+k} - \bar{u})$$

$$\text{für } k = 0, 1, \dots, \frac{n}{4}$$

Unter der Nullhypothese

$$H_0 : r_k^* = 0$$

(r_k^* ist das betragsgrößte r_k) konnte die Testgröße berechnet werden:

$$t_z = r_k^* \cdot \sqrt{\frac{n-2}{1-(r_k^*)^2}}$$

Nach der Festlegung der Irrtumswahrscheinlichkeit α (meist 0,05) konnte der kritische Wert $t_{n-2; 1-\alpha/2}$ ermittelt werden. Die Null-

hypothese wurde abgelehnt, wenn

$$|t_z| \geq t_{n-2; 1-\alpha/2}$$

Ansonsten konnte die Restkomponente als zufällig angesehen werden, es waren keine Korrelationen in der Reihe.

Durchführung des Signifikanztests

Wurde die Nullhypothese des Tests auf Zufälligkeit nicht abgelehnt, konnte ein Signifikanztest durchgeführt werden.

Für den Fall, dass der lineare Trend als optimale Trendfunktion bestimmt wurde, konnte ausgehend von der berechneten linearen Regressionsfunktion und der zugehörigen Restvarianz überprüft werden, ob die Variablen a_0 und a_1 einen signifikanten Einfluss auf die Variable x_t haben. Das heißt, es waren die Nullhypothesen

$$H_0^{(1)} : a_0 = 0 \quad \text{und}$$

$$H_0^{(2)} : a_1 = 0 \quad \text{zu testen. Nach der Wahl}$$

der Irrtumswahrscheinlichkeit α konnten die Testgrößen berechnet werden:

$$t_{linear}^{a_0} = \frac{\hat{a}_0}{\sqrt{\hat{s}_{linear}^2 \cdot \left(\frac{1}{n} + \frac{\bar{t}^2}{\sum_{t=1}^n (t - \bar{t})^2} \right)}} \quad \text{und}$$

$$t_{linear}^{a_1} = \frac{\hat{a}_1}{\sqrt{\hat{s}_{linear}^2 \cdot \sum_{t=1}^n (t - \bar{t})^2}}$$

Der kritische Wert war $t_{n-2; 1-\alpha/2}$. Für

$$|t_{linear}^{a_0}| \geq t_{n-2; 1-\alpha/2} \quad \text{wurde } H_0^{(1)} \text{ abge-}$$

lehnt und dies bedeutet, dass a_0 einen signifikanten Einfluss auf x_t hatte. War die Testgröße kleiner als der kritische Wert, konnte die Nullhypothese nicht abgelehnt werden.

Die Entscheidungen für $H_0^{(2)}$ waren analog zu treffen.

Wurde der quadratische Trend als optimale Trendfunktion bestimmt, konnte auf der Basis der berechneten quadratischen Regressionsfunktion und deren Restvarianz geprüft werden, ob die Variablen a_1 und a_2 einen signifikanten Einfluss auf x_t haben. Dieser Test erfolgte entsprechend den Test für den linearen Trend, abgesehen von wenigen Änderungen: Der kritische Wert war $t_{n-3;1-\alpha/2}$

und die Testgrößen waren folgendermaßen zu berechnen:

$$t_{quadr.}^{a_1} = \frac{\hat{a}_1}{\sqrt{\hat{S}_{quadr.}^2 \cdot \frac{c_{22}}{c_{11} \cdot c_{22} - c_{12}^2}}}$$

$$\text{und } t_{quadr.}^{a_2} = \frac{\hat{a}_2}{\sqrt{\hat{S}_{quadr.}^2 \cdot \frac{c_{11}}{c_{11} \cdot c_{22} - c_{12}^2}}}$$

$$\text{mit } c_{11} = \sum_{t=1}^n (t - \bar{t})^2$$

$$c_{22} = \sum_{t=1}^n (t^2 - \bar{t}^2)^2 \text{ und}$$

$$c_{12} = \sum_{t=1}^n (t - \bar{t}) \cdot (t^2 - \bar{t}^2)$$

Da die anderen oben beschriebenen Trendfunktionen nicht zur Anwendung kamen, wird auf eine Darstellung der Signifikanzprüfung an dieser Stelle verzichtet.

Durchführung verschiedener Prognosen

Es wurden verschiedene Prognoseverfahren angewendet, wie zum Beispiel die naive, die lokale und die globale Prognose. Der Prognosewert zurzeit n bei h Schritten in die Zukunft wird mit $\hat{x}_{n,h}$ bezeichnet.

Die naive Prognose

$$\hat{x}_{n,h} = x_n + h \cdot (x_n - x_{n-1})$$

berücksichtigt nur die Veränderung zwischen den Beobachtungswerten der letzten beiden Jahre und geht davon aus, dass diese auch in der Zukunft gleich bleibt. Bei einer lokalen Prognose wird der Trend über wenige aktuelle Werte der Zeitreihe geschätzt und anschließend in die Zukunft fortgesetzt:

$$\hat{x}_{n,h} = \hat{m}_{n+h}^{lokal}$$

Bei der globalen Prognose

$$\hat{x}_{n,h} = \hat{m}_{n+h}$$

wird der Trend über die gesamte Zeitreihe bestimmt.

Umsetzung der Berechnungen mittels VBA-Programmierung³⁾

Für beide Zeitreihen erfolgte die Bestimmung der optimalen Trendfunktion, die Durchführung des Tests auf Zufälligkeit, des Signifikanztests sowie verschiedener Prognosen mittels Microsoft Excel. Es wurden Makros in der Programmiersprache VBA entwickelt, welche alle Berechnungen automatisch ausführen und anschließend die Ergebnisse in einem neuen Tabellenblatt ausgeben. Diese erstellten Makros können auch in den nächsten Jahren genutzt werden, um Prognosen zu berechnen.

Auswertung

Für die beiden gegebenen Zeitreihen konnten folgende Funktionen als optimale Trendfunktionen bestimmt werden, welche den Verlauf der Zeitreihe sehr gut wiedergaben: Für die Bruttowertschöpfung war die quadratische

Trendfunktion und für die Zahl der Erwerbstätigen die lineare Trendfunktion am besten geeignet.

Die Residuen konnten als Realisierungen von Zufallsgrößen betrachtet werden: Der Prozess der Datengenerierung kann als zufällig angesehen werden, denn zur Ermittlung des Ergebnisses für die Bruttowertschöpfung und die Zahl der Erwerbstätigen in der Gesundheitswirtschaft in Sachsen sind teilweise Schätzungen und individuelle Entscheidungen notwendig. Daher konnte der Test auf Zufälligkeit angewendet werden und das Ergebnis war, dass es keine Korrelationen in den trendbereinigten Reihen gibt.

Somit konnte der entsprechende Signifikanztest durchgeführt werden und dieser ergab, dass ein signifikanter Trend vorlag und eine qualifizierte Prognose möglich war.

Es wurden verschiedene Prognosen durchgeführt: eine naive Prognose, mehrere lokale lineare Prognosen mit unterschiedlicher Stützstellenanzahl sowie globale Prognosen. Aufgrund der Kürze der vorhandenen Zeitreihen wurden nur für einen Zeitraum von fünf Jahren Werte prognostiziert.

Die naive Prognose erzielte bei beiden Zeitreihen hohe Werte, denn die Veränderung in

3) VBA: Visual Basic for Application

Tab. 2 Berechnungs-¹⁾ und Prognosewerte zur Bruttowertschöpfung²⁾ in der Gesundheitswirtschaft 2006 bis 2008

Jahr	Berechnungswerte	Prognosewerte				
		naive Prognose	lokale lineare Prognose mit der folgenden Anzahl an Stützstellen			globale quadratische Prognose
			4	5	6	
Millionen €						
2006	8 809	8 717	8 685	8 740	8 687	8 740
2007	8 970	9 004	8 952	9 035	8 959	9 057
2008	9 288	9 290	9 219	9 329	9 231	9 384

1) Schätzung Stand April 2011; alle Angaben vorläufig.

Datenquelle: AK VGRdL; Berechnungsstand August 2010/ Februar 2011.

2) in jeweiligen Preisen

Tab. 3 Prognosewerte zur Bruttowertschöpfung¹⁾ in der Gesundheitswirtschaft 2006 bis 2008 – relative Abweichung zu den Berechnungswerten²⁾

Jahr	Naive Prognose	Lokale lineare Prognose mit der folgenden Anzahl an Stützstellen			Globale quadratische Prognose
		4	5	6	
		%			
2006	-1,0	-1,4	-0,8	-1,4	-0,8
2007	0,4	-0,2	0,7	-0,1	1,0
2008	0,0	-0,8	0,4	-0,6	1,0

1) in jeweiligen Preisen

2) Schätzung Stand April 2011; alle Angaben vorläufig.

Datenquelle: AK VGRdL; Berechnungsstand August 2010/ Februar 2011.

den letzten beiden Jahren war relativ hoch. Die globalen linearen sowie die lokalen linearen Prognosen ergaben geringere und sinnvollere Werte. Bei einer globalen Prognose mit der quadratischen Trendfunktion (welche bei der Zeitreihe zur Bruttowertschöpfung in der Gesundheitswirtschaft in Sachsen als optimal bestimmt wurde) erhält man schnell sehr hohe und somit unrealistische Prognosewerte.

Die ermittelten Trendfunktionen wurden einem Praxistest unterzogen. Es wurden Prognosen mit den verkürzten Zeitreihen für die Jahre 1996 bis 2005 vorgenommen und anschließend wurden die prognostizierten Werte mit den tatsächlichen Werten verglichen. Die Ergebnisse sind in den Tabellen 2 bis 5 dargestellt. Es haben sich recht gute Schätzungen ergeben. Die relativen Abweichungen der Prognose- von den Berechnungswerten waren für die Bruttowertschöpfung nie größer als 1,4 Prozent (vgl. Tab. 3), für die Zahl der Erwerbstätigen wurde eine maximale Abweichung von 4,1 Prozent ermittelt (vgl. Tab. 5). Daraus resultierend konnten Empfehlungen zur Vorausberechnung der beiden Zeitreihen gegeben werden.

Dabei konnten die im ersten Schritt der Auswertung auf der Grundlage aller zur Verfügung stehenden Daten ermittelten Trendfunktionen bestätigt werden.

Die dort für die Bruttowertschöpfung ermittelte quadratische Trendfunktion liefert sehr gute Schätzungen, allerdings dürfte diese Trendfunktion sicher bald zu große Prognosewerte liefern. Deshalb sollte für die Zeitreihe zur Bruttowertschöpfung in der Gesundheitswirtschaft in Sachsen eine lokale lineare Prognose für eine verkürzte Zeitreihe ab dem Jahr 2001 oder 2002 vorgenommen werden, für die Zeitreihe zu den Erwerbstätigen in der Gesundheitswirtschaft in Sachsen ist eine globale lineare Prognose zu bevorzugen.

Zusammenfassung und Bewertung

Der hier vorgestellte Ansatz zur Prognose der Bruttowertschöpfung und der Zahl der Erwerbstätigen in der Gesundheitswirtschaft ist für Schätzungen dieser beiden Indikatoren grundsätzlich geeignet. Trotzdem sollten nur Prognosen für etwa drei bis fünf Jahre durchgeführt werden. Für diesen Prognosehorizont werden sich noch relativ zuverlässige Ergebnisse ermitteln lassen, die nicht allzu sehr von den exakten Werten abweichen. Allerdings sollte man den Prognosehorizont nicht viel größer wählen, da es sehr wahrscheinlich ist, dass diese Prognosen nicht mehr zutreffen werden.

Anne Schneider, Dipl.-Wirtschaftsmathematikerin

Tab. 4 Berechnungs-¹⁾ und Prognosewerte zur Zahl der Erwerbstätigen in der Gesundheitswirtschaft 2006 bis 2008

Jahr	Berechnungswerte	Prognosewerte				
		naive Prognose	lokale lineare Prognose mit der folgenden Anzahl an Stützstellen		globale quadratische Prognose	
			3	5	linear	logistisch
			1 000 Personen			
2006	238	235	235	240	240	236
2007	242	236	236	244	243	239
2008	247	237	237	247	246	241

1) Schätzung Stand April 2011; alle Angaben vorläufig.
Datenquelle: AK VGRdL; Berechnungsstand August 2010/ Februar 2011.

Tab. 5 Prognosewerte zur Zahl der Erwerbstätigen in der Gesundheitswirtschaft 2006 bis 2008 – relative Abweichung zu den Berechnungswerten¹⁾

Jahr	Naive Prognose	Lokale lineare Prognose mit der folgenden Anzahl an Stützstellen		Globale quadratische Prognose		
		3	5	linear	logistisch	
		%				
2006	-1,2	-1,1	1,0	0,6	-0,8	
2007	-2,5	-2,4	0,8	0,4	-1,4	
2008	-4,1	-4,0	0,3	-0,2	-2,5	

1) Schätzung Stand April 2011; alle Angaben vorläufig.
Datenquelle: AK VGRdL; Berechnungsstand August 2010/ Februar 2011.

Literatur- und Quellenverzeichnis:

- [1] Schneider, A.: Möglichkeiten und Grenzen der Anwendung von Methoden der Zeitreihenanalyse zur Vorausberechnung von Bruttowertschöpfung und Erwerbstätigenzahl der Gesundheitswirtschaft in Sachsen, Statistisches Landesamt Sachsen, Hochschule Zittau/Görlitz, Diplomarbeit 2011.
- [2] Frie, B.; Muno, K.; Speich, W.-D.: Gesundheitswirtschaft und Wertschöpfungsansatz nach WZ 2008. In: Statistik in Sachsen, 2/2011, S. 30 - 41.