



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

Longitudinale Auswertungen

am Beispiel Lungentumore

Ausgeführt durch das
Institut für Statistik und Wahrscheinlichkeitstheorie
der Technischen Universität Wien
Wiedner Hauptstrasse 8-10
A-1040 Vienna, AUSTRIA

unter der Leitung von
Ao. Univ.-Prof. Dipl.-Ing. Dr.techn. Peter Filzmoser

Wien, am 12.09.2014

Inhaltsverzeichnis

1	Kurzfassung	2
2	Datenbasis	3
2.1	Testdatensatz	8
2.2	Bevölkerungsentwicklung	8
3	Theoretische Grundlagen	11
3.1	Chow Test	12
3.2	Adaption für SV Daten	16
3.3	Modellentwicklung	18
3.4	Praktische Vorgehensweise	22

1 Kurzfassung

Unter der Annahme, dass longitudinale Trends einem linearen Modell folgen, kann auf einheitliche Struktur vor und nach einem potentiellen Bruchpunkt mit Hilfe von F-Tests überprüft werden. Die einheitliche Struktur wird dabei als die Möglichkeit interpretiert, zwei Gruppen von Daten, meist zwei Zeitabschnitte einer Zeitreihe, mit dem gleichen Modell zu beschreiben. Ist dies nicht möglich, so muss der Grund dafür fachlich hinterfragt werden, da direkte Vergleiche der Daten zu unterschiedlichen Zeitpunkten nicht mehr zulässig sind.

Um diese Tests durchführen zu können, sind strenge Vorgaben an die Datenstruktur notwendig, die im Allgemeinen in den SV-Daten nicht erfüllt werden. Mit Hilfe einiger Anpassungen der klassischen Tests kann aber ein geeigneter Permutationstest definiert werden, der unter Zuhilfenahme der spezifischen Struktur der longitudinalen MBDS Daten der österreichischen Sozialversicherungsträger die Struktur eben dieser Daten ermöglicht.

Da diese Abschnitte nicht auf zeitliche Abschnitte der gleichen Zeitreihe begrenzt sind, ist auch ein Test zwischen unterschiedlichen Regionen oder Krankenanstalten im gleichen Zeitintervall möglich.

Neben den theoretischen Erkenntnissen über Möglichkeiten der Strukturanalyse wird durch die Auswertungen im Anhang viel datengetriebenes Wissen über das Abrechnungssystem geschaffen. Werden Statistiken miteinander verglichen, so muss berücksichtigt werden, ob zwischen den Vergleichsperioden ein Bruchpunkt liegt. Ist dies der Fall, so sollte darauf hingewiesen werden und die Natur des Bruchpunktes genauer analysiert werden. Gleichzeitig, liefern die Bruchpunkte auch eine Liste von interessanten Vergleichsperioden. Ein Beispiel, wie diese aufbereitet werden können wird in Abschnitt 3.4 gegeben.

Zwei Strukturbrüche sind hier klar hervorzuheben. Einmal tritt ein Bruch von 2001 auf 2002 in nahezu allen Zeitreihen auf. Hier hat offensichtlich eine massive Systemumstellung stattgefunden. Ein weiterer häufiger Bruch findet von 2008 auf 2009 statt.

Es gibt einen eindeutigen Strukturbruch in nahezu allen Zeitreihen von 2001

auf 2002. Im Bereich der LDF Punkte tritt zusätzlich ein weiterer kleinerer Bruch zwischen 2008 und 2010 auf. Abgesehen von diesen Brüchen bilden die LDF Punkte im Allgemeinen im Modell konsistente Zeitreihen. Zwei Untergruppen, die LDF-Punkte spezieller Bereich und die Zusatzpunkte für Mehrfachleistungen heben sich von den anderen Zeitreihen ab. Sie weisen in fast jedem Jahr einen Strukturbruch auf. Auch wenn eine klare Tendenz erkennbar ist, so dürfte das verwendete lineare Modell nicht geeignet sein um diese Zeitreihen zu analysieren.

Die Analysen der Zeitreihen der Krankenanstalten ist aufwendiger. Der Hauptzweck der Tabellen aus diesem Abschnitt ist, schnell Hypothesen überprüfen zu können, etwa dass sich eine spezialisierte Krankenanstalt signifikant anders verhält, als der regionale Durchschnitt oder der österreichische Durchschnitt. Ein Beispiel für eine Auffälligkeit ist, dass in der Zeitreihe Belagstage die bedeutendste Region, Wien K9, 2008 für Belagstage klar anders modelliert wird als Österreich gesamt (Seite 25 Anhang). Analysiert man die einzelnen Krankenanstalten, so stellt man schnell fest, dass die zwei, in diesem Bereich bedeutenden Krankenanstalten, das AKH (K901) und das Sozialmedizinisches Zentrum Baumgartner Höhe (K971) sich entsprechend des österreichischen Trends verhalten, viele kleine und mittelgroße Krankenanstalten aber klar abweichen.

Diese Analyse kann nun im dritten Abschnitt auf Seite 56 fortgesetzt werden. Die p-Werte zeigen sehr schnell, welche Krankenanstalten vergleichbar sind und welche nicht. Hier ist es aufgrund der vorherigen Ergebnisse wenig überraschend, dass die Anstalten K901 und K971 für Belagstage einen p-Wert von 0.99 aufweisen. Grundsätzlich besteht die Möglichkeit, die Anstalten in ähnliche Gruppen zusammenzufassen. Das muss aber für jede betrachtete Größe und für jedes Jahr gesondert erfolgen.

2 Datenbasis

Zur Verfügung gestellt wurden MBDS Daten von 2001 bis 2011 mit Informationen aus 4 verschiedenen Satzarten in Form von csv Dateien. Die deskriptive Beschreibung der Daten ist in den Berichten GAP-DRG 2 - Bericht Rohdaten vom 22.1. durch Florian

Endel erfolgt. Die Inhalte der Satzarten gliedern sich in

- Aufenthalte
- Verlegungen
- Diagnosen
- Leistungen

Die Informationen können entsprechend verknüpft werden und daraus auf dem jeweils gewünschten Level aggregiert werden.

Aufgrund der unterschiedlichen Datenstruktur über die Jahre wurden die Daten für jedes Jahr eingelesen, zusammengefügt, auf eine einheitliche Struktur gebracht und in transformierter, einheitlicher Form gespeichert, sodass eine konsistente Datenstruktur für die Jahre 2001 bis 2011 in Form von csv Dateien zur Verfügung steht. Zur besseren Auswertung wurden diese Basisdaten in die GAP-DRG Daten eingespielt.

Als einheitliche Basis steht folgende Struktur für weiterführende Auswertungen zur Verfügung.

- Satzart 1 - Aufenthalte

Jahr	Krankenanstalten Nummer
Aufnahmenummer	Aufnahmedatum
Aufnahmeart	Entlassungsart
Entlassungsdatum	Belagstage
Alter bei Entlassung	Geschlecht
Staatsbürgerschaft	Postleitzahl
Postleitzahl Ausland	Herkunft
Kostenträger	LDF Gruppe
LDF Knoten	LDF Tageskomponente
LDF Leistungskomponente	Dauerausreißer unten TK
Dauerausreißer unten LK	Zusatzpunkte Dauerausreißer oben
Zusatzpunkte intensiv	Zusatzpunkte Mehrfachleistungen
Zusatzpunkte spezieller Bereich	Punkte total

- Satzart 2 - Diagnosen

Jahr	Krankenanstalten Nummer
Aufnahmenummer	Diagnosetyp
ICD Subkategorie	

- Satzart 3 - Leistungen

Jahr	Krankenanstalten Nummer
Aufnahmenummer	MEL
Datum	Anzahl

- Satzart 4 - Verlegungen

Jahr	Krankenanstalten Nummer
Aufnahmenummer	Funktionscode
Subcode	Entlassungskennzeichen
Aufnahmedatum	Verlegungsdatum
Belagstage	Bezeichnung Funktionscode
Bezeichnung Fachgruppe	

Die im folgenden vorgestellte Theorie kann grundsätzlich für jede Zeitreihe angewandt werden, die gewisse Kriterien erfüllt, auf die später genauer eingegangen wird. Für Testzwecke wurden spezifische Zeitreihen erstellt. Da die Daten zwar zeitliche Informationen beinhalten, aber noch keine Zeitreihen darstellen, müssen diese erst erstellt werden. Die zu Testzwecken erstellte Zeitreihe der Daten wurde für die Variablen

- Anzahl Aufenthalte
- Anzahl der Belagstage
- LDF Tageskomponente
- LDF Leistungskomponente
- Punkte Belagsdauerausreißer unten TK
- Punkte Belagsdauerausreißer unten LK

- Zusatzpunkte Belagsdauerausreißer oben
- Zusatzpunkte intensiv
- Zusatzpunkte Mehrfachleistungen
- Punkte spezieller Bereich
- Punkte total
- Anzahl Leistungen

aufbereitet. Alle Variablen können nach Krankenanstalten, Kostenträger und Diagnosen sowie nach Art der Leistung gefiltert werden, wie aus der Konstruktion der zur Verfügung stehenden Datensätzen ersichtlich ist.

Als Basis für die Zeitreihen dienen die Intervalle jeweils von 1. bis 15. jeden Monats und vom 16. bis zum Monatsende. Jedem Intervall werden alle Aufenthalte zugeordnet, von denen mindestens ein Aufenthaltstag im betrachteten Intervall liegt. Dadurch wird jeder Aufenthalt zwischen einem und n Zeitintervallen zugeordnet. Jedem Aufenthalt sind eine eindeutige Anzahl an Belagstage, eine eindeutige Anzahl an LDF Punkten etc. zugeordnet. Mit Hilfe dieser Verknüpfung kann über jedes Zeitintervall jede gewünschte Variable aggregiert werden, entweder durch summieren oder durch eindeutiges und nicht-eindeutiges zählen. Aufgrund dieser Konstruktion ist klar, dass die betrachteten Werte, summiert über die Zeitintervalle ungleich den betrachteten Werten aggregiert über den Gesamtzeitraum sind.

Aus diesem Grund wird zusätzlich ein Faktor berechnet, der den Anteil der betrachteten Größe für das jeweilige Intervall bestimmt. Die Tabellen 1 und 2 veranschaulichen das Zählverfahren.

	id.1	aufdat	entdat	belagstage	punkte_total
9	19135796	2009-09-10	2009-12-28	108	42792

Tabelle 1: Ein beliebiger Datensatz, der über mehrere Zeitintervalle der Zeitreihe hinweggeht.

id	Lower	Upper	Anzahl	Punkte			Anteil
				Belagstage	total	Faktor	Punkte total
233	2009-09-10	2009-09-15	6	108	42792	6/108	2377.33
234	2009-09-16	2009-09-30	15	108	42792	15/108	5943.33
235	2009-10-01	2009-10-15	15	108	42792	15/108	5943.33
236	2009-10-16	2009-10-31	16	108	42792	16/108	6339.55
237	2009-11-01	2009-11-15	15	108	42792	15/108	5943.33
238	2009-11-16	2009-11-30	15	108	42792	15/108	5943.33
239	2009-12-01	2009-12-15	15	108	42792	15/108	5943.33
240	2009-12-16	2009-12-28	13	108	42792	13/108	5150.89

Tabelle 2: Diese Tabelle verbindet den Datensatz aus Tabelle 1 mit der zugrunde liegenden Zeiteinteilung der Intervalle der Zeitreihe. Es wird geprüft ob die Intervallgrenzen lower und upper eine Schnittmenge mit dem Intervall Aufnahme datum (aufdat) und Entlassungsdatum (entdat) haben. Für jedes Intervall, das eine Schnittmenge hat, wird der ursprüngliche Datensatz mit dem Intervall zusammengeführt. Die linke Hälfte der Tabelle zeigt die Zeitintervalle sowie die Anzahl der überschneidenden Tage, die rechte Hälfte den vervielfachten ursprünglichen Datensatz mit den errechneten Faktoren und den gewichteten LDF Punkten.

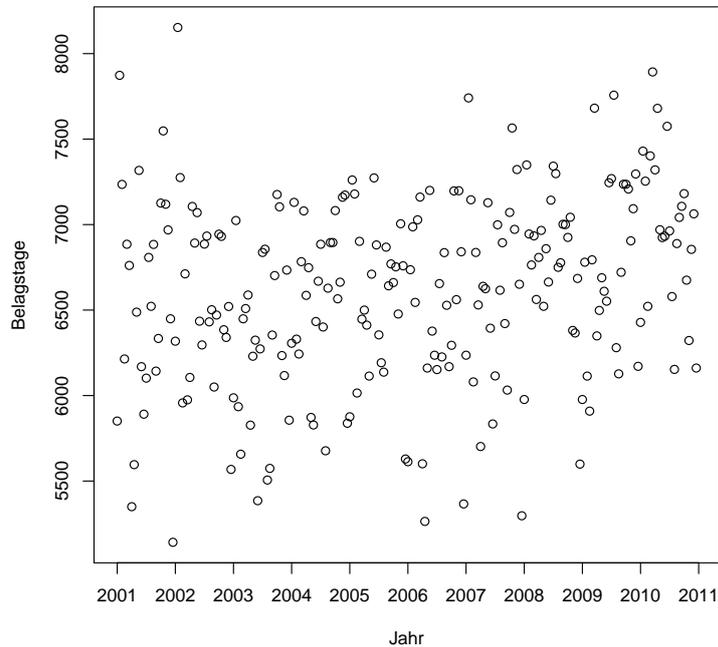


Abbildung 1: Entwicklung Belagstage der Zeitreihe Lungentumore

Jeder Datenpunkt entspricht der Summe der Belagstage aller Aufenthalte im zugehörigen Zeitintervall. Pro Jahr gibt es 24 Zeitintervalle, jeweils die erste und die zweite Hälfte jedes Monats. Einige der Zeitintervalle sind in Tabelle 2 dargestellt.

Ein wichtiger Aspekt im Rahmen der Verknüpfung der Datensätze ist, dass die Struktur der Aufenthaltsnummern ab 2006 verändert. Vor 2006 werden für ganz Österreich eindeutige Aufenthaltsnummern verwendet, ab 2006 werden in unterschiedlichen Krankenanstalten die gleichen Aufenthaltsnummern verwendet. Es ist daher wichtig, die Verbindung über Aufenthaltsnummern, Jahr und Krankenanstalt gleichzeitig herzustellen.

2.1 Testdatensatz

Über das beschriebene Verfahren können Zeitreihen verschiedenster Variablen auf unterschiedlichen Aggregationsstufen berechnet werden. Zu Testzwecken wurden unterschiedliche Variablen zum Thema Lungentumore auf Halbmonatsbasis ausgewertet. Im Rahmen der Auswertung wurden alle Aufenthalte berücksichtigt, die mit Diagnosen aus Tabelle 3 zusammenhängen.

Code	Bezeichnung
C33	Bösartige Neubildung der Trachea Bösartige Neubildung der Luftröhre
C34.-	Bösartige Neubildung der Bronchien und der Lunge
C34.0	Bösartige Neubildung: Hauptbronchus
C34.1	Bösartige Neubildung: Oberlappen (-Bronchus)
C34.2	Bösartige Neubildung: Mittellappen (-Bronchus)
C34.3	Bösartige Neubildung: Unterlappen (-Bronchus)
C34.8	Bösartige Neubildung: Bronchus und Lunge, mehrere Teilbereiche überlappend
C34.9	Bösartige Neubildung: Bronchus oder Lunge, nicht näher bezeichnet

Tabelle 3: Berücksichtigte ICD 10 Diagnosen für den Testdatensatz. Die Diagnosen wurden durch Experten festgelegt.

2.2 Bevölkerungsentwicklung

Ein Ziel des Projektes ist es, den Einfluss der Bevölkerungsentwicklung auf die Entwicklung der Zeitreihen zu messen. Dazu müssen auch Zeitreihen für die Bevölkerungsentwicklung mit den gleichen Zeitpunkten bzw. Intervallen erstellt werden.

Als Grundlage für die Zeitreihen der Bevölkerungsentwicklung werden durch den Hauptverband der österreichischen Sozialversicherungsträger Jahresdaten der Bevölkerung auf Bezirksebene von 2002 bis 2013 zur Verfügung gestellt.

Um eine für die erstellten Zeitreihen nutzbare Datenbasis zu schaffen, werden smoothed splines nach Chambers und Hastie (1992) eingesetzt. Dabei werden kubische Funktionen durch die existierende Jahresdatenpunkte gelegt und dadurch die Werte an den benötigten Zeitpunkten, insbesondere für das fehlende Jahr 2001 geschätzt.

Grundsätzlich können durch Adaptionen des angewandten statistischen Tests beliebig viele beschreibende Variablen verwendet werden. Auf diese Thematik wird später genauer eingegangen. Es ist jedoch sinnvoll, nur eine begrenzte Anzahl an Altersvariablen zu verwenden. Als Test wurden 6 Zeitreihen mit jeweils männlicher Bevölkerung unter 50, zwischen 50 und 80 und über 80 verwendet. Tabelle 4 stellt die Korrelationen zwischen den beschreibenden Zeitreihen dar. Wir sehen, dass die Bevölkerung über 50, egal ob männlich oder weiblich stark untereinander und auch mit der zeitlichen Entwicklung korreliert. Im Allgemeinen korreliert auch die männliche mit der weiblichen Bevölkerung sehr stark.

Das hat zwei Konsequenzen. Erstens wird durch eine höhere Granularität, der Bevölkerungsstruktur, keine wesentlich bessere Beschreibung der Daten gelingen können. Die Bevölkerungsgruppe $m3$ etwa könnte völlig vernachlässigt werden, ohne Informationen im linearen Modell zu verlieren.

Zweitens ist es nicht möglich, Änderungen in der Struktur auf eindeutige Bevölkerungsgruppen zurückzuführen. Auch wenn das Einbeziehen der Bevölkerungsentwicklung in ein Modell die Qualität der Beschreibung wesentlich verbessert, so kann nur die Aussage getroffen werden, dass die Bevölkerungsentwicklung für eine Änderung verantwortlich ist, nicht welcher Teil der Bevölkerung, da die unterschiedlichen Bevölkerungsgruppen annähernd gleich viel Information in das Modell einbringen. Wiederum kann man am Beispiel $m3$ klar sehen, dass wir nie unterscheiden können, ob $m3$ oder $m2$ für die bessere Modellbeschreibung verantwortlich sind.

	jahr	monat	m1	m2	m3	w1	w2	w3
jahr	1.00	0.00	-0.31	0.99	0.99	-0.04	0.93	0.98
monat	0.00	1.00	-0.06	0.10	0.10	-0.02	0.09	0.09
m1	-0.31	-0.06	1.00	-0.42	-0.39	0.96	-0.61	-0.13
m2	0.99	0.10	-0.42	1.00	1.00	-0.16	0.97	0.95
m3	0.99	0.10	-0.39	1.00	1.00	-0.12	0.96	0.96
w1	-0.04	-0.02	0.96	-0.16	-0.12	1.00	-0.38	0.15
w2	0.93	0.09	-0.61	0.97	0.96	-0.38	1.00	0.86
w3	0.98	0.09	-0.13	0.95	0.96	0.15	0.86	1.00

Tabelle 4: Korrelation zwischen den beschreibenden Größen. m1 bis w3 stehen für männliche und weibliche Altersgruppen die variabel gewählt werden können. Diese spezielle Korrelationen basieren auf den Altersgruppen unter 50, 51 bis 80 und über 80.

Das größte Problem aber ist, dass die Entwicklung der Bevölkerungsgruppen, in manchen Untergruppen einen starken linearen Trend aufweist. Dadurch ergibt sich auch ein starker linearer Zusammenhang mit der zeitlichen Entwicklung und damit mit den Jahren. Abbildung 2 stellt diesen Zusammenhang für die Gruppe der männlichen Bevölkerung zwischen 50 und 80 Jahren dar. Die Korrelation zwischen den Jahren und der Bevölkerung beträgt für diesen extremen Fall sogar 0.99. Daher sollte die Bevölkerungsstruktur in unseren linearen Modellen vernachlässigt werden. In Abschnitt 3.3 wird das Thema und eine zentrale Problematik des Inkludierens der Altersstruktur ausführlich diskutiert.

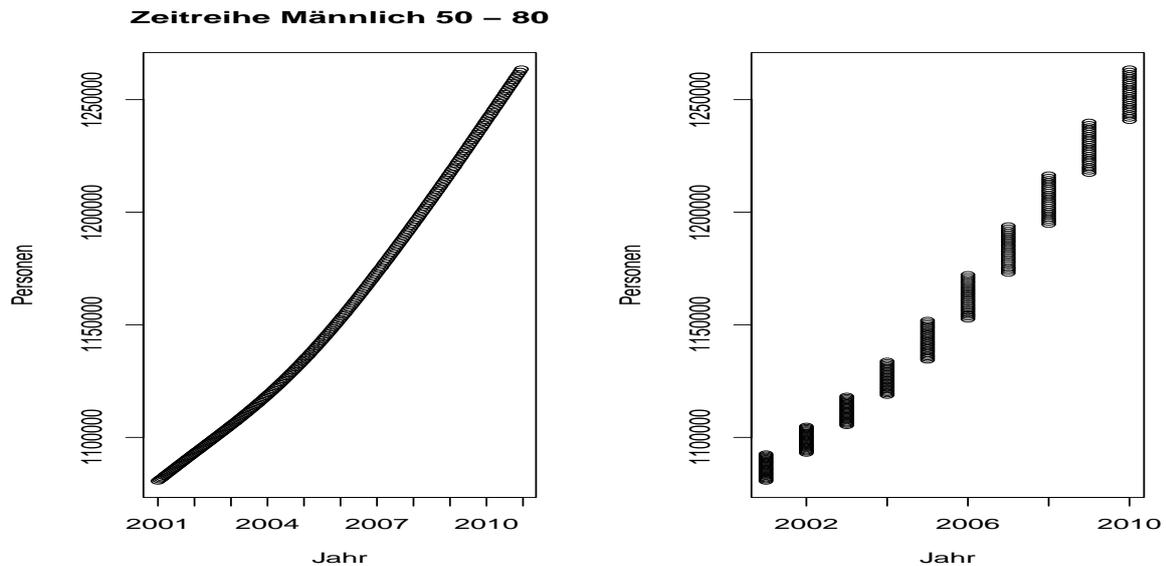


Abbildung 2: Zusammenhang Bevölkerungsentwicklung und Zeit. Die Anzahl der Personen folgen stark einem linearen Trend. Links wird die Anzahl der Personen mit einem Index, der aus Jahr und Monat besteht verglichen, rechts rein mit den Jahren. Dadurch werden rechts jeweils 12 Werte pro Jahr dargestellt.

3 Theoretische Grundlagen

Im Allgemeinen wird ein Strukturbruch im statistischen Sinn als eine signifikante Änderung der Koeffizienten im beschreibenden Modell betrachtet. Wir gehen davon aus, dass eine lineare Beschreibung der Zeitreihe grundsätzlich möglich ist. Die Nullhypothese lautet nun immer, dass die Daten vor und nach einem potentiellen Bruchpunkt mit dem gleichen Modell beschrieben werden können. Diese Hypothese wird nicht getestet, indem die Koeffizienten direkt verglichen werden sondern die Verteilung der Residuen analysiert wird.

Zusätzlich wird ein weiterer Test durchgeführt, der auf Heteroskedastizität testet. Warum dieser zweite Test notwendig ist, wird anschließend an die Einführung des Strukturtests erklärt.

3.1 Chow Test

In Chow (1960) wird der im Allgemeinen als Chow-Test bekannte Test eingeführt. Wir gehen von einem linearen Regressionsmodell $y = Xb_0 + u$ aus und schätzen die Koeffizienten b_0 für unsere Daten über das klassische Modell. Zusätzlich trennen wir unsere Daten X in X_1 und X_2 , y in y_1 und y_2 , die für Daten vor und nach einem potentiellen Bruchpunkt stehen, und schätzen für diese Daten wiederum die Koeffizienten b_1 und b_2 .

Die Nullhypothese nimmt an, dass $b_0 = b_1 = b_2$ gilt, d.h. dass alle drei Regressionsmodell gleich beschrieben werden. Unter dieser Nullhypothese kann gezeigt werden, dass die Chow Statistik zentral F-verteilt ist.

$$F_C = \frac{\|X_1 b_1 - X_1 b_0\|^2 + \|X_2 b_2 - X_2 b_0\|^2}{\|y_1 - X_1 b_1\|^2 + \|y_2 - X_2 b_2\|^2} \frac{(m + n - 2p)}{p} \sim F_{p, m+n-2p} \quad (1)$$

Dabei werden einige Annahmen benötigt, um diese zentrale F-Verteilung der Teststatistik zu ermöglichen. Es wird davon ausgegangen, dass die Anzahl der Variablen des Modells gleich p mit $p < m$ und $p < n$ gilt. n und m sind die Anzahl der Beobachtungen in X_1 und X_2 . Um die zentrale F-Verteilung zu gewährleisten, benötigen wir zusätzlich unabhängige gleich normalverteilte Fehler unserer Regressionsmodelle. Auf dieses Problem wird später genauer eingegangen. Auch wird im Rahmen des Chow Tests davon ausgegangen, dass der potentielle Bruchpunkt, der X in X_1 und X_2 trennt, bekannt ist.

Ist der Bruchpunkt a-priori nicht bekannt, muss eine Statistik mit variablem Bruchpunkt gefunden werden. Das Problem ist, dass der Bruchpunkt nur auf die Alternativhypothese Einfluss hat. Eine Problematik die in Davies (1977) diskutiert wird.

Ein direkter Ansatz, mehrere mögliche Bruchpunkte zu identifizieren, ist für jeden möglichen Bruchpunkt die F_C Statistik zu berechnen und diese zu aggregieren. In dem R-Paket `strucchange`, eingeführt in Zeileis, A. and Leisch, F. and Hornik, K. and Kleiber, C. (2001), basierend auf Andrews und Ploberger (1994), werden diverse Möglichkeiten der Aggregation eingeführt. Eine häufig verwendete Statistik wird über

das Supremum der einzelnen Chow Statistiken bestimmt.

$$\text{sup}F = \sup_{k < i < n-k} F_{C,i} \quad (2)$$

Diese Statistik ist offensichtlich nicht mehr F verteilt. Sie stammt aber aus einer Klasse von Statistiken, welche aus Aggregationen von F Statistiken besteht und für die Untersuchungen von Strukturbrüchen von zentraler Bedeutung ist. In Andrews, 1993 wird diese Klasse der Verteilungsfunktionen genauer untersucht. Die asymptotischen Ergebnisse ermöglichen die Bestimmung der kritischen Werte für diese aggregierten Statistiken von nicht unabhängigen stochastischen Größen. Im Rahmen des R-Paketes werden diese aus Hansen (1997), wo die kritischen Werte über eine Simulation, basierend auf eben diesen asymptotischen Ergebnissen bestimmt werden, verwendet.

Um die Ergebnisse aus Hansen (1997) verwenden zu können sind die gleichen Voraussetzungen notwendig um die Verteilung der entsprechenden Statistiken zu gewährleisten, bzw. um diese berechnen zu können. Werden die Voraussetzungen nicht erfüllt, so sind die einzelnen Statistiken nicht mehr zentral F verteilt. Damit ist die Aggregation kein Element der betrachteten Familie von Verteilungsfunktionen.

Zusätzlich zu den Annahmen von Gleichung (1) sehen wir, dass die Modelle immer vollen Rang haben müssen, da sonst die Dimensionen der Koeffizienten nicht mehr übereinstimmen und die Chow Statistik nicht mehr wohldefiniert ist.

Aufgrund dieser Annahmen kann weder die supF Statistik, noch die Chow-Statistik selbst direkt verwendet werden um Bruchpunkte in den SV Daten zu bestimmen. Der Zusammenhang in den SV Daten ist kein linearer. Dadurch sind die Residuen des linearen Regressionsmodells auch nicht unabhängig. In kurzen Zeitreihen fällt dies nicht auf, nach wenigen Jahren sind die Effekte in den Residuen der linearen Modelle jedoch groß genug um die Nullhypothese zu verwerfen. Der Zähler und der Nenner aus Gleichung (1) sind daher aus den oben genannten Gründen keine unabhängigen χ^2 Statistiken und daher folgt F_C keiner zentralen F-Verteilung.

Darüber hinaus, wird nach Bruchpunkten im Laufe von 10 Jahren gesucht. Dabei ist nicht bekannt, welches Datum genau als Bruchpunkt untersucht werden soll, bzw. ob ein Bruchpunkt existiert und ob dieser eindeutig ist. Die supF Statistik bietet hier zwar

klare Vorteile hinsichtlich der nicht bekannten Bruchpunkte, aufgrund des Argumentes bezüglich der nicht-zentralen F-Verteilung von $F_{C,i}$ kann das angewandte Verfahren aber nicht direkt eingesetzt werden. Die Simulation von Hansen (1997) beruht auf den Eigenschaften der kombinierten F-Verteilungen und simuliert nicht für jede die kombinierten, abhängigen gemeinsamen Statistiken die Verteilungsfunktion neu. Auch kann der Algorithmus von Hansen (1997) nicht entsprechend adaptiert werden da die asymptotischen Ergebnis aus Andrews (1993) nicht anwendbar sind.

Auch wenn die Verteilungen der Chow Statistik und der supF Statistik nicht explizit bestimmt werden können, so können zumindest die kritischen Werte für spezifische supF Statistiken und Chow Statistiken über Permutationstests approximiert werden. Dadurch wird gleichzeitig ein Vergleich zwischen den theoretischen und den approximierten Quantilen ermöglicht.

Ein Vergleich der theoretischen kritischen Werte der Chow Statistik, der dem 95% Quantil der entsprechenden F-Verteilung entspricht, und der über den Permutationstest bestimmten kritischen Werte wird für eine Zeitreihe, in der die LDF Tageskomponente sowohl durch die zeitliche Entwicklung, als auch durch die Bevölkerungsentwicklung beschrieben wird, in den Abbildungen 4 und 5 für die Chow Statistik dargestellt. Für die supF Statistik wird in Hansen (1997) keine Tabelle kritischer Werte angegeben. Der kritische Wert wurde daher auf Basis der Zeitreihen simuliert indem ein Wert approximiert wurde, der möglichst nahe einer Statistik liegt, die mit einem p-Wert von 0.05 ausgewiesen wird. Da der Simulationsaufwand für die supF Statistik um ein vielfaches über dem Aufwand der Chow Statistik liegt wurde eine entsprechend kleinere Permutationsanzahl gewählt. Es ist zu beachten, dass die Simulation ausschließlich auf die Unterschiede zwischen den theoretischen kritischen Werten und den über Simulation bestimmten Quantilen hinweisen soll. Der kritische Wert hängt stark von der Anzahl der Freiheitsgrade bzw. vom gewählten Modell und der Anzahl der Beobachtungen, sowie von der Breite des Zeitfensters in dem der Bruchpunkt liegen kann ab. Ein weiterer wesentlicher Einflussfaktor ist der Grad der Beschreibung den das Modell erreicht. Für die Simulation wurde eine Breite von 8 Zeitpunkten gewählt.

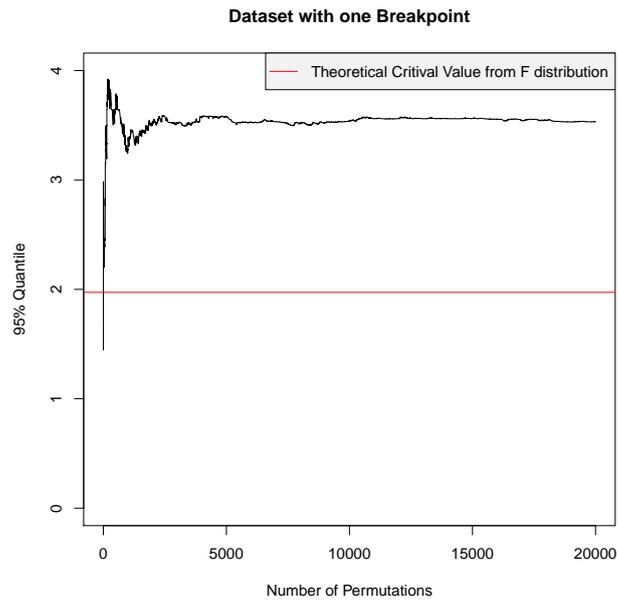


Abbildung 3: Kritischer Wert der Chow Statistik gegen simulierter kritischer Wert. Verglichen werden der kritische Wert der F Statistik der im klassischen Chow Test verwendet wird mit dem über Permutationstest simulierten Wert für den Fall eines einzelnen Bruchpunktes.

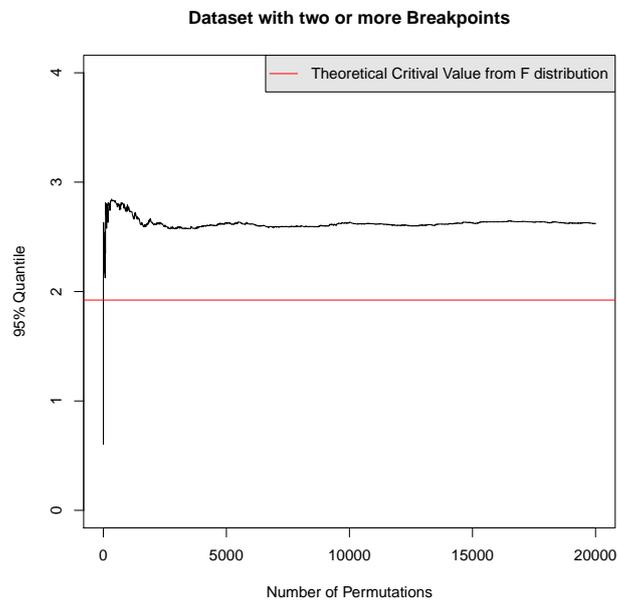


Abbildung 4: Kritischer Wert der Chow Statistik gegen simulierter kritischer Wert. Verglichen werden der kritische Wert der F Statistik der im klassischen Chow Test verwendet wird mit dem über Permutationstest simulierten Wert für den Fall von zwei oder mehr Bruchpunkten.

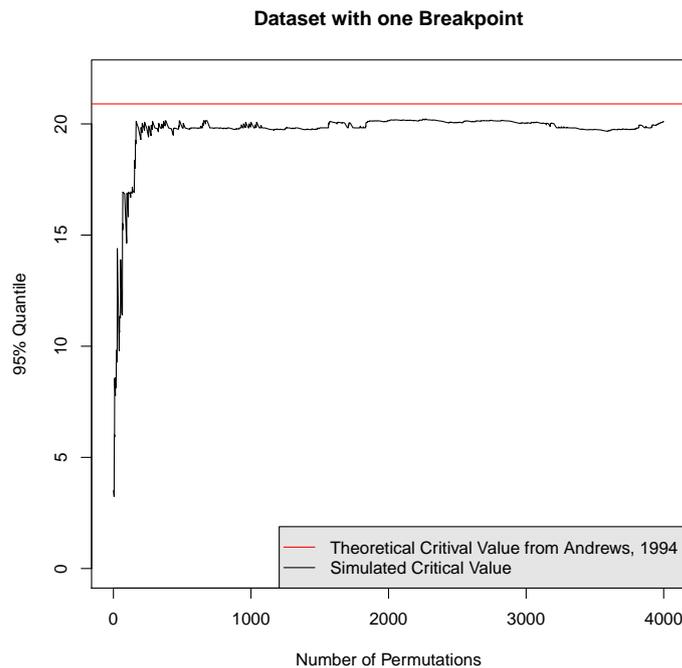


Abbildung 5: Kritischer Wert der supF Statistik gegen simulierter kritischer Wert. Verglichen werden der kritische Wert der supF Statistik aus Andrews, 1994 mit dem über Permutationstest simulierten Wert für den Fall eines eindeutigen Bruchpunktes

3.2 Adaption für SV Daten

Wir haben zwei zentrale Probleme. Da die Abhängigkeit der Residuen nicht bekannt ist, ist es sinnvoll, die kritischen Werte der unbekanntem Verteilungsfunktion direkt mit Hilfe eines Permutationstests zu approximieren. Andererseits müssen wir auch von bekannten Bruchpunkten ausgehen, da es zu rechenintensiv ist die supF Statistik für Intervalle um jeden möglichen Bruchpunkt zu simulieren.

Um das Problem der unbekanntem Bruchpunkte zu beseitigen, gehen wir davon aus, dass unterjährig keine Strukturbrüche auftreten können. Nur mit dem Jahreswechsel kann eine Änderung eintreten. Gleichzeitig testen wir nur auf Strukturbrüche innerhalb von 2 Jahren. Damit wird gleichzeitig ein weiteres Probleme beseitigt.

Es wurde bereits auf den nicht exakt linearen Zusammenhang der Daten hingewiesen. Daher wird nach einer gewissen Zeit auch immer ein Strukturbruch gefunden werden, da die Beschreibung über das lineare Modell nicht mehr ausreicht. Durch das Beschreiben der Daten in jeweils zwei Jahren entsteht ein stückweise lineares Modell.

Wir erhalten eine gewisse Toleranz und umgehen damit Probleme, die aus den nicht exakten Modellannahmen heraus entstehen. Einerseits das Phänomen, dass an den Rändern von zu langen Zeitreihen Bruchpunkte zu finden sind. Andererseits könnten nicht-eindeutige Bruchpunkte gefunden werden. Angenommen, in einer 4 periodigen Zeitreihe, t_1, \dots, t_4 werden möglichst große Intervalle gesucht, in denen kein Bruchpunkt auftritt.

Für die Intervalle $[t_1, t_3]$ und $[t_2, t_4]$ jeweils kein Strukturbruch erkannt. Innerhalb von $[t_1, t_4]$ wird aber ein Bruchpunkt ausgewiesen. Dadurch ist nicht mehr klar, ob der Strukturbruch in t_2 oder t_3 auftritt. Mit Hilfe einer Maximum Likelihood Schätzung könnte nun der wahrscheinlichste Bruchpunkt ermittelt werden. Dadurch würde ein Punkt fälschlicherweise als Bruchpunkt ausgewiesen werden, denn der Grund für solche Phänomene ist wiederum die nicht exakte Modellbeschreibung. Durch die Betrachtung von 2 Jahren, wird in diesem Beispiel sowohl der Strukturbruch zum Zeitpunkt t_2 , als auch zum Zeitpunkt t_3 abgelehnt. Somit werden kleine nicht-lineare Änderungen über lange Zeiträume akzeptiert, solange diese lokal gut genug linear approximiert werden können.

Auf das zweite große Problem wurde im Rahmen der Bedingungen für die Chow Statistik in Gleichung (1) hingewiesen. Es wird davon ausgegangen, dass alle Modelle vollen Rang haben. In unserem Fall haben nie alle Modelle vollen Rang. Da die Modelle $y_1 = X_1 b_1 + \epsilon_1$ und $y_2 = X_2 b_2 + \epsilon_2$ jeweils nur Daten aus einem Jahr umfassen, spielt die beschreibende Variable Jahr keine Rolle und wird in den Intercept aufgenommen. Gleichzeitig ist der entsprechende Koeffizient im Modell $y_0 = X_0 b_0 + \epsilon_0$ im Allgemeinen ungleich null. Zusätzlich kann der volle Rang aufgrund der starken Abhängigkeiten in der Altersstruktur nicht gewährleistet werden. Tabelle 4 gibt ein Beispiel für die Ursache des Auftretens der Singularitäten in den linearen Modellen.

Die Dimensionen von b_0 , b_1 und b_2 aus Gleichung (1) stimmen also im Allgemeinen nicht überein. Der Zähler von F_C lässt sich aber als

$$\left\| \begin{array}{c} X_1 b_1 - X_1 b_0 \\ X_2 b_2 - X_2 b_0 \end{array} \right\|^2 = \left\| \begin{array}{c} y_1 - X_1 b_0 \\ y_2 - X_2 b_0 \end{array} \right\|^2 - \left\| \begin{array}{c} y_1 - X_1 b_1 \\ y_2 - X_2 b_2 \end{array} \right\|^2 \quad (3)$$

schreiben. Da die rechte Seite von Gleichung (3) der Differenz der Summe der quadratischen Residuen des gemeinsamen Modells und der getrennten Modelle entspricht, kann eine repräsentative Statistik leichter errechnet werden. b_0 wird für jede beliebige Permutation gleich geschätzt. Also bleibt die Summe der quadratischen Fehler konstant. Daher muss nur

$$\|y_1 - X_1 b_1\|^2 + \|y_2 - X_2 b_2\|^2 \quad (4)$$

bestimmt werden und daraus die Ausreißer im Sinne der 95% Quantile bestimmt werden. Zu beachten ist, dass die Ausreißer nach unten gesucht werden, wie leicht aus dem negativen Vorzeichen des letzten Anteils der quadratischen Summen in Gleichung (3) nachvollzogen werden kann. Um bessere Interpretierbarkeit und Vergleichbarkeit der Statistiken zu gewährleisten betrachten wir nicht direkt die Summe sondern den Anteil an der Summe der quadratischen Fehler im allgemeinen Modell.

$$\frac{\|y_1 - X_1 b_1\|^2 + \|y_2 - X_2 b_2\|^2}{\|y - X b_0\|^2} \quad (5)$$

Das Problem der linearen Abhängigkeit der beschreibenden Größen, insbesondere der Altersstruktur, kann mit Hilfe von Maximum Likelihood Schätzungen für die Regressionskoeffizienten oder mit alternativen Modellen wie der PLS Regression (Wold, 1966) gelöst werden. Für unsere konkrete Situation wird dies jedoch nicht notwendig sein.

3.3 Modellentwicklung

Neben der Altersstruktur steht uns die id, das Monat und das Jahr der Zeitpunkte als beschreibende Variablen zur Verfügung. Um ein möglichst geeignetes Modell zu finden, untersuchen wir die Zeitreihe der LDF-Intensivpunkte genauer.

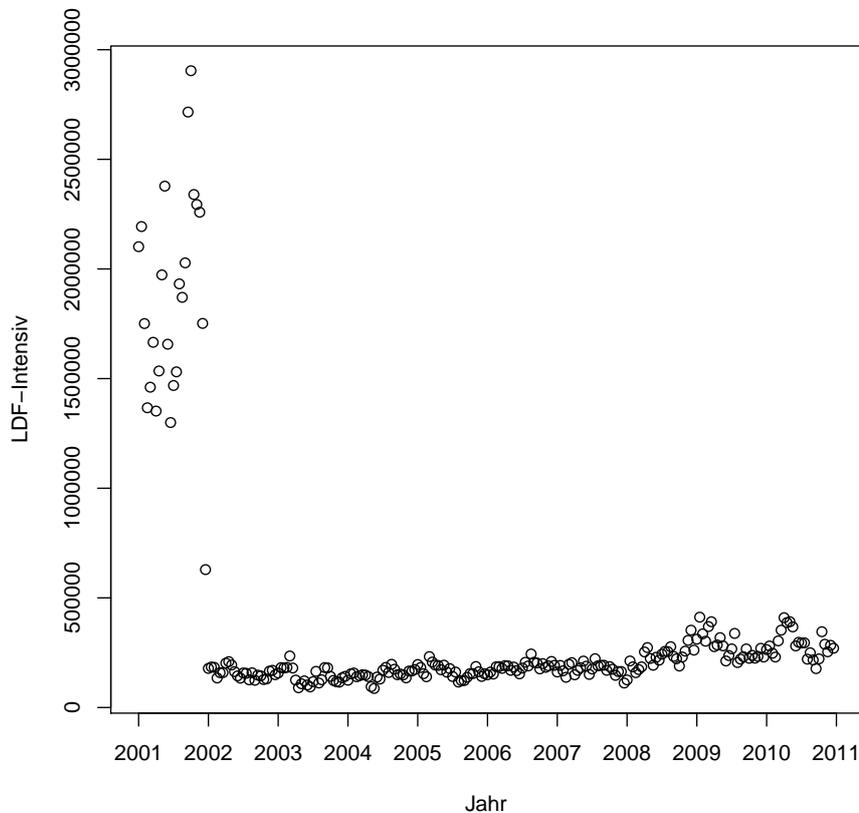


Abbildung 6: Bruchpunkte der Zeitreihe LDF Zusatzpunkte intensiv. Jeder Datenpunkt entspricht der Summe der LDF Zusatzpunkte intensiv aller Aufenthalte im zugehörigen Zeitintervall. Pro Jahr gibt es 24 Zeitintervalle, jeweils die erste und die zweite Hälfte jedes Monats.

Wir sehen offensichtlich unterschiedliche Modelle, die sicher nicht rein über die id sinnvoll beschrieben werden können. Beschreiben wir diese Zeitreihe rein über Jahr und Monat so erhalten wir immer noch ein sehr schlechtes Modell. Da wir aber ein stückweise lineares Modell verwenden und nur zwei Jahre gleichzeitig betrachten kann über das Modell

$$LDF_{intensiv} \sim Intercept + Jahr + Monat \tag{6}$$

jeden beliebigen Sprung in den Daten modelliert werden. Sehr gut zu sehen ist dies an den ersten zwei Jahren der Zeitreihe in Abbildung 7. In rot wird das lineare Modell aus Jahr und Monat zusammengesetzt, in blau wird die Altersstruktur inkludiert.

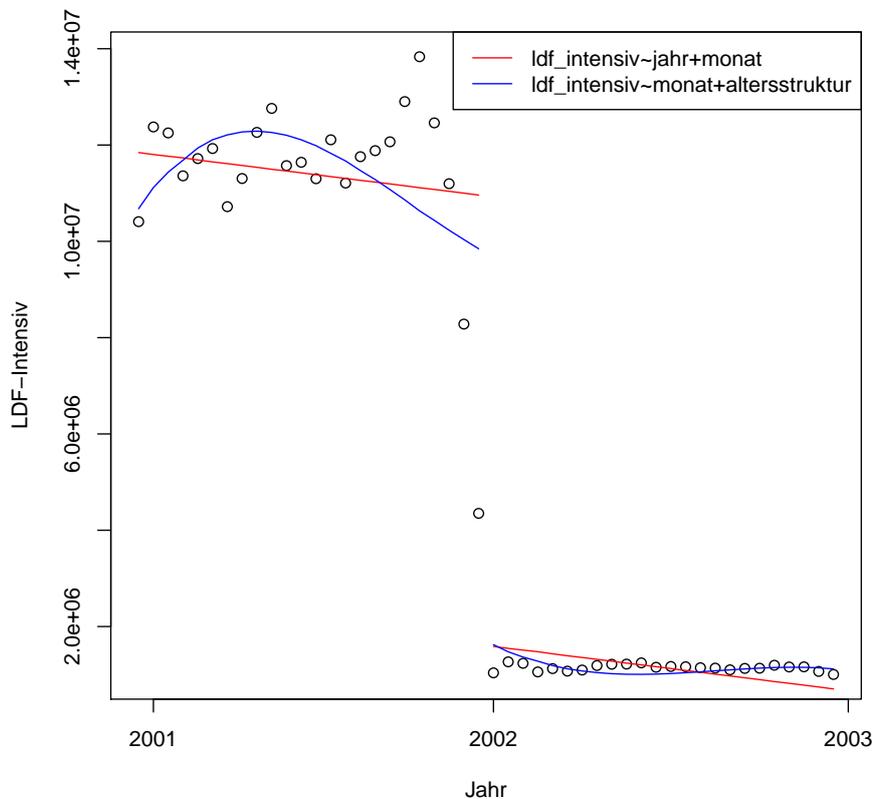


Abbildung 7: Lineare Beschreibung der Zeitreihe LDF-Intensiv. Die Beschreibung erfolgt in einem gemeinsamen Modell für beide Jahre. Die Beschreibung unter Berücksichtigung der Altersstruktur in blau inkludiert beinhaltet das Jahr nicht.

Will man auf Strukturbrüchen zwischen Jahren testen, darf das Jahr selbst in Zeitreihen, die nur über zwei Jahre gehen nicht in das Modell einbezogen werden. An dieser Stelle muss auf die hohen Korrelationen zwischen den Variablen der Altersstruktur aus Tabelle 4 hingewiesen werden. Insgesamt kann 99% der Variabilität der Variable Jahr linear aus der Altersstruktur beschrieben werden. Dadurch ist es für das Modell irrelevant ob die Altersstruktur oder das Jahr in das Modell aufgenommen werden. In keinem der beiden Fälle kann der Strukturbruch mehr erkannt werden, da er mit modelliert wird.

Das genaue Gegenteil ist das Fall, wenn das Modell zu genau beschrieben wird. Eben das passiert, wenn die Altersstruktur in das Modell miteinbezogen wird ohne dass große Änderungen wie der Sprung im ersten Abschnitt der Zeitreihe auftreten. Das

Modell wird dabei sensitiv für kleine Änderungen in den Koeffizienten. Das Beispiel aus Abschnitt 3.4 veranschaulicht diese Problematik.

Ein weiteres Thema, das im ersten Abschnitt der Zeitreihe klar ersichtlich wird ist das Problem der Heteroskedastizität. Auch für den Permutationstest ist die Homoskedastizität erforderlich. Insbesondere große Unterschiede wie sie in diesem Fall auftreten können sonst die Strukturbrüche verschleiern. Ein fiktives Beispiel wird in Abbildung 8 dargestellt. Links wird ein Sprung und eine Verdoppelung des Regressionskoeffizienten simuliert mit jeweils dem gleichen normalverteilten Fehler. Rechts wird der zweite Fehler so weit vergrößert, dass das Modell keine adäquate Beschreibung der Variablen mehr darstellt. In diesen Fällen versagt der Test auf Strukturbruch. Aus eben diesem Grund wird gleichzeitig ein Test auf Homoskedastizität durchgeführt.

Es werden also einmal die Strukturbrüche in Form von Änderungen der Koeffizienten des Regressionsmodells ausgewiesen und einmal die Strukturbrüche in Form von Änderungen der Fehlerverteilung. Tabelle 5 stellt das Testergebnis für diese Zeitreihe dar.

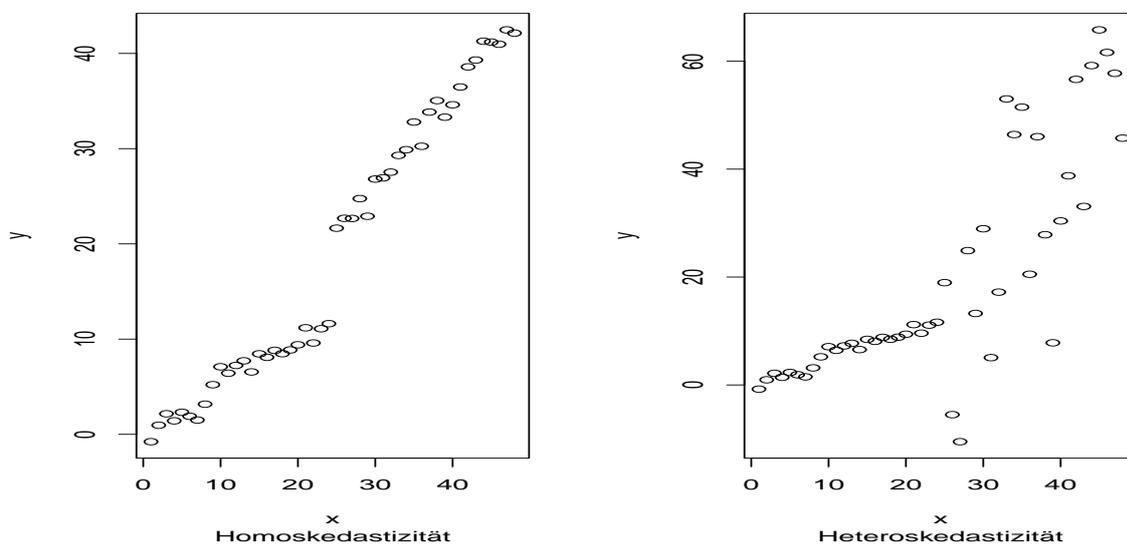


Abbildung 8: Es werden zwei verschiedene Arten von fiktiven Bruchpunkten dargestellt. Links bleibt die Fehlerverteilung konstant, rechts ändert sich die Varianz der Fehler mit dem Bruchpunkt.

	Bruchpunkt			Heteroskedasitizät		
	criticalVal	pValue	Stat	critVal	pValue	Stat
Model 1	88.10	0.00	10.05	1.57	0.21	1.32
Model 2	88.26	0.10	91.39	1.59	0.00	15.10

Tabelle 5: Vergleich Modell 1 und Modell 2 aus Abbildung 7

3.4 Praktische Vorgehensweise

Ein wesentlicher Vorteil der eingeführten Tests ist, dass die Voraussetzung der Unabhängigkeit der Fehler sowie das Problem linearer Abhängigkeiten nicht mehr notwendig sind. Dadurch können beliebige Datensätze auf gemeinsame Struktur getestet werden.

Zunächst werden die Daten jedes Jahres gegen die Daten der jeweils angrenzenden Jahre wie in klassischen Tests auf Strukturbrüche getestet. Das Ergebnis wird in Tabelle 6 und Abbildung 9 dargestellt. Die Strukturbrüche werden in roten vertikalen Linien dargestellt, die Änderungen in der Verteilung der Fehler in strichlierten blauen vertikalen Linien.

Um das Problem der optimalen Modellwahl besser zu verstehen, wird in Abbildung 10 und Tabelle 7 die gleiche Zeitreihe für das Modell LDF-Belagsdauererisierer $\text{oben} \sim \text{Monat} + \text{Alterstruktur}$ ausgewertet. In der Zeitreihe sind zwei Strukturbrüche in Form von Sprüngen gut zu erkennen. Von 2001 auf 2002 ein offensichtlicher extremer Sprung und von 2008 auf 2009 ein kleiner Sprung, bei dem sich auch die Varianz verändert.

Aus der zweiten Abbildung ist gut zu erkennen, dass die Alterstruktur alle Sprünge problemlos modellieren kann, wie in Abschnitt 3.3 beschrieben. Übrig bleiben nur die unterschiedliche Fehlervarianz, die teilweise weiterhin erkannt wird. Durch die Modelländerung kann es allerdings auch leicht passieren, dass sich die Fehlerverteilung so verändert, dass die Änderungen in der Varianz ebenfalls nicht mehr erkannt werden können, wie es auch hier bei der Änderung im Jahr 2008 der Fall ist.

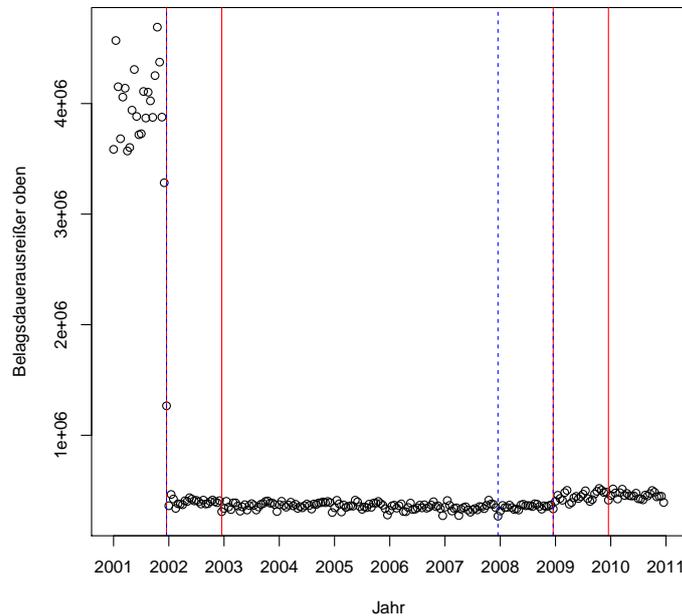


Abbildung 9: Bruchpunkte der Zeitreihe Belagsdauerausreißer oben nach dem Modell $BDA\text{-}oben \sim \text{Monat}$.

Die Punkte der Zeitreihe entsprechen der Summe bzw. der Anzahl der betrachteten Variable jeweils eines halben Monats. Die Zeitreihe jedes Jahres wird mit der Zeitreihe des Folgejahres mit Hilfe eines linearen Modells verglichen. Die Brüche werden durch vertikale Linien in rot für Änderungen in den Koeffizienten der linearen Beschreibung und in blau für Änderungen in der Fehlervarianz dargestellt.

Jahr 1	Jahr 2	Bruch Statistik	Kritischer Wert	p-Wert		Varianz Statistik	Kritischer Wert	p-Wert	
2001	2002	6.05	87.08	0.00	***	20.22	1.10	0.00	***
2002	2003	75.93	88.98	0.00	***	1.12	1.55	0.61	
2003	2004	98.13	87.79	0.66		1.15	1.58	0.54	
2004	2005	97.11	86.32	0.53		1.30	1.68	0.37	
2005	2006	94.32	87.46	0.28		1.14	1.64	0.58	
2006	2007	97.63	87.60	0.59		1.22	1.72	0.45	
2007	2008	93.77	88.64	0.21		1.87	1.83	0.04	**
2008	2009	26.05	84.60	0.00	***	2.08	1.36	0.00	***
2009	2010	91.18	86.98	0.14		1.27	1.43	0.19	

Tabelle 6: Tests auf Strukturbrüche und Änderungen in der Varianz von 2001 bis 2010 der Zeitreihe Belagsdauerausreißer oben. Verglichen werden jeweils die Zeitreihen aus zwei Jahren. Einmal wird über ein lineares Modell getestet, ob die Koeffizienten der Beschreibung konstant bleiben. Als zweites wird getestet, ob die Varianz der Fehler konstant bleibt. Neben dem p-Werten der Tests wird ein zugehöriges Signifikanzlevel angegeben.

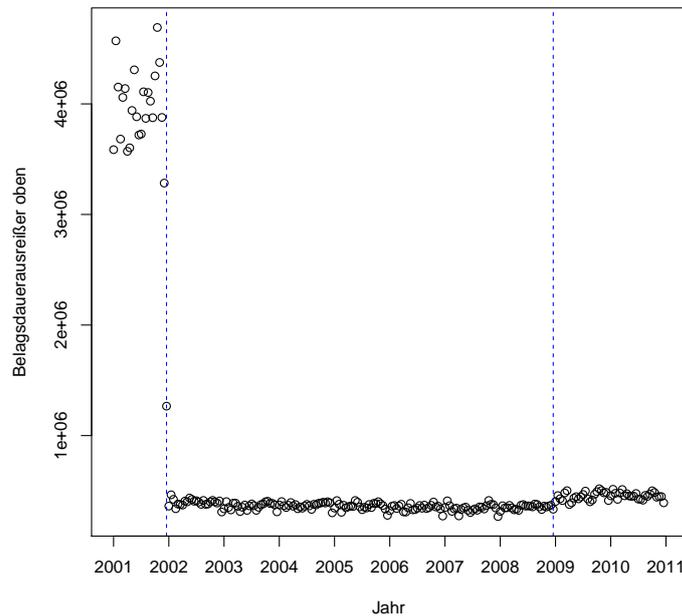


Abbildung 10: Bruchpunkte der Zeitreihe LDF-Intensiv nach dem Modell LDF-Intensiv~Monat+Altersstruktur.

Die Punkte der Zeitreihe entsprechen der Summe bzw. der Anzahl der betrachteten Variable jeweils eines halben Monats. Die Zeitreihe jedes Jahres wird mit der Zeitreihe des Folgejahres mit Hilfe eines linearen Modells verglichen. Die Brüche werden durch vertikale Linien in rot für Änderungen in den Koeffizienten der linearen Beschreibung und in blau für Änderungen in der Fehlervarianz dargestellt.

Jahr 1	Jahr 2	Bruch Statistik	Kritischer Wert	p-Wert		Varianz Statistik	Kritischer Wert	p-Wert	
2001	2002	58.73	55.35	0.08	.	0.00	4.45	0.00	***
2002	2003	90.21	63.96	0.92		0.00	1.71	0.79	
2003	2004	85.13	65.15	0.81		0.00	1.80	0.53	
2004	2005	86.13	65.70	0.75		0.00	1.84	0.44	
2005	2006	84.43	67.63	0.63		0.00	1.65	0.65	
2006	2007	73.40	64.63	0.25		0.00	1.86	0.75	
2007	2008	77.73	65.44	0.32		0.00	2.22	0.22	
2008	2009	97.49	69.49	1.00		0.00	1.70	0.00	***
2009	2010	99.73	74.53	1.00		0.00	1.48	0.17	

Tabelle 7: Tests auf Strukturbrüche und Änderungen in der Varianz von 2001 bis 2010 der Zeitreihe Belagsdauerausreißer oben. Verglichen werden jeweils die Zeitreihen aus zwei Jahren. Einmal wird über ein lineares Modell getestet, ob die Koeffizienten der Beschreibung konstant bleiben. Als zweites wird getestet, ob die Varianz der Fehler konstant bleibt. Neben dem p-Werten der Tests wird ein zugehöriges Signifikanzlevel angegeben.

Um die Unterschiede zwischen den jeweils betrachteten Jahren besser verstehen zu können, wird die Verteilung der betrachteten Statistik auf die häufigsten Diagnosen zur Verfügung gestellt. Abbildung 11 bis 13 stellen diese Informationen für den Bruchpunkt 2008 auf 2009 der Zeitreihe Belagsdauerausreißer oben, die in Abbildung 9 abgebildet ist, dar. Die Auswahl der betrachteten Diagnosen erfolgt in der Reihenfolge der Häufigkeit. In diesem Fall werden die 6 häufigsten Diagnosen dargestellt.

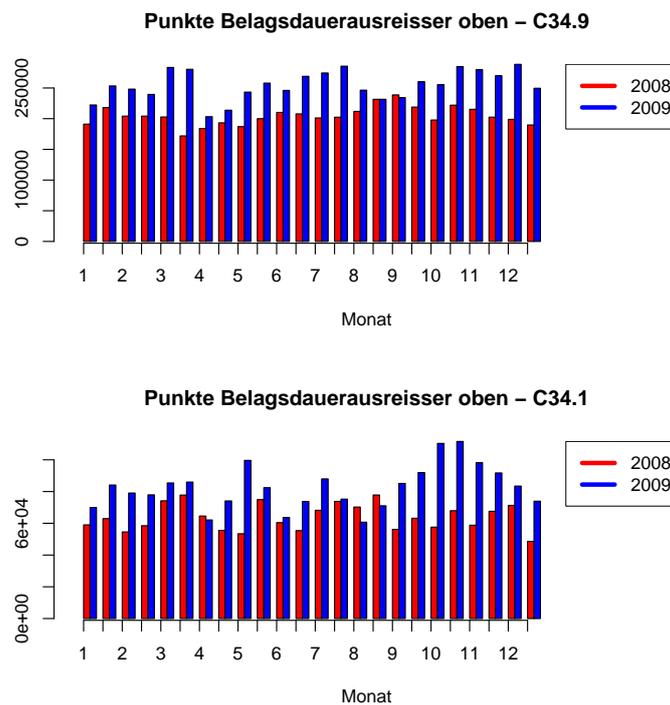


Abbildung 11: Entwicklung der Zeitreihe Belagsdauerausreißer oben aufgeschlüsselt nach Diagnosen. Die Balken vergleichen direkt die Anzahl der Punkte Belagsdauerausreißer oben zwischen den Jahren 2008 und 2009. Es werden jeweils die gleichen Monate bzw. Zeitintervalle der Jahre miteinander verglichen.

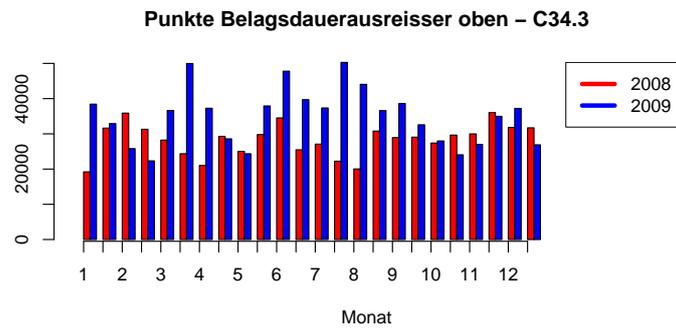
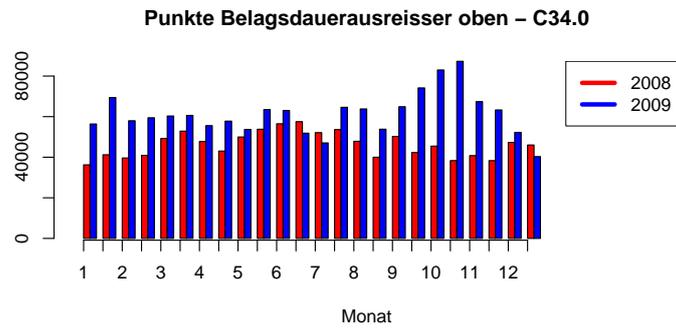


Abbildung 12: Entwicklung der Zeitreihe Belagsdauerausreißer oben aufgeschlüsselt nach Diagnosen. Die Balken vergleichen direkt die Anzahl der Punkte Belagsdauerausreißer oben zwischen den Jahren 2008 und 2009. Es werden jeweils die gleichen Monate bzw. Zeitintervalle der Jahre miteinander verglichen.

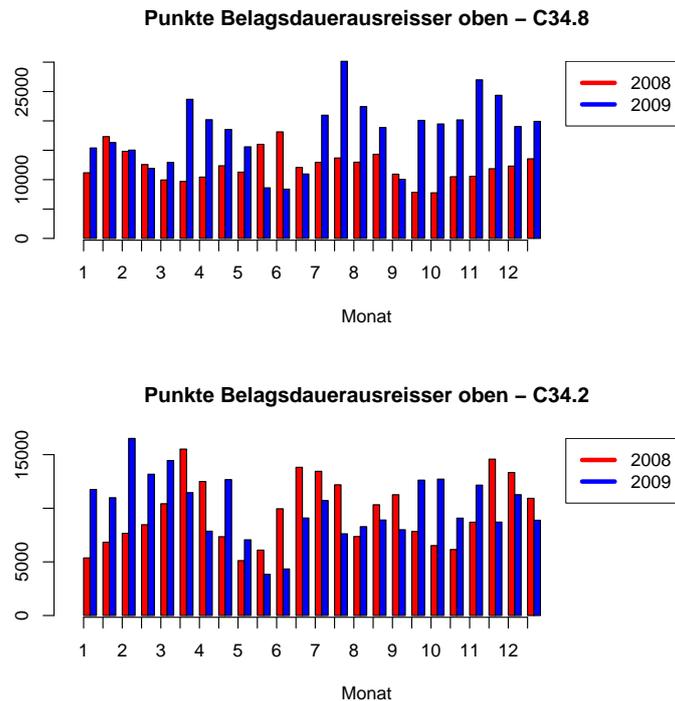


Abbildung 13: Entwicklung der Zeitreihe Belagsdauerausreißer oben aufgeschlüsselt nach Diagnosen. Die Balken vergleichen direkt die Anzahl der Punkte Belagsdauerausreißer oben zwischen den Jahren 2008 und 2009. Es werden jeweils die gleichen Monate bzw. Zeitintervalle der Jahre miteinander verglichen.

Das eingeführte Verfahren überprüft im wesentlichen, ob zwei Datengruppen mit dem gleichen linearen Modell sinnvoll beschrieben werden können. Diese Datengruppen müssen nicht zwangsläufig aufeinander folgende Zeitabschnitte sein. Genauso kann getestet werden, ob zwei Regionen oder zwei Krankenanstalten dem gleichen Modell folgen. Sinnvoll ist es hier natürlich die Zeit einzuschränken um dem Einfluss von zeitlich bedingte Brüchen ausgesetzt zu sein. Ein Beispiel ist durch Tabelle 8 gegeben.

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
K335_S	1.491	1.082	1.101	1.213	1.377	1.230	1.254	1.257	1.189	1.208
K335_R	0.755	0.974	0.597	0.466	0.948	0.540	0.465	0.999	0.993	0.998
K335_A	0.956	0.216	0.443	0.621	0.921	0.852	0.542	0.008	0.928	0.002

Tabelle 8: Tests auf unterschiedliche Struktur zwischen Krankenanstalten und Regionen von 2001 bis 2010: Belagstage \sim Monat

*_S ... Statistik, *_R ... Vergleich mit Region, *_A ... Vergleich mit Österreich gesamt

Für jede Krankenanstalt werden für jedes Jahr 3 Kennzahlen angegeben. Die

Zeilennamen repräsentieren zunächst die Krankenanstaltennummern, in Tabelle 8 K335. Die erste Kennzahl, „S“, gibt die betrachtete Statistik in 1000 an um ein Gefühl für die Relevanz der Krankenanstalt im betrachteten Bereich zu geben. Die zweite Kennzahl, „R“, stellt den p-Wert des durchgeführten Strukturtests zwischen der Krankenanstalt und dem gesamten Bundesland dar. Die dritte Kennzahl, „A“, wiederholt diesen Test für ganz Österreich. Der fehlende Wert „-“ bedeutet in diesem Zusammenhang, dass nicht für jeden Zeitpunkt des betrachteten Jahres für die betrachtete Krankenanstalt Aufenthalte registriert wurden, bzw. die betrachtete Statistik nicht existiert.

In Tabelle 8 gibt es in zwei Jahren, nämlich 2008 und 2010 einen signifikanten Unterschied zwischen dem beschreibenden Modell der Krankenanstalt und gesamt Österreich.

Zuletzt wird noch ein Test zwischen Krankenanstalten durchgeführt. Dabei muss abermals jedes Jahr gesondert betrachtet werden, da sonst die lineare Struktur innerhalb der Krankenanstalten nicht mehr gegeben ist. Darüber hinaus ist der Test derzeit auf Vergleiche innerhalb der Regionen eingeschränkt. Dies liegt an dem nicht linear wachsenden Aufwand mit zusätzlichen Vergleichen. Grundsätzlich können so alle Krankenanstalten miteinander verglichen werden, der Rechenaufwand benötigt jedoch bei geeigneter Permutationszahl mehrere Monate. Ein Beispielergebnis wird in Tabelle 9 dargestellt.

Aus dieser Beispieltabelle des Tests auf Belagstage im Jahr 2010 kann etwa erkannt werden, dass einige Krankenanstalten niedrige p-Werte aufweisen, im Test gegen andere Krankenanstalten mit niedrigen p-Werten aber fast immer hohe p-Werte haben. Aus diesen Ergebnissen könnten Krankenanstalten auch weiter in Gruppen gefasst werden, die unterschiedliches Abrechnungsverhalten aufweisen, das vermutlich darauf zurückzuführen ist, dass nicht alle Krankenanstalten Lungentumore behandeln. Ein gutes Beispiel ergibt sich aus der ersten Spalte. Die Krankenanstalt 303 ist nur gegenüber den Anstalten 347, 347, 354 und 356 gegenüber auffällig. Untereinander hingegen erscheinen diese Krankenanstalten als sehr homogen.

	K303	K315	K316	K319	K321	K326	K334	K335	K338	K347
K303	-	-	-	-	-	-	-	-	-	-
K315	0.524	-	-	-	-	-	-	-	-	-
K316	0.679	0.980	-	-	-	-	-	-	-	-
K319	0.959	0.463	0.622	-	-	-	-	-	-	-
K321	0.542	0.147	0.216	0.332	-	-	-	-	-	-
K326	0.143	0.564	0.484	0.086	0.043	-	-	-	-	-
K334	0.320	0.839	0.779	0.269	0.091	0.935	-	-	-	-
K335	0.851	0.227	0.366	0.901	0.404	0.036	0.126	-	-	-
K338	0.779	0.877	0.964	0.754	0.235	0.353	0.623	0.537	-	-
K347	0.019	0.128	0.115	0.016	0.004	0.730	0.491	0.002	0.072	-
K354	0.008	0.044	0.061	0.004	0.000	0.528	0.312	0.000	0.028	0.907
K356	0.036	0.227	0.177	0.021	0.019	0.851	0.639	0.009	0.130	0.968
K377	0.535	0.997	0.967	0.470	0.200	0.711	0.916	0.287	0.851	0.261
K378	0.309	0.863	0.785	0.231	0.094	0.859	0.998	0.093	0.622	0.370
K379	0.473	0.985	0.925	0.410	0.177	0.715	0.934	0.227	0.806	0.231
K380	0.961	0.455	0.612	0.999	0.368	0.106	0.248	0.901	0.782	0.009
K382	0.872	0.767	0.886	0.854	0.284	0.238	0.476	0.611	0.974	0.047
K383	0.573	0.995	0.996	0.539	0.203	0.576	0.839	0.342	0.921	0.151

Tabelle 9: Vergleich der Strukturen zwischen Krankenanstalten über p-Werte im Modell $\text{Belagstage} \sim \text{Monat}$ im Jahr 2010 eingeschränkt auf einen Teil der Krankenanstalten der Gruppe K3.

Literatur

- [1] D.K.K. Andrews. “Test for Parameter Instability and Structural Change With Unknown Change Point”. *Econometrica* 61 (1993), 821–856.
- [2] D.K.K. Andrews und W. Ploberger. “Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative”. *Econometrica* 62 (1994), 1383–1414.
- [3] J. M. Chambers und T. J. Hastie. *Statistical Models in S*. Wadsworth & Brooks/Cole., 1992.
- [4] Gregory C. Chow. “Tests of equality between sets of coefficients in two linear regressions”. *Econometrica* 28 (1960), 591–605.
- [5] R.B. Davies. “Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative”. *Biometrika* 64 (1977), 247–254.
- [6] B.E. Hansen. “Approximate asymptotic p values for structural-change tests.” *Journal of Business & Economic Statistics* 15 (1997), 60–67.

- [7] H. Wold. “Estimation of principal components and related models by iterative least squares”. *Krishnaiah, P.R. Multivariate Analysis* (1966), 391–420.
- [8] Zeileis, A. and Leisch, F. and Hornik, K. and Kleiber, C. “strucchange. An R package for testing for structural change in linear regression models”. *Adaptive Information Systems and Modelling in Economics and Management Science* 55 (2001).