Zusatzfolien zum Referatteil C zu einer Frage aus der Zuhörerschaft der FG Si ZNO anlässlich der Präsentation von C. Knüsli während der Sitzung vom 1.7.2021 in Andelfingen / ZH

Nachtrag C. Knüsli z. Hd. von M. Ott,

Präsident FG Si ZNO vom 14.7.2021

Frage vom 1.7.2021 zur Nachweisführung eines Sprungs (level shift) des Trends der Säuglingssterblichkeitsrate (SSt-Rate) in der Schweiz bei Zeitpunkt 1986/1987

Frage vom 1.7.2021 von Herrn … : *«Wenn es einen Bruch (level shift) des Verlaufes der SSt-Rate in der Schweiz zwischen 1969 und 2019 gibt: Weshalb kann gesagt werden, dass dieser Bruch genau bei 1986/1987 stattgefunden hat? Würde sich nicht auch sonst irgendeine beliebige «Zacke» im Verlauf der Kurve der SSt-Rate des BFS zwischen 1969 und 2019 für einen solchen mathematischen Nachweis anbieten?»* (Diese Frage wird im Folgenden verkürzt als **«Frage vom 1.7.2021**» bezeichnet)

Zusammenfassende Antwort vom 14.7.2021 von Dr. C. Knüsli, Basel, und Dr. H. Scherb, München:

Es kann mit mindestens drei – von einander unabhängigen – mathematischen Methoden gezeigt werden, dass im zur Diskussion stehenden 51-jährigen Verlauf ein maximaler Sprung (level shift) genau zum Zeitpunkt 1986/1987 besteht. Zu diesem Zeitpunkt ist die statistische Wahrscheinlichkeit für einen Sprung (level shift) am höchsten im Vergleich zu den übrigen 48 theoretisch möglichen Zeitpunkten.

Die detaillierte Antwort wird wie folgt gegeben …

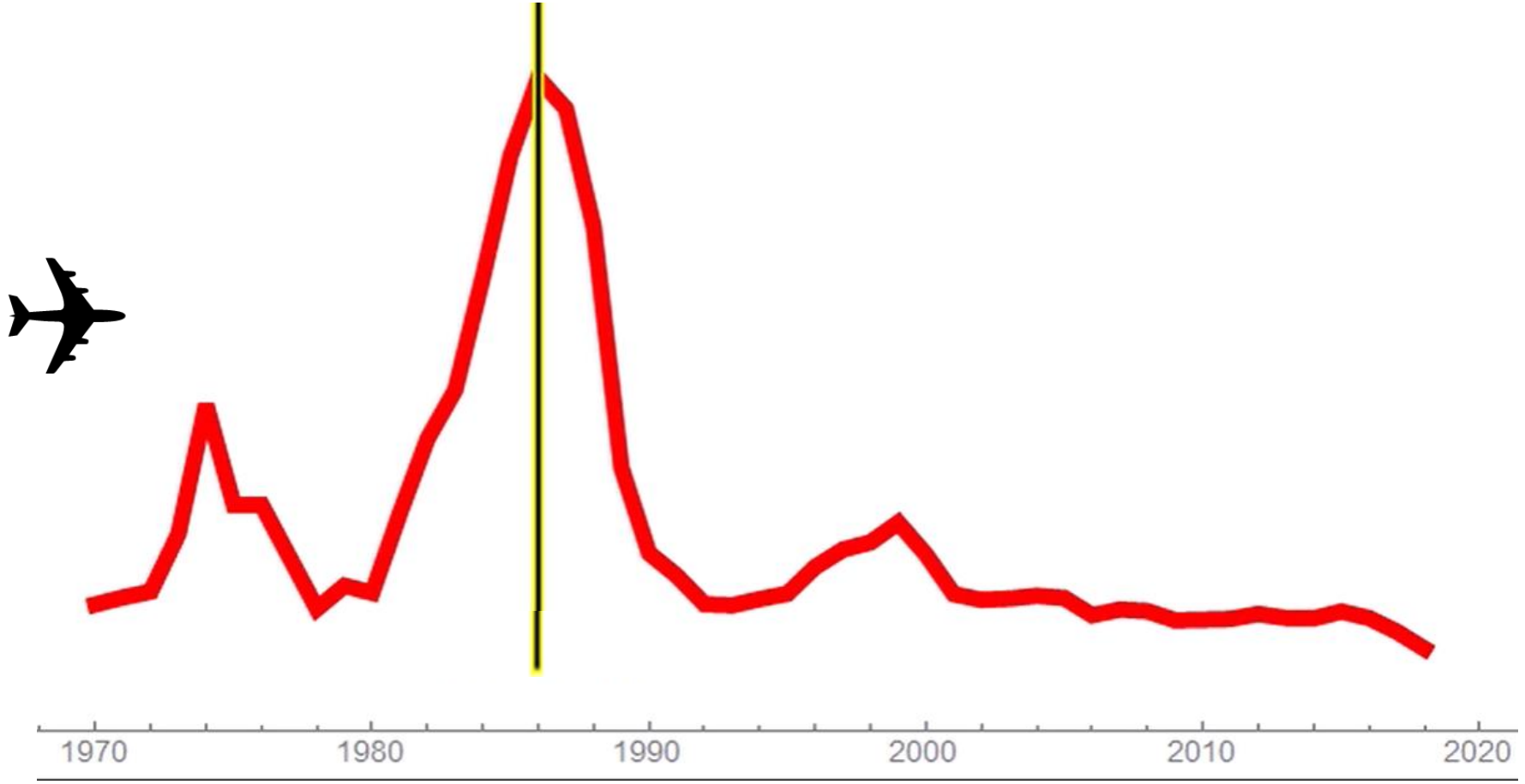
1. anhand eines intuitiv aufzufassenden analogen Vergleiches ***für informierte Laien*** (Folien 3 bis 6),

2. anhand von 19 sachspezifischen Darstellungen (Folien 9 bis 27) ***für Spezialisten mit epidemiologisch-mathematischer Ausbildung.***

1. Intuitiv aufzufassender **analoger Vergleich** für informierte Laien zur „Frage vom 1.7.2021“

- Grundsätzlich: Es geht um ein ***Optimierungsproblem***, welches nur ***eine «beste» Lösung*** zulässt.
- Stellen Sie sich vor, Sie müssen von A. («1969») nach B. («2019») über einen hohen Schweizer Gebirgspass fliegen, dessen genaue Lage im Verlauf der Flugstrecke Ihnen nicht bekannt ist.
- Der Pilot teilt Ihnen mit, dass er den Passüberflug aufgrund der knappen Treibstoffreserve nur dann sicher schaffen würde, wenn er den Beginn des Steigfluges (ab erreichter mittlerer Flughöhe) über den Gebirgspass zum einzig möglichen, korrekten Zeitpunkt einplane.
- Sollte er mit dem Steigflug jedoch zu früh beginnen, würde sich das Flugzeug über dem Pass wegen Treibstoffmangel bereits wieder im Sinkflug befinden und am Pass wegen zu geringer Höhe wahrscheinlich zerschellen – ebenso bei ungenügender Höhe des Flugzeugs, sollte er mit dem Steigflug zu spät beginnen.
- Der Radar stellt dem Piloten folgendes Profilbild bezüglich der Hindernis-höhe auf dem Flug von A. nach B. zur Verfügung (siehe nächste Folie):

Vergleich: Radar-Hindernisprofil für Flug von A.(«1969») nach B.(«2019»)



Könnten Sie dem Piloten zustimmen, wenn er vorschlägt, den Gebirgspass-Überflug am besten bei «1986» einzuplanen ?

- Ja: mit welcher Begründung?

- Nein: mit welcher Begründung?

1. Intuitiv aufzufassender **analoger Vergleich** für informierte Laien zur „Frage vom 1.7.2021“

- Auch die fachlich-mathematische Beantwortung der Frage nach der Lage eines Trendsprungs (level shift) der Säuglingssterblichkeitsrate in der Schweiz orientiert sich einerseits an der Bedingung, dass die bestmögliche (= statistisch am höchsten signifikante) Sprungvariante erfasst werden muss (vergleichbar mit der Bedingung des limitierten Flugtreibstoffs im analogen Beispiel des Fluges von A. nach B.), sowie andererseits an einer «abtastenden mathematischen Radarfunktion» für alle 48 in Frage kommenden Zeitpunkte für den Sprung (level shift).

- **Die drei erwähnten Berechnungsvarianten ergeben alle übereinstimmend für den Zeitpunkt 1986/1987 die höchste Wahrscheinlichkeit des Sprungs (level shift) des Trends der Säuglingssterblichkeitsrate in der Schweiz im Verlaufe der Jahre 1969 bis 2019.**

2. Es folgen 19 sachspezifischen Darstellungen (Folien 9 bis 27) zur Beantwortung der «Frage vom 1.7.2021» für Spezialisten mit epidemiologisch-mathematischer Ausbildung

Übersicht zur Strukturierung der Folien zu den drei benützten «change-point methods»:

a. Folie 9 bis 12: quadratic logistic trend

b. Folie 13 bis 16: ***logistic regression***

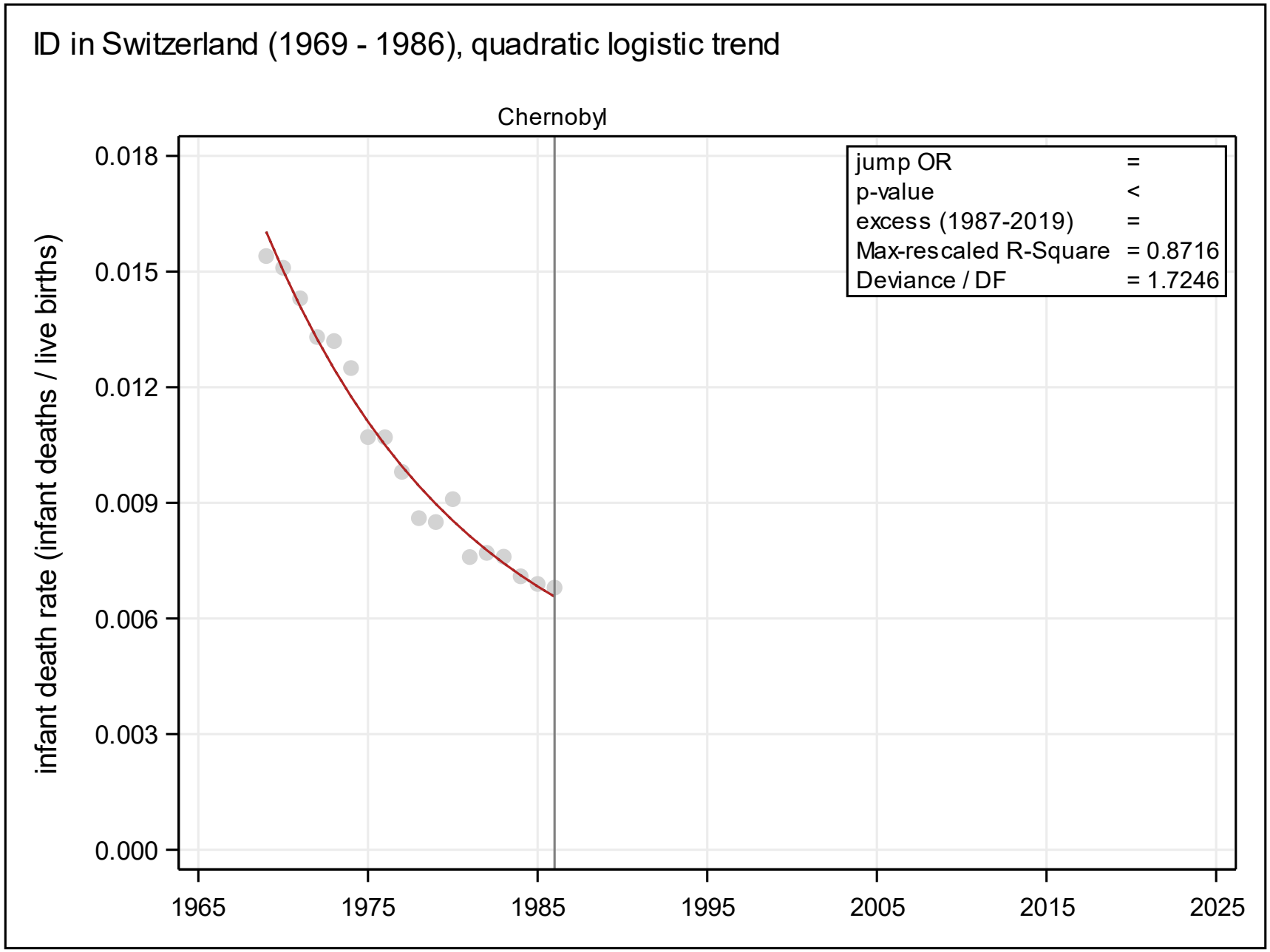
c. Folie 17 bis 20: ***Markov chain Monte Carlo (MCMC)*** for change points

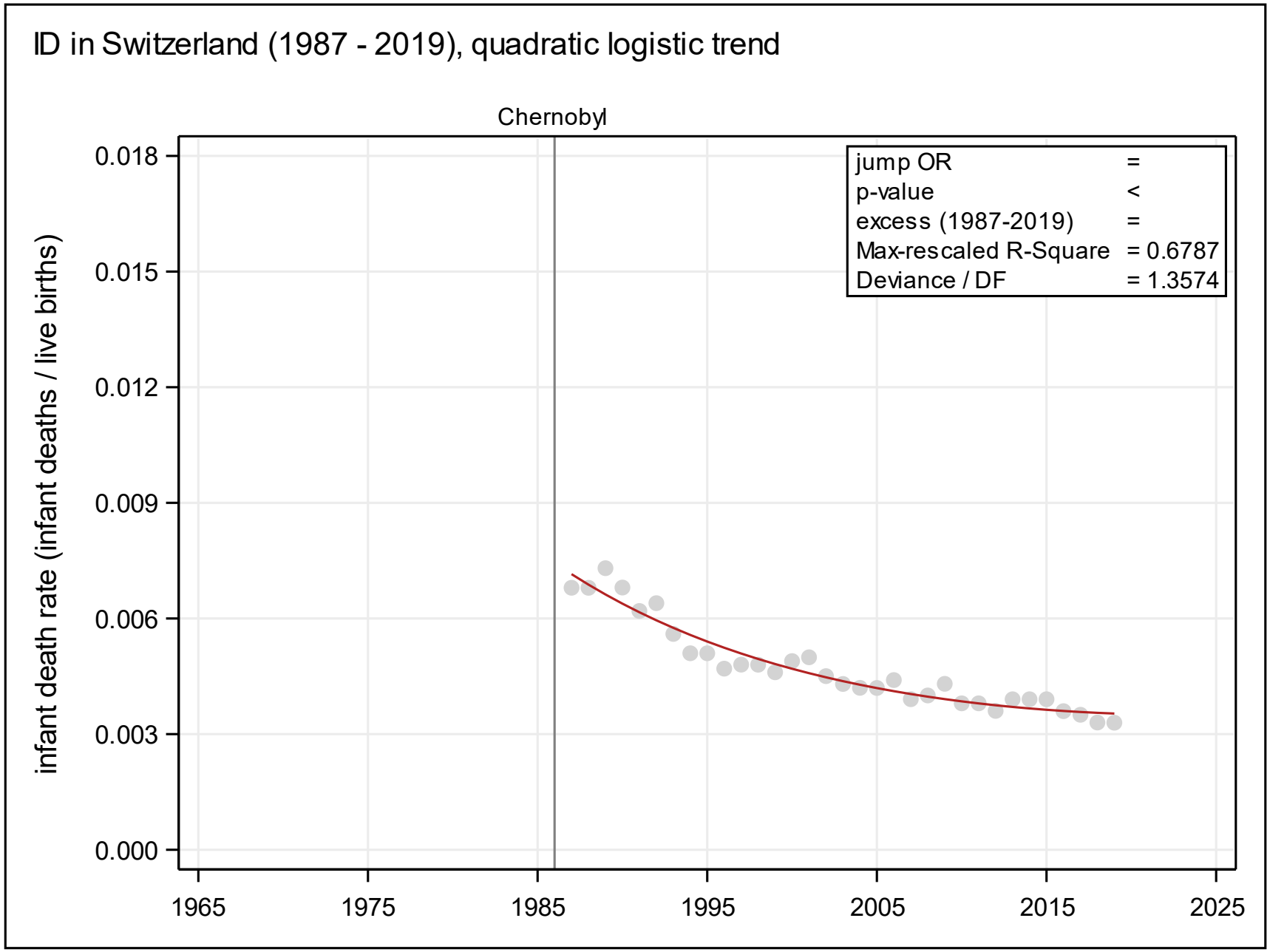
d. Folie 21 bis 23: ***non-linear logistic regression***

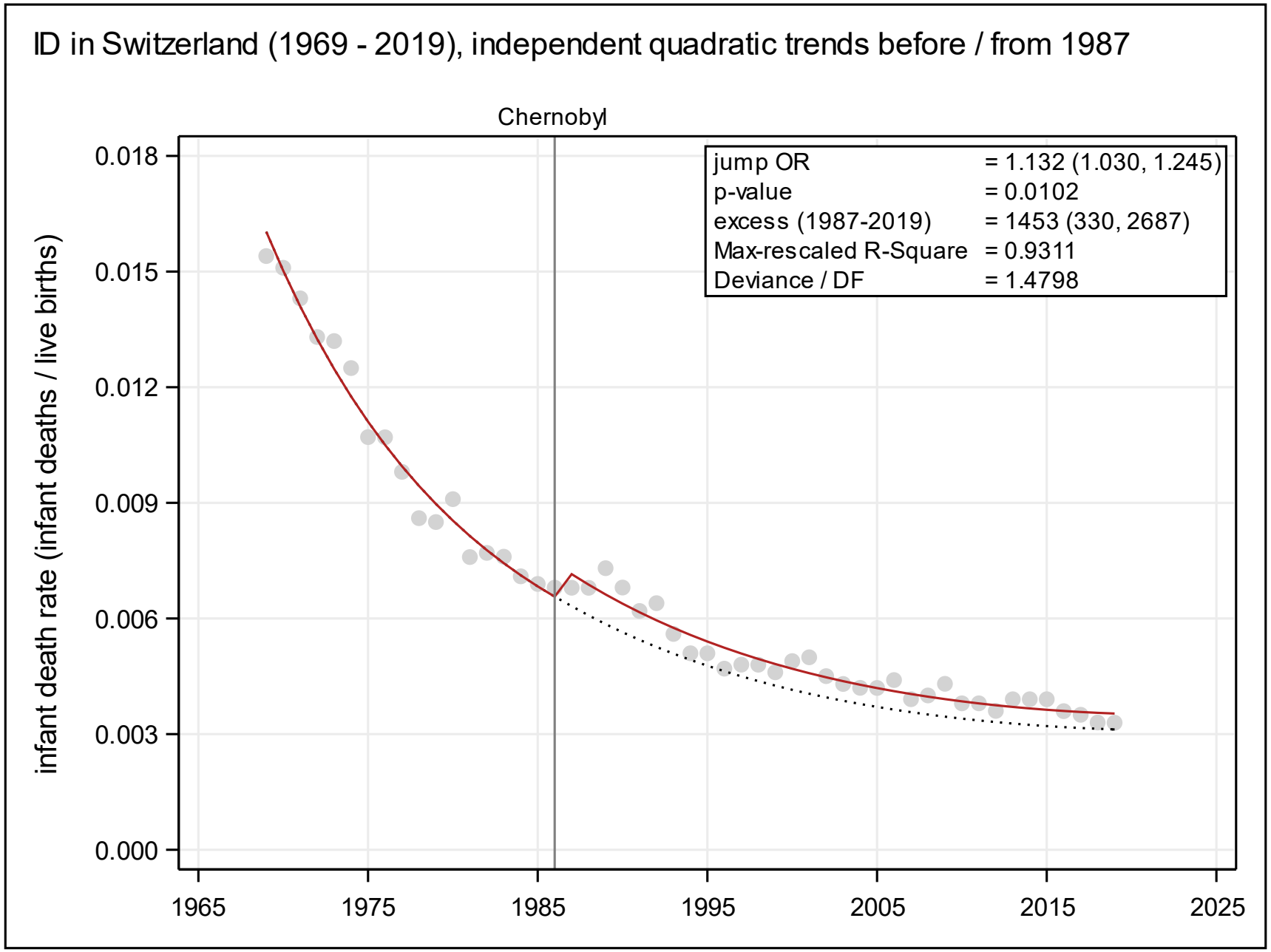
e. Folie 24 bis 27: comparison of the jump odds ratios for three change-point methods

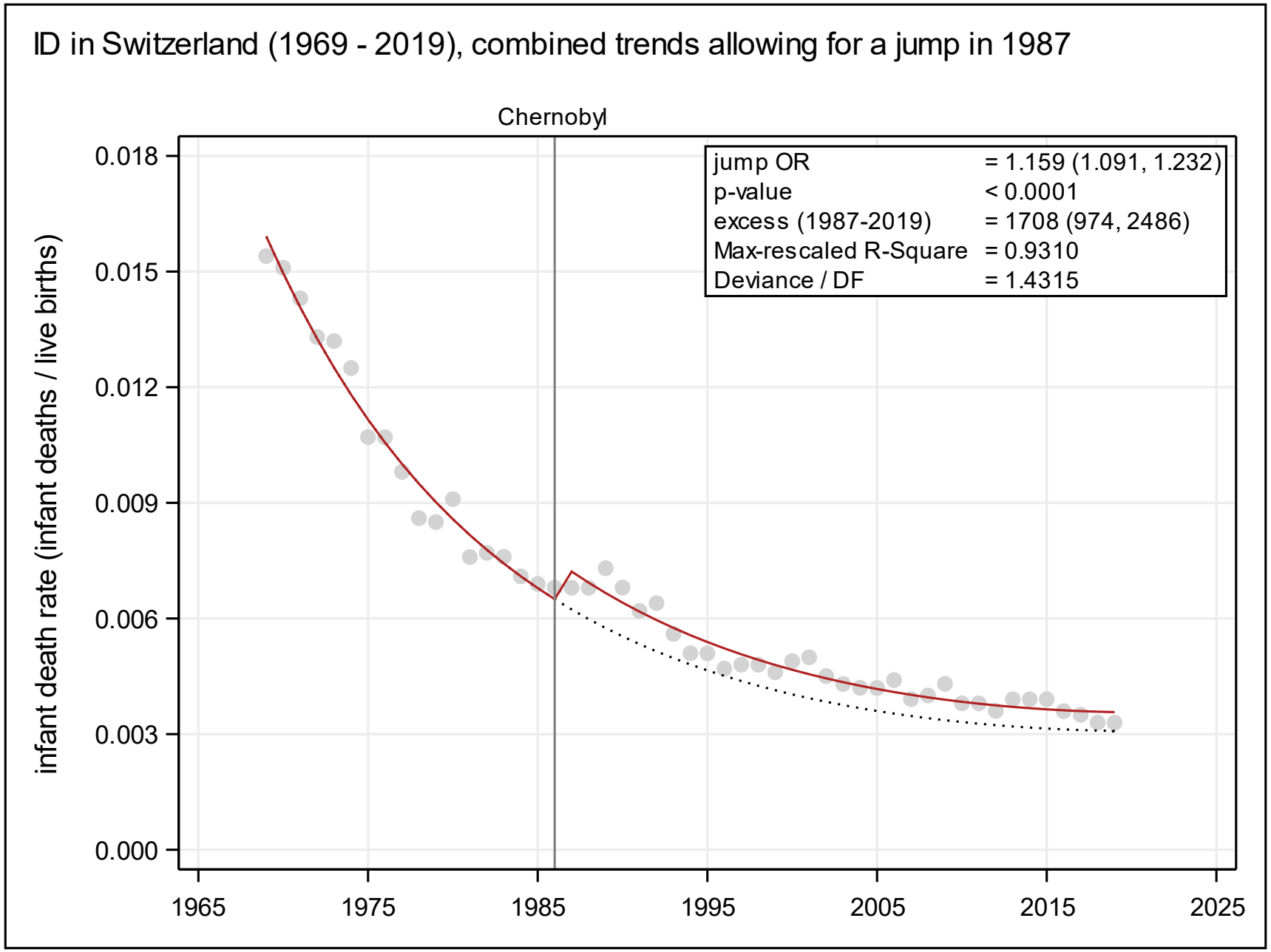
Change-point estimation by three methods for the secular infant death trend in Switzerland 1969-2019 in context with Chernobyl 1986

Hagen Scherb in cooperation with Claudio Knüsli, July 2021


ID in Switzerland (1969 - 1986), quadratic logistic trend
Chernobyl
infant death rate (infant deaths / live births)
0.018
0.015
0.012
0.009
0.006
0.003
0.000
1965
1975
1985
1995
2005
2015
2025
jump OR =
p-value <
excess (1987-2019) =
Max-rescaled R-Square = 0.8716
Deviance / DF = 1.7246


ID in Switzerland (1987 - 2019), quadratic logistic trend
Chernobyl
jump OR =
p-value <
excess (1987-2019) =
Max-rescaled R-Square = 0.6787
Deviance / DF = 1.3574
infant death rate (infant deaths / live births)
0.018
0.015
0.012
0.009
0.006
0.003
0.000
1965
1975
1985
1995
2005
2015
2025


ID in Switzerland (1969 - 2019), independent quadratic trends before / from 1987
Chernobyl
jump OR = 1.132 (1.030, 1.245)
p-value = 0.0102
excess (1987-2019) = 1453 (330, 2687)
Max-rescaled R-Square = 0.9311
Deviance / DF = 1.4798
infant death rate (infant deaths / live births)
0.018
0.015
0.012
0.009
0.006
0.003
0.000
1965
1975
1985
1995
2005
2015
2025


ID in Switzerland (1969 - 2019), combined trends allowing for a jump in 1987
Chernobyl
jump OR = 1.159 (1.091, 1.232)
p-value < 0.0001
excess (1987-2019) = 1708 (974, 2486)
Max-rescaled R-Square = 0.9310
Deviance / DF = 1.4315
infant death rate (infant deaths / live births)
0.018
0.015
0.012
0.009
0.006
0.003
0.000
1965
1975
1985
1995
2005
2015
2025

Deviance as a goodness of fit measure for logistic regression

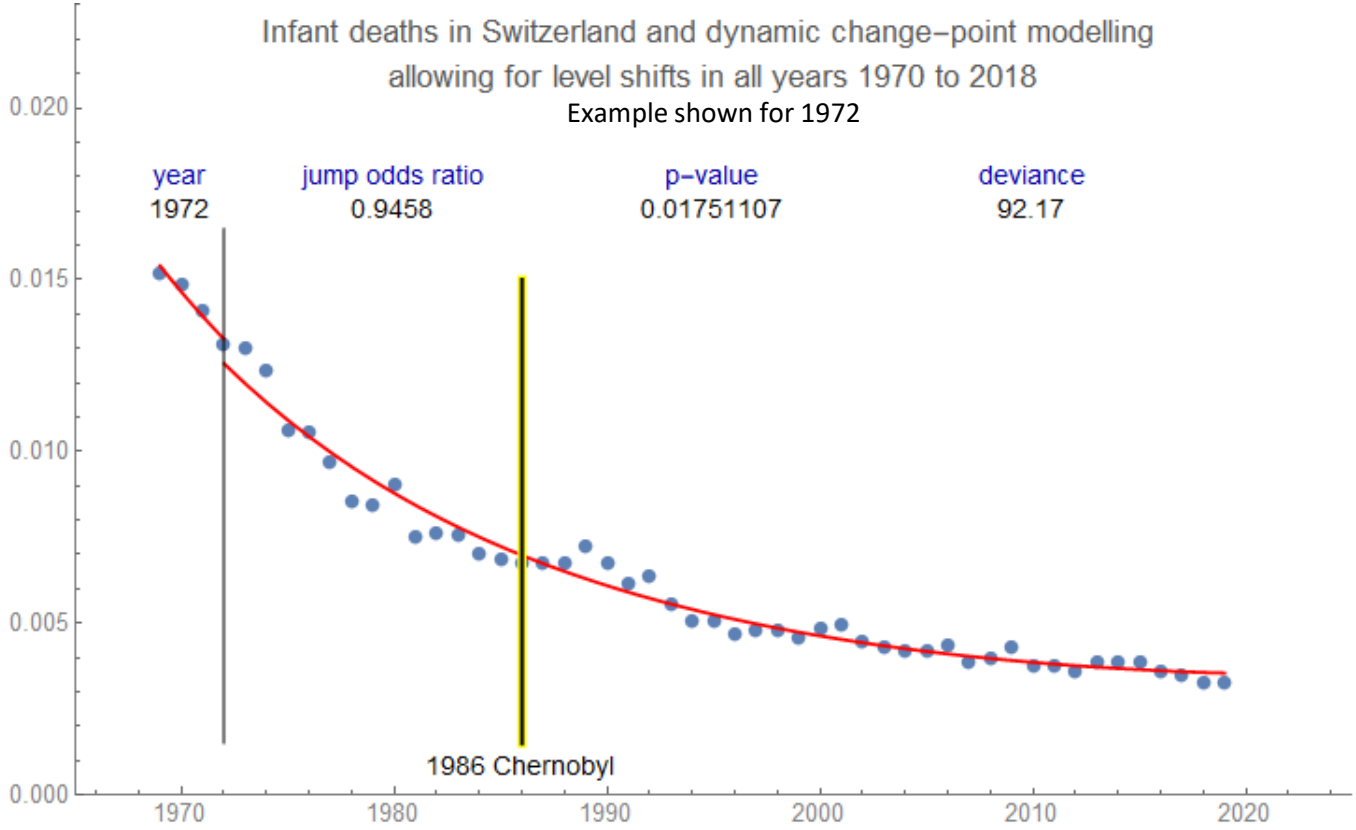
The formula for the deviance is

$$2 * \sum_{i=1}^{n} observed(i) * \log\left(\frac{observed(i)}{expected(i)}\right)$$

where each *i* is a cell in the 2-way contingency table with each row being a profile and each column being one of the two categories of the dependent variable. *Observed(i)* is the observed frequency and expected(i) is the expected frequency based on the fitted model. If *Observed(i) =0*, the entire term in the summation is set to 0. The degrees of freedom is the number of profiles minus the number of estimated parameters.

The closer the ratio *deviance/degrees of freedom* to the value of 1.0, the better the model fits to the data.

Deviance as a goodness of fit measure for logistic regression



Für jedes Jahr von 1970 bis 2018 kann ein Sprungmodell aus jeweils **zwei optimalen Kurvensegmenten** angepasst werden, und die Sprunghöhe (d.h. die jump odds ratio als Maß für den level-shift) und die Deviance werden notiert.

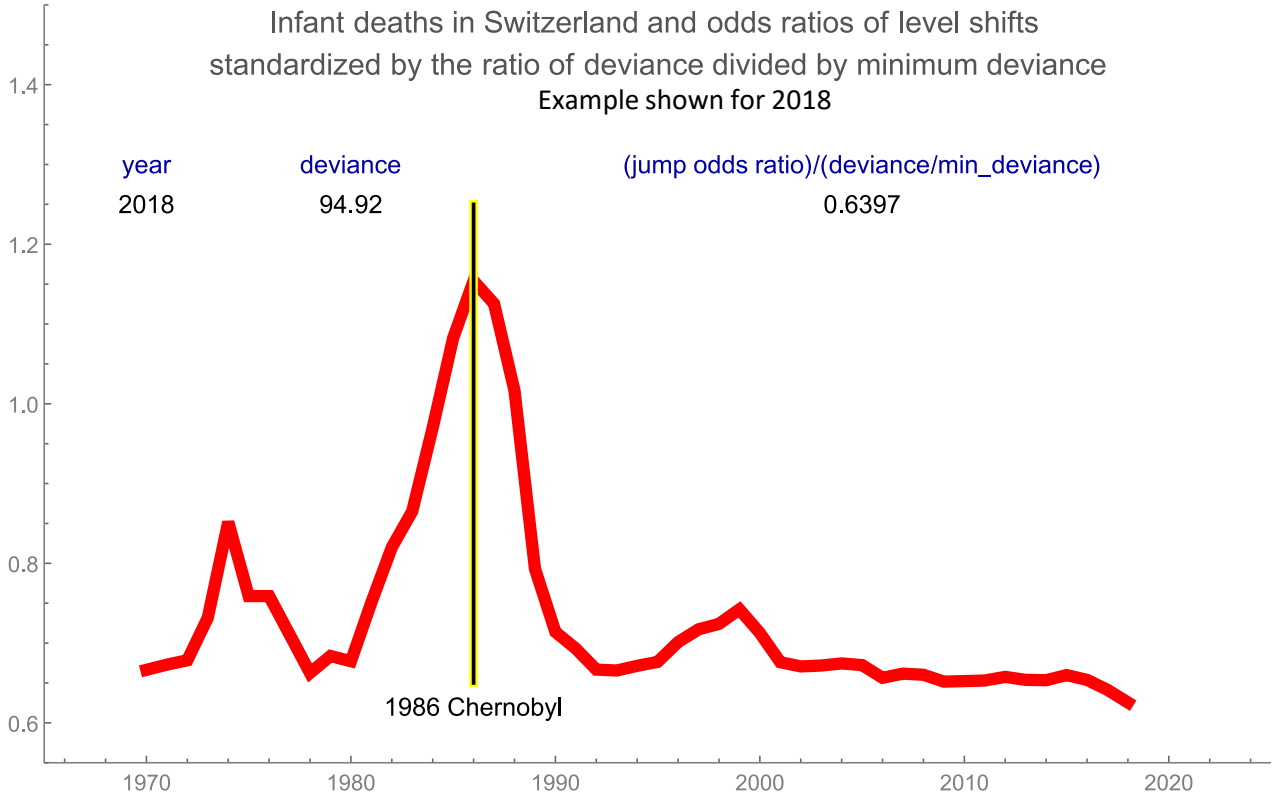
In der Abbildung werden exemplarisch die beiden Kurvensegmente für das Jahr 1972 dargestellt.

Deviance as a goodness of fit measure for logistic regression

year	jump OR	p-value	deviance	year	jump OR	p-value	deviance
1969	.	.	.	1995	1.001	0.97558900	97.8
1970	0.975	0.28905700	96.7	1996	1.028	0.33755600	96.9
1971	0.955	0.04414960	93.8	1997	1.041	0.17284100	96.0
1972	0.946	0.01751110	92.2	1998	1.047	0.13480800	95.6
1973	0.908	0.00007243	82.0	1999	1.060	0.06592530	94.5
1974	0.868	0.00000004	67.4	2000	1.040	0.24101200	96.5
1975	0.885	0.00000578	77.1	2001	1.000	0.99393100	97.8
1976	0.880	0.00000456	76.7	2002	0.992	0.82030700	97.8
1977	0.898	0.00016309	83.6	2003	0.994	0.86597600	97.8
1978	0.961	0.16102200	95.9	2004	0.998	0.95656500	97.8
1979	1.010	0.71874400	97.7	2005	0.995	0.89020400	97.8
1980	1.002	0.94376800	97.8	2006	0.961	0.31179600	96.8
1981	1.061	0.03589680	93.4	2007	0.975	0.51792500	97.4
1982	1.091	0.00171039	88.0	2008	0.972	0.46273100	97.3
1983	1.105	0.00025653	84.4	2009	0.929	0.05700030	94.2
1984	1.130	0.00000525	77.0	2010	0.938	0.09589470	95.1
1985	1.148	0.00000015	70.1	2011	0.942	0.11660900	95.4
1986	1.157	0.00000002	66.3	2012	0.963	0.31965100	96.8
1987	1.151	0.00000004	67.6	2013	0.946	0.13894800	95.6
1988	1.133	0.00000091	73.7	2014	0.923	0.03544820	93.4
1989	1.075	0.00415610	89.6	2015	0.892	0.00378281	89.4
1990	1.036	0.16569900	95.9	2016	0.887	0.00460060	89.7
1991	1.019	0.44933300	97.3	2017	0.881	0.00872346	90.8
1992	0.981	0.45269600	97.3	2018	0.898	0.09223150	94.9
1993	0.977	0.37683000	97.0	2019	.	.	.
1994	0.993	0.78844300	97.8	2020	.	.	.

Für jedes Jahr von 1970 bis 2018 kann ein Sprungmodell aus jeweils **zwei optimalen Kurvensegmenten** angepasst werden, und die Sprunghöhe (d.h. die jump odds ratio als Maß für den level-shift) und die Deviance werden notiert.

Deviance as a goodness of fit measure for logistic regression



Für jedes Jahr von 1970 bis 2018 kann ein Sprungmodell aus jeweils **zwei optimalen Kurvensegmenten** angepasst werden, und die Sprunghöhe (d.h. die jump odds ratio als Maß für den level-shift) und die Deviance werden notiert.

Die rote Kurve mit Maximum in 1986 zeigt den Quotienten der normierten Sprunghöhen: dividiert durch die normierten Deviancen (=Deviance / Minimum Deviance).

Das eindeutige Maximum der normierten Sprunghöhen findet sich im Punkt 1986.

Außer in 1986 gibt es kein weiteres Jahr mit höherem Sprung bei kleinerer Deviance.

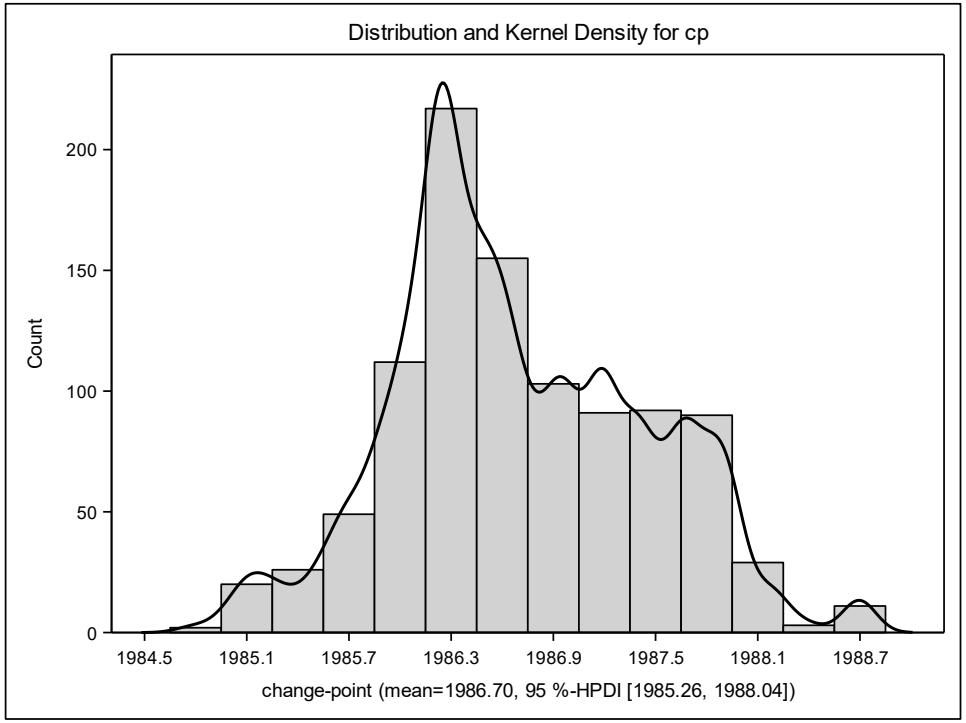
Markov chain Monte Carlo (MCMC) for Change-Points

The MCMC procedure is a general-purpose Markov chain Monte Carlo (MCMC) simulation procedure that is designed to fit Bayesian models. Bayesian statistics is different from traditional statistical methods such as frequentist or classical methods. In essence, Bayesian statistics treats parameters as unknown random variables, and it makes inferences based on the posterior distributions of the parameters.
See: https://support.sas.com/documentation/onlinedoc/stat/131/mcmc.pdf .

For a delineation and an application of MCMC to change-point estimation see: The secondary sex ratio in Italy over the past eighty years (1940 to 2019) and potential impact of radiological contamination after atmospheric nuclear testing and after Chernobyl: temporal change-point analysis using Markov Chain Monte Carlo. | Semantic Scholar

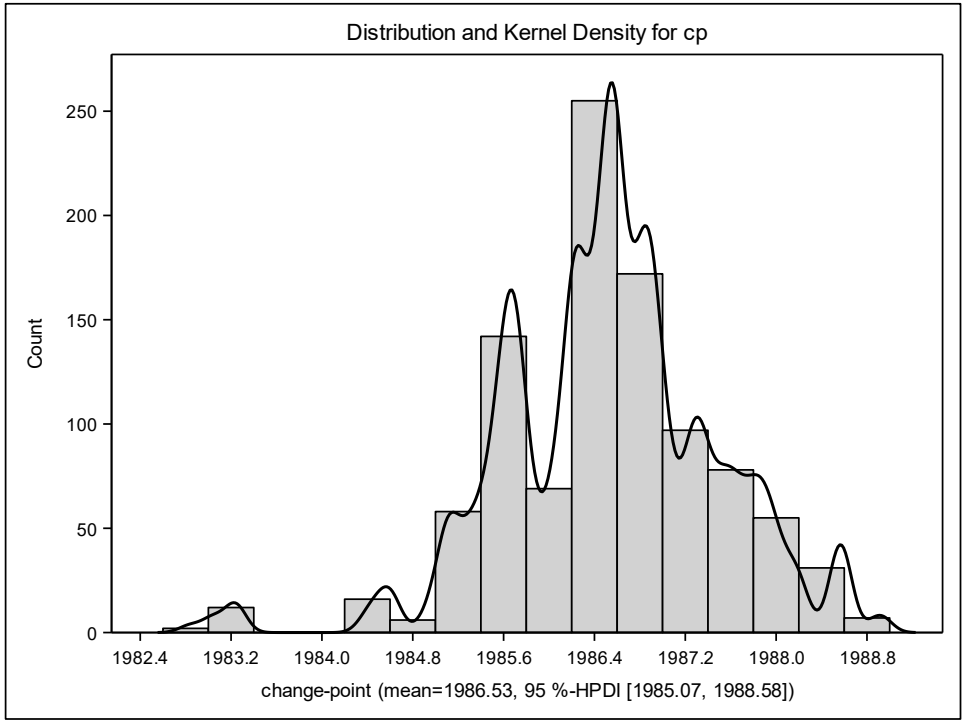
Markov chain Monte Carlo (MCMC) for Change-Points (cp)

Posterior Summaries and Intervals: prior cp ~ Normal(1987, v=2);					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
alpha	1000	-5.35830	0.01630	-5.39120	-5.33000
jump	1000	0.14310	0.01660	0.11280	0.17320
t	1000	-0.03160	0.00003	-0.03170	-0.03160
t2	1000	0.00057	0.00004	0.00049	0.00065
time-point	1000	1986.70	0.74410	1985.26	1988.04
s2	1000	0.00417	0.00091	0.00274	0.00608

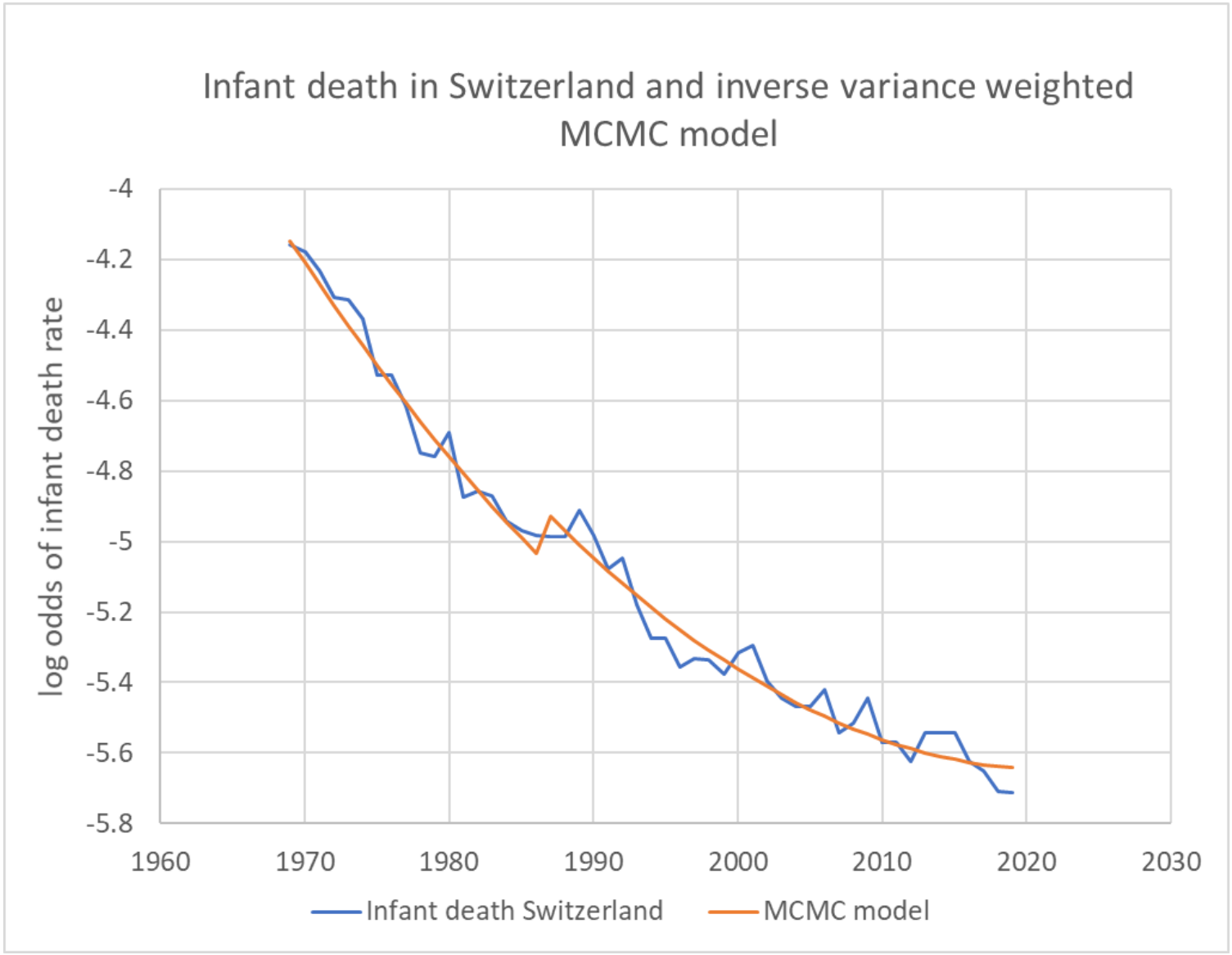


Markov chain Monte Carlo (MCMC) for Change-Points (cp)

Posterior Summaries and Intervals: prior cp ~ Uniform(1977, 1997);					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
alpha	1000	-5.36430	0.01680	-5.39660	-5.33440
jump	1000	0.14520	0.01600	0.11960	0.18090
t	1000	-0.03160	0.00004	-0.03170	-0.03160
t2	1000	0.00058	0.00004	0.00050	0.00064
time-point	1000	1986.53	0.95490	1985.07	1988.58
s2	1000	0.00418	0.00087	0.00282	0.00609



Conclusion: For a wide range of plausible prior distributions, the MCMC approach is consistent with logistic regression. An advantage is that a 95 % highest posterior density interval (95% HPDI) for the timepoint of the level-shift may be obtained.



Nonlinear logistic regression – as a 3rd possibility for CP estimation

Fit the function

$$a + bt + ct^2 + \frac{jump}{exp(s1(-(t - t2))) + 1}$$

to the log odds of the infant death

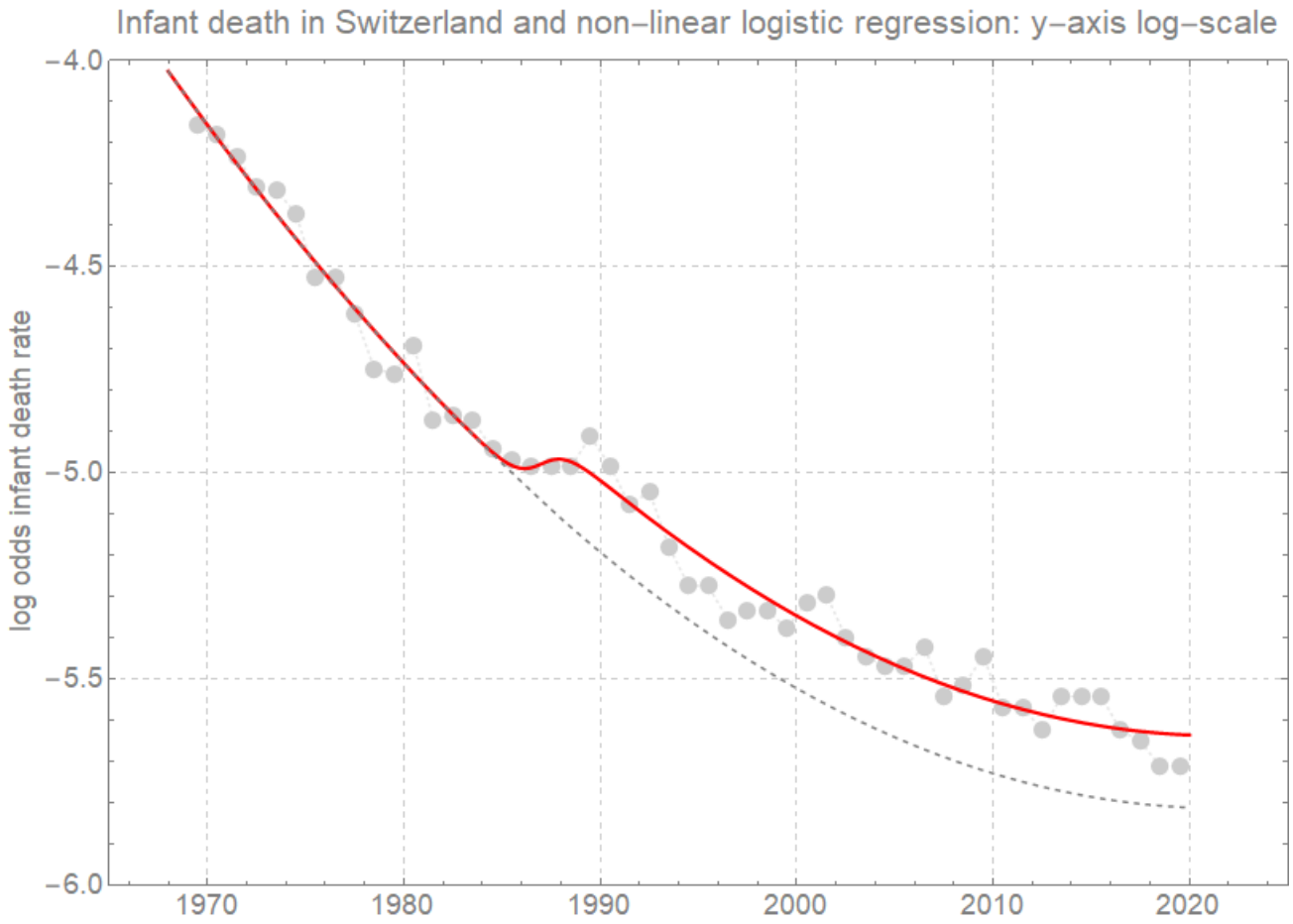
Nonlinear logistic regression – as a 3rd possibility for CP estimation

Result

Parameter Table of the non-linear logistic regression				
Parameter	**Estimate**	**Standard Error**	**t-Statistic**	**P-Value**
a	2523.15000	218.79300	11.53210	0.000000
b	-2.50184	0.21861	-11.44450	0.000000
c	0.00062	0.00005	11.33300	0.000000
jump	0.17159	0.04712	3.64180	0.000697
t2	1986.99000	0.70065	2835.91000	0.000000
s1	1.47796	1.45294	1.01722	0.314486

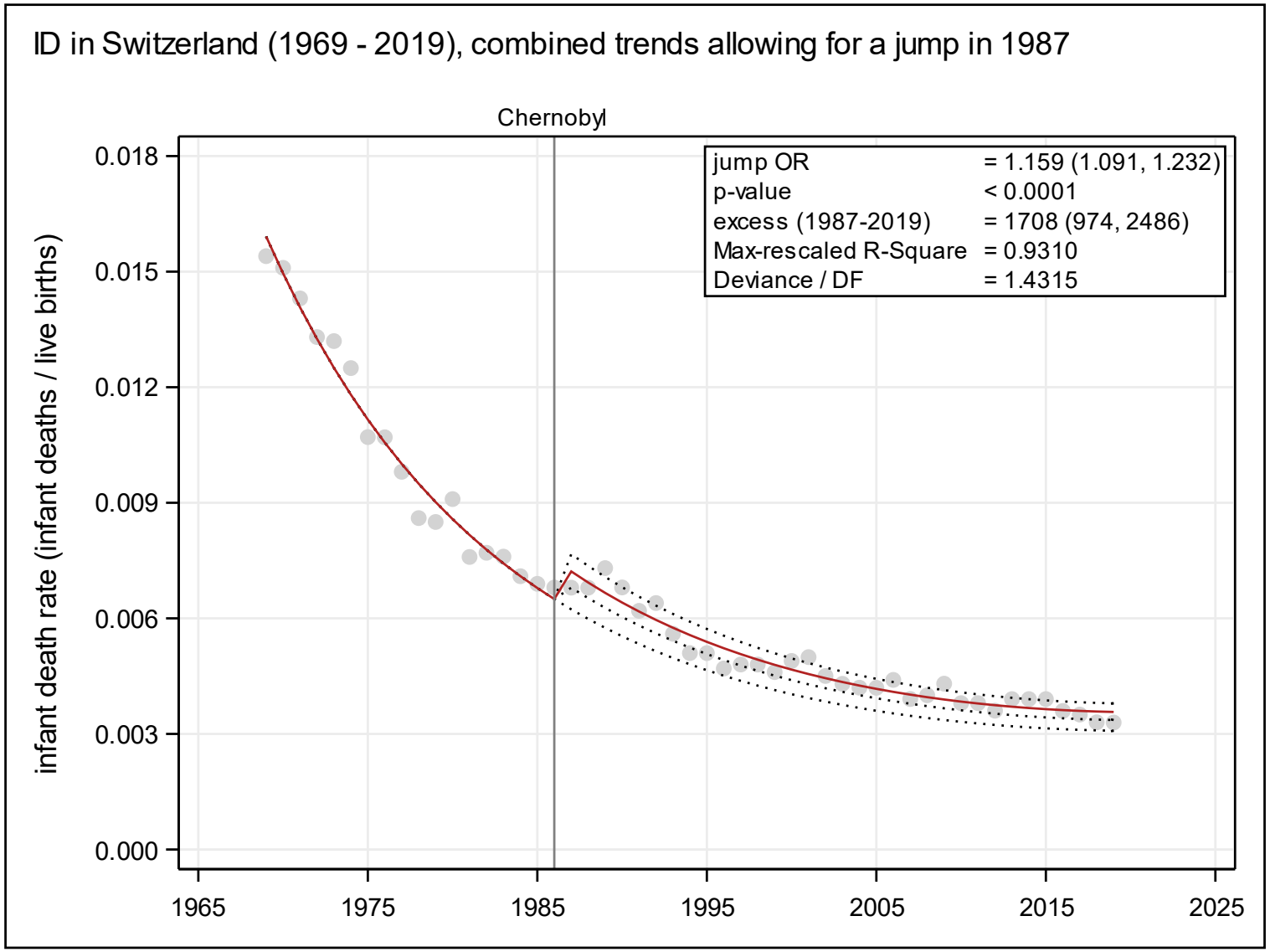
Nonlinear logistic regression – as a 3rd possibility for CP estimation

Result (note dynamical zooming for small infant death probabilities by log-scale of the y-axis)



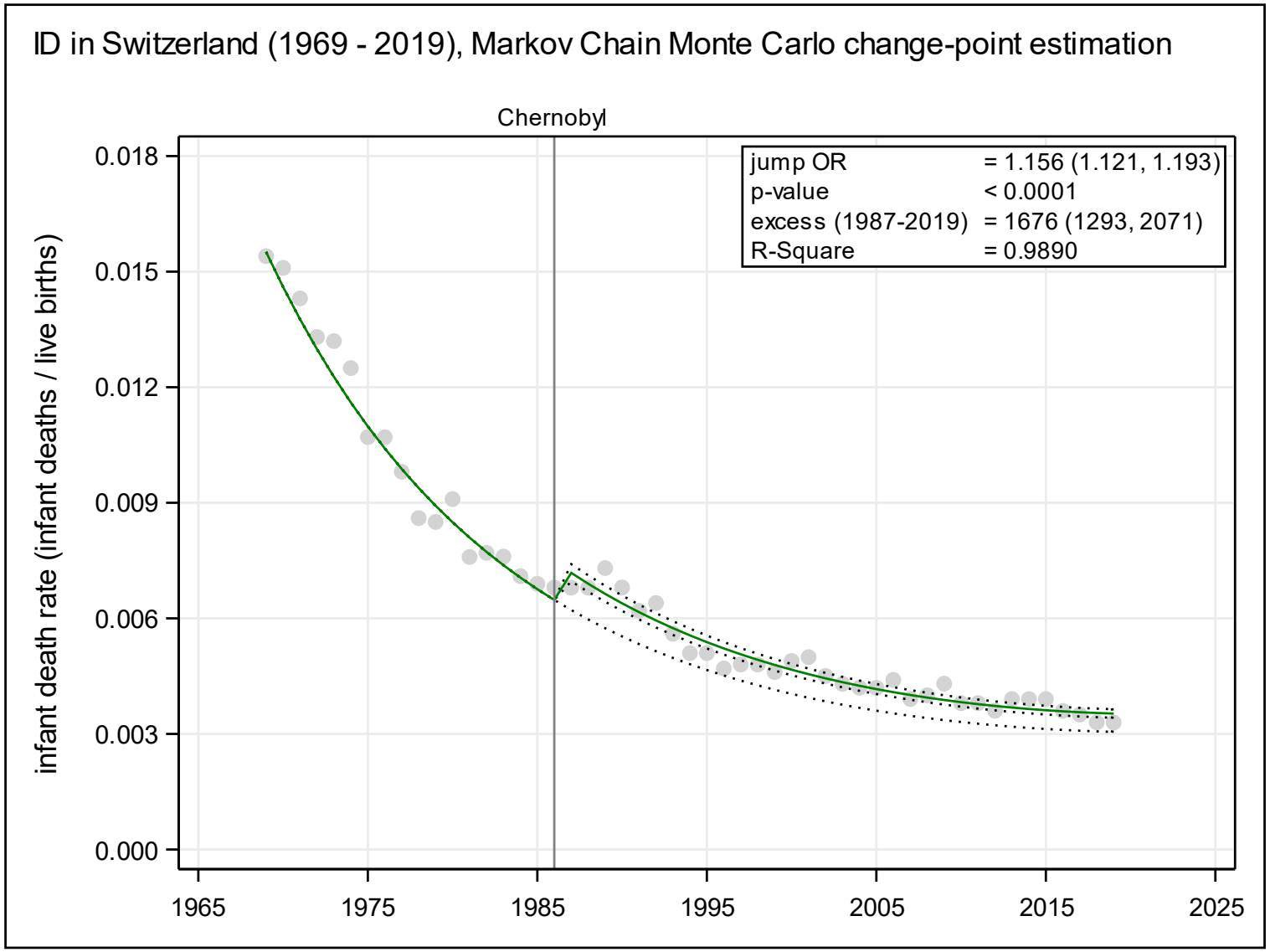
Graphical comparison of three change-point methods

Linear Logistic Regression with 95%-confidence band for the level shift



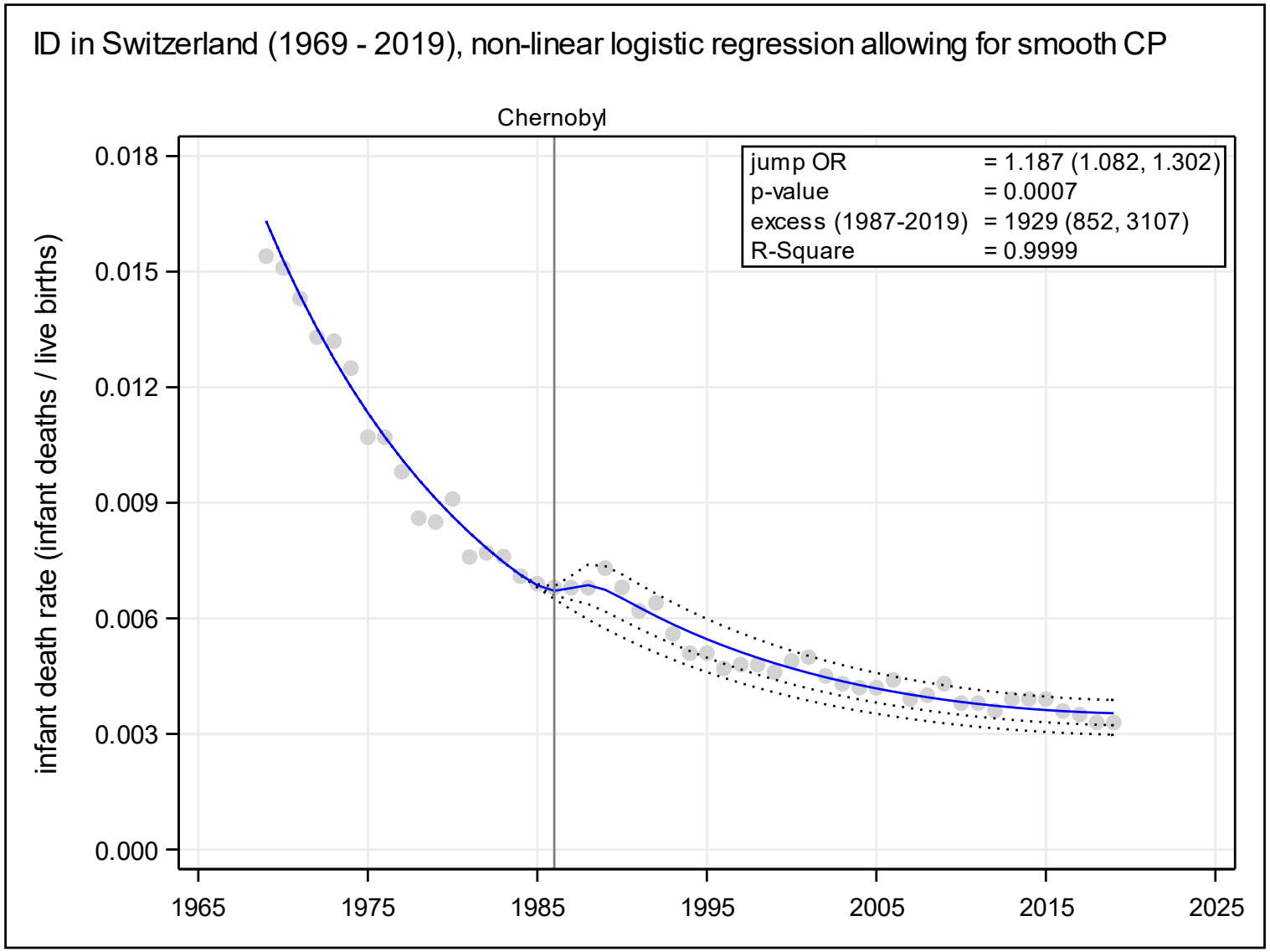
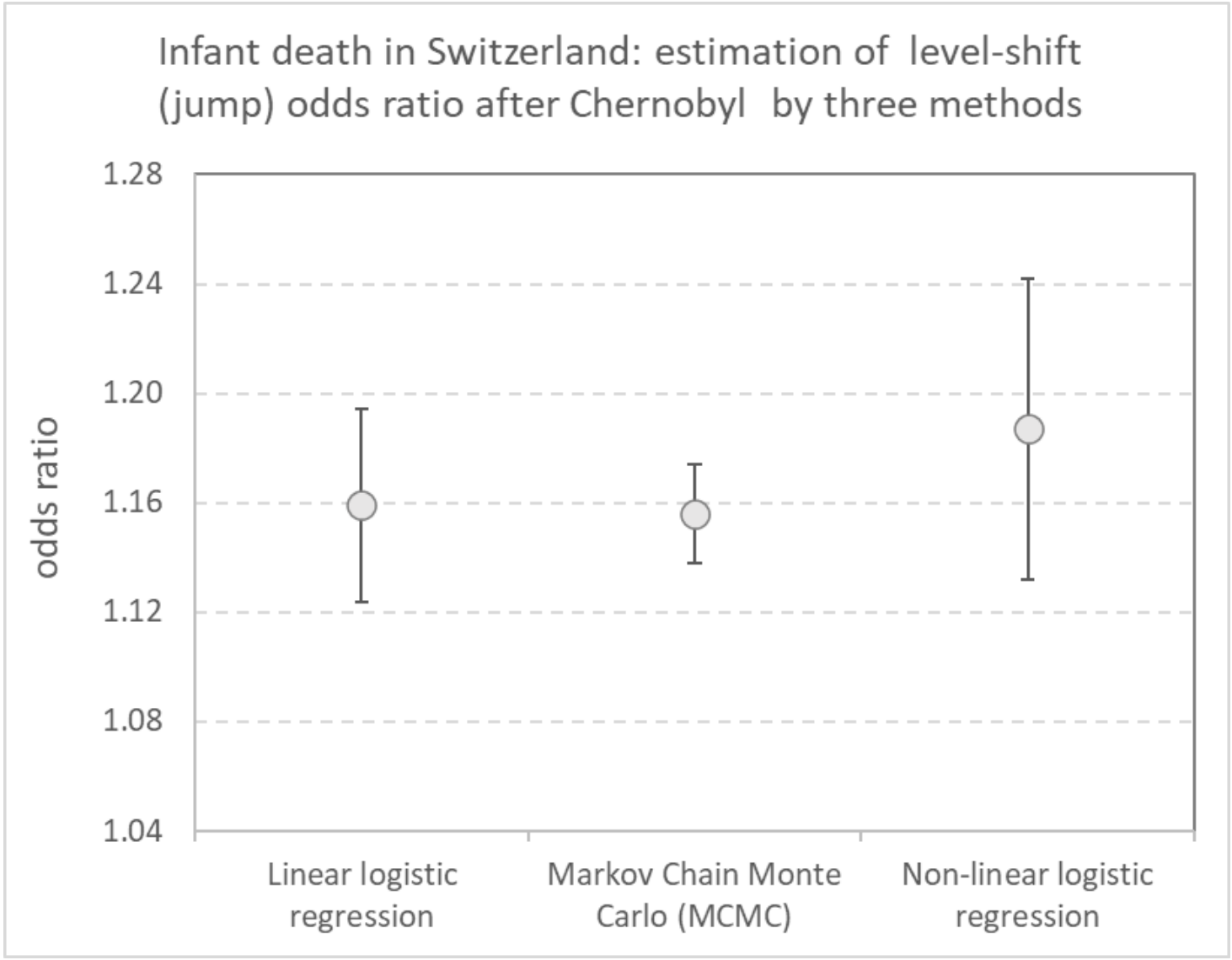
Graphical comparison of three change-point methods

Markov Chain Monte Carlo (MCMC) with 95%-confidence band for the level shift



Graphical comparison of three change-point methods

Non-linear Logistic Regression with 95%-confidence band for the level shift



Graphical comparison of three change-point methods