Structural Change in Regression Models: Detection, Estimation, and Model Goodness-of-Fit

Matilde Machado Gonçalves
Master in Mathematical Engineering
Department of Mathematics
Faculty of Sciences of the University of Porto
2026

Structural Change in Regression Models: Detection, Estimation, and Model Goodness-of-Fit

Matilde Machado Gonçalves

Dissertation carried out as part of the Master in Mathematical Engineering

Department of Mathematics

2026

Supervisor

Ana Rita Pires Gaio, Assistant Professor, Faculty of Sciences of the University of Porto

Acknowledgements

Começo por agradecer ao meu pai, o meu porto seguro. A pessoa que melhor me conhece, que me consegue ler só com um olhar, e que sem que eu diga nada sabe o que preciso. Foi ele que me ensinou que a calma é também uma forma de força, que me apoiou em todos os momentos difíceis e para quem eu corria quando precisava de um abraço. A nossa união é o maior presente que a vida me deu.

A toda a minha família, que de perto ou de longe, sempre acreditou em mim mais do que eu própria, por vezes acreditei. Em especial, à minha ”boadrasta” e à minha irmã, obrigada por fazerem parte da minha vida, por serem a melhor distração depois dos dias de trabalho e por, apesar do cansaço, me fazerem sempre sorrir. À minha avó materna, obrigada por teres sido a minha professora no momento em que mais precisei, e por me teres ajudado a construir a confiança que me faltava. À minha avó paterna, obrigada pelo apoio constante e por nunca teres deixado de estar do meu lado. Sem vocês nada disto seria possível.

Às grandes amigas que a faculdade me deu, Margarida e Eduarda, não tenho palavras suficientes para vos agradecer. Sem vocês, o meu percurso não teria sido o mesmo, não apenas nesta tese, mas todo o caminho até aqui. Agradeço-vos do fundo do coração, e sei, com uma certeza que não precisa de explicações, que vos levo comigo até ao fim da vida.

Agradeço a todos os professores que me acompanharam ao longo do meu percurso académico, obrigada por partilharem conhecimento mas também paciência e dedicação. Em particular à Professora Rita Gaio, o meu profundo agradecimento por todo o apoio, disponibilidade e dedicação ao longo deste ano.

Por último, agradeço à minha mãe. Foi ela que me ensinou a lutar, por mim, pelos meus sonhos e pela certeza de que só se vive uma vez, e por isso vale a pena viver com verdade e com coragem. Tenho o maior orgulho de poder dizer que ela é minha mãe e vivo todos os dias com o desejo de que, onde quer que ela esteja, se sinta orgulhosa de quem sou e de quem me estou a tornar. Esta tese, como tudo o resto que faço, é também para ela.

Resumo

Mudança estrutural refere-se a uma situação em que os elementos de um modelo estatístico não são estáveis ao longo da amostra, de tal forma que uma ou mais componentes do modelo, tipicamente os parâmetros de regressão, a função média condicional, ou a variância, sofrem alterações em localizações desconhecidas. Este fenómeno surge naturalmente em diversas áreas, como a genómica, as finanças, a medicina, ou as ciências sociais. Esta tese estabelece um tratamento unificado da estimação, deteção, e teste para a qualidade de ajustamento para modelos de regressão com pontos de corte, abrangendo tanto os modelos de regressão linear como Modelos Lineares Generalizados (GLM).

A estimação do ponto de corte é abordada através do estimador de Mínimos Quadrados Globais (Global Least Squares, GLS), no contexto linear e, para GLMs através do algoritmo de Linearização Iterativa, que evita a pesquisa exaustiva sobre as partições admissíveis que seria exigida por um algoritmo análogo, i.e. Máxima Verosimilhança Global. A deteção de mudanças estruturais é desenvolvida através de cinco famílias de testes, unificadas sob uma base comum: testes baseados na estatística F, o teste da razão de verosimilhanças, um teste de razão de verosimilhança empírica estendido ao contexto GLM, testes de flutuação baseados em somas cumulativas, e testes baseados em scores, juntamente com um procedimento sequencial para o caso de múltiplos pontos de corte. Com base no modelo de ponto de corte detetado, é desenvolvido um teste para a qualidade de ajustamento baseado em projeções, para avaliar se a especificação segmentada ajustada descreve adequadamente a estrutura da média condicional da resposta. Um estudo de simulação de Monte Carlo avalia o desempenho dos testes de hipoteses e procedimentos de estimação em amostras de tamanho finito, em modelos de regressão linear, logística, e de Poisson, sob diferentes intensidades de sinal e localizações do ponto de corte. Os resultados fornecem orientação prática para a seleção entre procedimentos de deteção e estimação, de acordo com o contexto, posição de ponto de corte, tipo de regressão e tempo de implementação computacional. Os resultados teóricos suportam, na generalidade, os resultados práticos obtidos ao longo do estudo de simulação.

Palavras-chave: Mudança estrutural, ponto de corte, regressão com ponto de corte, modelos lineares generalizados, verossimilhança empírica, testes de hipóteses, qualidade de ajustamento.

Abstract

Structural change refers to a situation in which elements of a statistical model are not stable over the sample, so that one or more components of the model, typically regression parameters, the conditional mean function, or the variance, shift at unknown locations. This phenomenon arises naturally in several fields, such as genomics, finance, medicine, or the social sciences. This thesis develops a unified treatment of estimation, detection, and goodness-of-fit testing for changepoint regression models, spanning both the linear regression models as well as the Generalized Linear Model (GLM) frameworks.

Breakpoint estimation is addressed through the Global Least Squares (GLS) estimator in the linear setting and, across GLMs, through the Iterative Linearization algorithm, which avoids the exhaustive search over admissible partitions required by a Global Maximum Likelihood analogue. Detection of structural change is developed across five families of tests, unified under a common framework: F-statistic based tests, the classical Likelihood Ratio Test, an Empirical Likelihood ratio test extended to the GLM setting, Cumulative Sum fluctuation tests, and score-based tests, together with a sequential procedure for the multiple-breakpoint case. Building on the detected changepoint model, a projection-based goodness-of-fit testing framework is developed to assess whether the fitted segmented specification adequately describes the conditional mean structure of the data. A Monte Carlo simulation study evaluates the finite-sample performance of the estimation and testing procedures across the linear, logistic, and Poisson regression families, under varying signal strength and breakpoint location. The results provide practical guidance for selecting among detection and estimation procedures according to regression setting, break type, and computational budget. The theoretical results mostly support the practical findings across the simulation procedure.

Keywords: Structural change, breakpoint, changepoint regression, generalized linear models, empirical likelihood, hypothesis tests, Goodness-of-Fit.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CUSUM Cumulative Sum. 4, 23, 24, 26, 41– 49, 61– 63, 73– 77, 83, 101

DGP Data Generating Process. 67, 75, 80, 84

EDF Exponential Dispersion Family. 3, 9– 11, 41, 44, 46– 48, 55, 61, 62, 97, 105

EL Empirical Likelihood. 4, 26, 34– 41, 61– 63, 73– 77, 83, 84

GLM Generalized Linear Model. 3– 5, 9– 11, 17, 20– 23, 25, 29, 32, 34, 38– 41, 44, 46– 51, 54– 56, 61– 64, 67, 69, 71, 82, 83, 90, 101– 103

GLMs Generalized Linear Models. 3, 10, 38, 40, 62, 71, 83

GLS Global Least Squares. 3, 17– 20, 22, 25, 60, 77, 79, 80, 82, 84, 92

GML Global Maximum Likelihood. 20, 84

GoF Goodness-of-Fit. 64, 68, 82

IRLS Iteratively Reweighted Least Squares. 4, 14, 17, 20, 21, 32, 33, 38, 41, 44– 46, 48– 52, 54– 56, 58, 59, 62, 84, 90, 91, 93, 97, 99, 101– 104

LRT Likelihood Ratio Test. 4, 23, 24, 26, 29, 31, 32, 34, 37– 41, 54, 55, 61– 63, 73– 77, 83, 84

ML Maximum Likelihood. 21, 22, 38, 91

MOSUM Moving Sum. 41, 47– 49, 61, 63, 73– 77, 83

OLS Ordinary Least Squares. 17, 18, 30, 35, 38, 41– 43, 45, 54, 61, 91, 92, 95, 96, 98, 100, 101

PBT Projection-based test. 67– 70, 83

RMEP Residual Marked Empirical Process. 65

RMPP Residual Marked Process Based on Projections. 66, 68

RMSE Root Mean Squared Error. 78, 79, 81

SD Standard Deviation. 78– 81

1 Introduction

1.1 Motivation

Structural change refers to a situation in which elements of a statistical model are not stable over the sample, so that one or more components of the model, typically regression parameters, the conditional mean function, or the variance, shift at unknown locations. In regression settings, this instability is often described through changepoints (also called break dates, breakpoints, structural change, thresholds, transition points and switch points). The associated relationship with breakpoints is said to be piecewise, segmented, broken-line, multi-phase or regime-change regressions. These terms are used interchangeably throughout this thesis. In a regression model exhibiting different trends along the predictors' domain, the goal is to detect whether changes have occurred, how many changepoints are there and to estimate the regression parameters.

Changepoint detection has become an important task in many applications. Detecting changepoints is common in many statistical and related areas including genomics (Wang et al., 2011, Aston and Kirch, 2012, Liu et al., 2020), financial revenue returns (Spokoiny, 2009, Bai, 2010, Cho and Fryzlewicz, 2015) among many others Zhou et al. (2025). In fact, structural changes arise naturally whenever populations, measurement systems, or behavioral mechanisms evolve. In medicine, the approval of a new treatment protocol may shift the relationship between a risk factor and patient outcomes, so that data collected before and after the change no longer follow the same pattern. In psychology, a traumatic life event may abruptly alter the trajectory of stress and mental health indicators, rendering a model estimated over the full sample a poor description of either phase. In finance, changes in market conditions may induce breaks in return dynamics. The common thread is that a single model, fit uniformly across the entire sample, would silently absorb these shifts into biased estimates, obscuring rather than revealing the underlying structure.

Inference with an unknown changepoint is fundamentally more difficult than classical fixed-parameter methods used to take into account non-linear effects, such as polynomial regression, regression splines and non-parametric smoothing, which are not suitable because the changepoints are fixed a priori (regression splines) or are not considered at all (smoothing splines and polynomial regression). When the break location is unknown, it

becomes a nuisance parameter that is typically not identified under the null hypothesis of no change. This destroys standard likelihood theory: likelihood ratio, Wald, and score statistics no longer have their usual χ^2 limits, and critical values must be obtained through nonstandard asymptotics or resampling. This phenomenon underlies the use of supremum or average-type tests over a range of admissible break locations. Furthermore, allowing for multiple breaks introduces a discrete, high dimensional segmentation problem: one must estimate both the number of breaks and their locations. Even in linear regression, searching over all partitions is computationally expensive, and statistical accuracy requires separating true breaks from spurious fluctuations. Practical procedures must balance statistical power, computational feasibility, and robustness to heteroskedasticity or other forms of misspecification.

Historically, the first tests for structural change were not designed to estimate a break point explicitly (Chow, 1960, Brown et al., 1975). The primary reason is that the asymptotic distribution theory for break date estimators, whether obtained by least-squares or maximum likelihood, was largely unavailable at the time, and tractable results existed only for a handful of special cases, such as those explored by Hawkins (1977) and Kim and Siegmund (1989).

A natural starting point is the two-phase linear regression model, in which the regression relationship changes at an unknown threshold. Early contributions by Quandt (1958, 1960) developed likelihood-based estimation and testing for two-regime systems, introducing the idea of scanning candidate breakpoints and selecting the one that maximizes a local evidence measure. This approach would become central to the field. The fundamental difficulty that the break is unidentified under the null hypothesis was addressed formally by Andrews (1993) and Andrews and Ploberger (1994), who established large-sample theory for parameter instability tests based on supremum and related functionals of local test statistics. Andrews et al. (1996) complemented this by deriving finite-sample optimal tests for the normal linear regression model, and Andrews (2001) extended inference to general parametric models with unknown breakpoints.

The extension to multiple structural breaks was developed systematically by Bai (1997) and Bai and Perron (1998), who provided key asymptotic theory for estimating and testing multiple breakpoints in linear models, including the widely used sup-type tests and sequential procedures. Bai and Perron (2003a) consolidated this framework, clarifying

consistency, convergence rates, and practical implementation. A key computational contribution came from Hawkins (2001), who showed that exhaustive search over all admissible partitions within the exponential family can be performed in an execution time that is only quadratic in the number of candidate changepoints, making the exact multiple-break estimation practically feasible. Subsequent work refined testing and estimation across broader environments, with Zeileis et al. (2003), Zeileis (2007) and Zeileis et al. (2010) contributing stability diagnostics and computational tools, and Perron and Zhu (2005), Lavielle (2005), Qu and Perron (2007), and Kejriwal and Perron (2008) addressing further generalizations and refinements of multiple-break inference.

Many applications involve non-Gaussian responses, where Generalized Linear Models (GLMs) provide a natural framework through an exponential-family likelihood and a link function. In this setting, a structural change may correspond to a shift in regression coefficients, a change in dispersion, or more generally a change in the conditional mean structure on the scale of the linear predictor. Compared to linear regression, changepoint estimation in GLMs introduces additional technical and computational difficulties. The objective function is no longer quadratic, iteratively reweighted fitting algorithms interact non-trivially with segmentation, and the distributional form can amplify the effects of leverage points or separation, most acutely in logistic regression. These challenges have motivated both classical parametric developments, and, more recently, work on multiple changepoint detection in high-dimensional GLM settings (Leonardi and Bühlmann, 2016).

1.2 Structure

This thesis is organized as follows. Chapter 2 introduces the changepoint regression model in two complementary settings. The linear partial change model is developed first, establishing the regime structure, the distinction between stable and regime-specific regressors, and the continuity and discontinuity conditions at the breakpoints. This framework is then extended to the GLM setting, where the conditional distribution of the response belongs to the Exponential Dispersion Family (EDF) and structural change enters through the linear predictor via a link function. The likelihood, score equations, Fisher information, and working quantities for the GLM changepoint model are derived and collected for reference throughout the thesis.

Chapter 3 develops estimation procedures for the changepoint regression model, treating the number of breaks m as known throughout. In the linear setting, the Global Least

Squares (GLS) estimator is obtained by exhaustive search over admissible partitions. In the GLM setting, the computational burden of Iteratively Reweighted Least Squares (IRLS) algorithm motivates an iterative linearization approach due to Muggeo (2003), which treats the breakpoints as parameters of the linear predictor and estimates them jointly with the regression coefficients through a Taylor expansion of the hinge function. The large-sample properties of the estimators, such as consistency, super-consistency of the breakpoint estimators, and asymptotic normality of the regression coefficients, are also established.

Chapter 4 develops a unified treatment of structural change tests across both the linear and GLM settings. Five families of tests for testing the existence of a breakpoint are covered: the F-based tests, the classical Likelihood Ratio Test (LRT), the Empirical Likelihood (EL) ratio test, the Cumulative Sum (CUSUM)-based tests, and, finally, the score-based tests. The chapter concludes with a unified comparison of all five families and a sequential procedure for estimating the number of breaks when m is unknown.

Chapter 5 develops a projection-based goodness-of-fit testing framework for the fitted segmented regression model. Having established in Chapter 4 that a break exists, the goal shifts to assessing whether the estimated changepoint model adequately describes the data. The framework applies to any parametric regression model, linear or Generalized Linear Model (GLM), standard or segmented, and yields consistent diagnostic tests for model adequacy.

Chapter 6 presents a simulation study comparing the finite-sample performance of the structural break detection methods developed in Chapters 4 and 5 across different scenarios with varying evidence for the existence of a break and different break locations. The evaluation of the estimation of the breakpoint through the methods explored in Chapter 3 is also conducted.

Finally, Chapter 7 presents an overall conclusion and possible directions for future work.

2 Changepoint Regression Model

This chapter introduces the changepoint regression model, in two complementary settings. Section 2.1 develops the framework within linear regression, while Section 2.2 extends this framework to the GLM setting. These formulations establish the theoretical foundation for the estimation and testing procedures developed in the Chapters 3 and 4.

2.1 Linear Changepoint Regression Model

Consider a sample of n observations partitioned into $m+1$ regimes by m unknown breakpoints. Let $\psi = (\psi_1, \ldots, \psi_m)$ denote the vector of breakpoints, representing the unknown locations at which the regression relationship may change, satisfying

$$\psi_0 \leq \psi_1 \leq \cdots \leq \psi_m \leq \psi_{m+1},$$

where $\psi_0 = 0$ and $\psi_{m+1} = n$ mark the beginning and the end of the sample, respectively. The j-th regime, dependent on the breakpoints, is

$$R_j(\psi) = \{i : \psi_{j-1} < i \leq \psi_j\}, \qquad j = 1, \ldots, m+1,$$

and the regimes $R_1(\psi), \ldots, R_{m+1}(\psi)$ form a partition of $\{1, \ldots, n\}$. The linear changepoint regression model is specified as

$$y_i = x_i'\beta + z_i'\delta_j + u_i, \quad i \in R_j(\psi), \quad j = 1, \ldots, m+1,$$

where i indexes the observations and j indexes the regimes.

Here, the dependent variable is denoted by y_i. The vector $x_i \in \mathbb{R}^p$ contains predictors whose associated coefficient vector $\beta \in \mathbb{R}^p$ is stable across all regimes. The vector $z_i \in \mathbb{R}^q$ contains regressors whose associated coefficient vector $\delta_j \in \mathbb{R}^q$ is allowed to vary across regimes and is, therefore, regime-specific. The error term u_i is assumed to satisfy standard regularity conditions: zero conditional mean, $E(u_i \mid x_i, z_i) = 0$; finite variance, $\mathrm{Var}(u_i) = \sigma_j^2 < \infty$ within regime j; no serial correlation across observations; and finite moments beyond the second order. Homoskedasticity across regimes is imposed only where explicitly required, and the design matrices are assumed to have full column rank within each admissible regime.

This model is termed a partial change model because not all parameters are subject to shifts. The coexistence of stable and unstable regressors is common in empirical applications and yields potential savings in degrees of freedom, a consideration of particularly relevance when multiple breakpoints are considered. The treatment of the intercept is a modeling choice, it can be assumed to be stable or varying across regimes, being included in x_i or z_i, respectively.

The breakpoint vector $\psi = (\psi_1, \ldots, \psi_m)$, the stable coefficients β, and the regime-specific coefficients $\delta_1, \ldots, \delta_{m+1}$ are all treated as unknown and must be estimated jointly from the n observations (y_i, x_i, z_i). In most practical applications, the number of breaks m is itself unknown and must be inferred from the data, with its true value denoted as m^0.

Examples: Consider the model with one breakpoint ($m = 1$) at $\psi_1 = 12$, so that $n = 20 = \psi_2$ and there are two regimes, indexed as $j = 1$ and $j = 2$.

- If all regressors are allowed to vary across regimes, then $x_i = 0$, and there is no intercept. The model reduces to

$$y_i = \begin{cases} z_i\delta_1 + u_i, & i = 1, \ldots, 12 \\ z_i\delta_2 + u_i, & i = 13, \ldots 20 \end{cases}$$

- If there is a constant intercept, one stable regressor x_i, and one regime-specific regressor z_i, the model becomes

$$y_i = \begin{cases} \beta_1 + \beta_2 x_i + z_i\delta_1 + u_i, & i = 1, \ldots, 12 \\ \beta_1 + \beta_2 x_i + z_i\delta_2 + u_i, & i = 13, \ldots 20 \end{cases}$$

- If the intercept is allowed to shift, so it is absorbed in z_i, there is still one regressor that is allowed to vary across regimes and another that is stable. The model becomes

$$y_i = \begin{cases} \alpha_1 + x_i\beta + z_i\delta_1 + u_i, & i = 1, \ldots, 12 \\ \alpha_2 + x_i\beta + z_i\delta_2 + u_i, & i = 13, \ldots 20 \end{cases}$$

2.1.1 Matrix notation

For a fixed partition $\psi = (\psi_1, \ldots, \psi_m)$, the model in matrix form is $Y = X\beta + Z\delta + U$, where

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n, \qquad U = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \in \mathbb{R}^n$$

The matrix of stable regressors X, with components $x_i \in \mathbb{R}^p$ and the associated coefficient vector of stable coefficients β are

$$X = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix} \in \mathbb{R}^{n \times p}, \qquad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \in \mathbb{R}^p$$

Unlike X, the matrix Z of regime-specific regressors depends on the breakpoints. For a given partition ψ, it is the block-diagonal matrix, given by

$$Z = \begin{pmatrix} Z_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Z_{m+1} \end{pmatrix} \in \mathbb{R}^{n \times q(m+1)}, \qquad Z_j = \begin{pmatrix} z'_{\psi_{j-1}+1} \\ \vdots \\ z'_{\psi_j} \end{pmatrix} \in \mathbb{R}^{(\psi_j - \psi_{j-1}) \times q}$$

The stacked vector of regime-specific coefficients is

$$\delta = \begin{pmatrix} \delta_1' \\ \vdots \\ \delta_{m+1}' \end{pmatrix} \in \mathbb{R}^{(m+1)q \times 1},$$

The full design matrix can be constructed as

$$X^* = (X \mid Z) = \begin{pmatrix} x_1' & Z_1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_n' & 0 & \ldots & Z_{m+1} \end{pmatrix} \in \mathbb{R}^{n \times (p+q(m+1))}$$

Let β^0 and δ^0 denote the true regression coefficients, and let $\psi^0 = (\psi_1^0, \ldots, \psi_m^0)$ denote the true break dates. The estimation objective is to recover $(\beta^0, \delta_1^0, \ldots, \delta_{m+1}^0, \psi_1^0, \ldots, \psi_m^0)$, from the observed data. The number of breaks m itself may be unknown, with true value denoted by m^0.

Example: Consider again the model with $m = 1$ at $\psi_1 = 12$ and $n = 20 = \psi_2$.

- Whenever all regressors are allowed to vary across regimes, $X = 0$ and there is no intercept, the matrix representation is

$$Z = \begin{pmatrix} z_1 & 0 \\ \vdots & \vdots \\ z_{12} & 0 \\ 0 & z_{13} \\ \vdots & \vdots \\ 0 & z_{20} \end{pmatrix}, \ \delta = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$$

- Whenever there is a constant intercept, one stable regressor x_i, and one regime-specific regressor z_i, the matrix representation is

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_{20} \end{pmatrix} \ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \ Z = \begin{pmatrix} z_1 & 0 \\ \vdots & \vdots \\ z_{12} & 0 \\ 0 & z_{13} \\ \vdots & \vdots \\ 0 & z_{20} \end{pmatrix}, \ \delta = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$$

- Whenever the intercept is allowed to shift, one stable regressor, x_i, and one regime-specific regressor, z_i, the model reduces to

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_{20} \end{pmatrix}, \ \beta = \beta_1, \ Z = \begin{pmatrix} 1 & z_1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & z_{12} & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & z_{13} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & z_{20} \end{pmatrix}, \ \delta = \begin{pmatrix} \alpha_1 \\ \delta_1 \\ \alpha_2 \\ \delta_2 \end{pmatrix}$$

2.1.2 Continuity and Discontinuity

The partial change model is, by default, a discontinuous specification. The fitted values in regimes j and $j+1$ are, respectively

$$\hat{y}_i^{(j)} = x_i'\beta + z_i'\delta_j, \qquad \hat{y}_i^{(j+1)} = x_i'\beta + z_i'\delta_{j+1}$$

Since β is constant across regimes, the difference between consecutive fitted values at the breakpoint is $\hat{y}_{\psi_j}^{(j+1)} - \hat{y}_{\psi_j}^{(j)} = z'_{\psi_j}(\delta_{j+1} - \delta_j)$. The model is said to be continuous at ψ_j if this difference is zero for each $j = 1, \ldots, m$. When this condition holds, the regression function exhibits a kink, with a change in slope but no jump in level, at the breakpoint. Many procedures restrict attention to the continuous case, though the discontinuous specification is more general and does not require the continuity constraints to be imposed a priori.

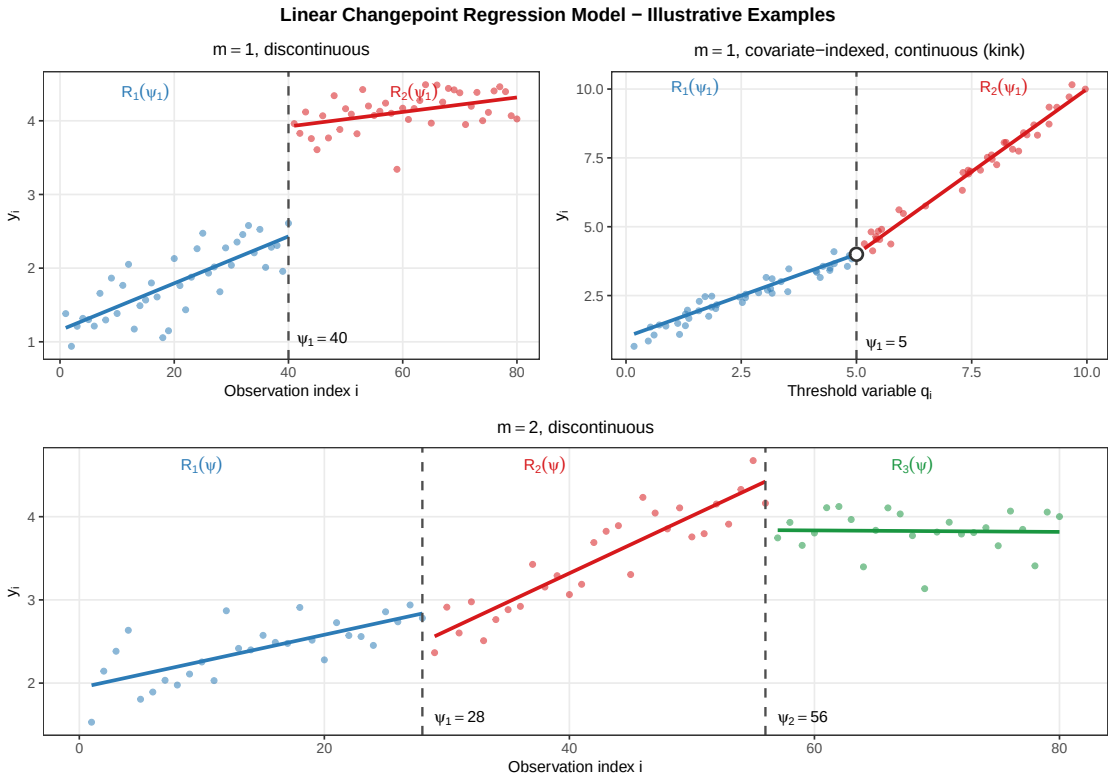

Figure 1: Examples of Linear Changepoint Regressions

2.2 Generalized Linear Changepoint Regression Model

The linear partial changepoint regression model assumes that the response y_i is continuous and that the conditional mean is a linear function of the regressors. While this covers a wide range of applications and admits a particularly tractable analytical structure, many empirical problems involve responses that are inherently non-Gaussian: binary outcomes in epidemiology and medicine, count data in ecology and economics, or positive-valued continuous measurements in actuarial science. In such cases, the normality and linearity assumptions underlying the linear model are inappropriate, and a more general framework is required. The present section extends the partial change framework to the GLM setting, in which the conditional distribution of y_i belongs to the Exponential Dispersion

Family (EDF) and structural change is introduced through the linear predictor via a known link function. As will become clear, the two frameworks are formally analogous in structure, with the linear model emerging as the special case of the GLM model under a normal distribution and the identity link.

2.2.1 Exponential Dispersion Family (EDF)

GLMs were introduced by Nelder and Wedderburn (1972) with the intent to extend classical linear regression by allowing the conditional distribution of y_i given the predictors to belong to the EDF. A distribution belongs to the EDF if its probability mass function or density can be written in canonical form as

$$f(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right),$$

where $\theta_i \in \mathbb{R}$ is the natural (canonical) parameter, $\phi > 0$ is the dispersion parameter, common across observations, $b(\cdot)$ is the cumulant function, a known twice-differentiable convex function that determines the mean and variance of y_i, $a(\phi)$ is a known positive function of the dispersion parameter, and $c(y,\phi)$ is a normalizing function that does not depend on θ.

Differentiating the log-density with respect to θ and using the fact that a valid density integrates to one yields

$$E(y_i \mid \theta_i) = b'(\theta_i), \qquad \mathrm{Var}(y_i \mid \theta_i) = a(\phi)\, b''(\theta_i)$$

The quantity $V(\mu_i) = b''(\theta_i)$, evaluated at $\mu_i = b'(\theta_i)$, is a defining characteristic of each distribution, called the variance function, and characterizes how the conditional variance depends on the conditional mean. Together, $V(\mu_i)$ and $a(\phi)$ determine the mean-variance relationship:

$$\mathrm{Var}(y_i \mid \theta_i) = a(\phi)\, V(\mu_i).$$

The following table summarizes the most commonly used members of the EDF, together with their cumulant functions, variance functions, and canonical parameters and dispersion specifications.

Distribution	$b(\theta)$	$V(\mu)$	**Canonical** θ	$a(\phi)$
Normal	$\theta^2/2$	1	μ	σ^2
Poisson	e^θ	μ	$\log\mu$	1
Bernoulli	$\log(1+e^\theta)$	$\mu(1-\mu)$	$\log\frac{\mu}{1-\mu}$	1
Gamma	$-\log(-\theta)$	μ^2	$-1/\mu$	ϕ
Inverse-Gaussian	$-\sqrt{-2\theta}$	μ^3	$-1/(2\mu^2)$	ϕ

Table 1: Common members of the exponential dispersion family.

2.2.2 GLM Changepoint Model

The linear partial changepoint regression model extends directly to the GLM framework, while retaining the same partial change structure, where x_i enters with stable coefficients across regimes and z_i is allowed to shift at the unknown breakpoints.

Suppose that, conditional on (x_i, z_i), the response y_i belongs to the EDF with conditional mean $\mu_i = E(y_i \mid x_i, z_i)$ and dispersion parameter ϕ. Let $g(\cdot)$ denote the known, monotone, twice-differentiable link function with inverse $g^{-1}(\cdot)$, and η_i the linear predictor, related to the conditional mean by

$$\eta_i = g(\mu_i) \quad \Leftrightarrow \quad \mu_i = g^{-1}(\eta_i)$$

When $g(\mu_i) = \theta_i$, the link is said to be canonical, and the natural parameter coincides with the linear predictor.

For a given set of m breakpoints $\psi = (\psi_1, \ldots, \psi_m)$, the GLM partial change model is specified as

$$\eta_i = x_i'\beta + z_i'\delta_j = g(\mu_i), \quad i \in R_j(\psi), \quad j = 1, \ldots, m+1,$$

where the regimes $R_j(\psi)$ are defined as in Section 2.1. The vectors x_i and z_i, the coefficient vectors β and δ_j, and the block-diagonal structure of the design matrix carry over without alteration from the linear setting. x_i enters with stable coefficients β common across all regimes, while z_i enters with regime-specific coefficients δ_j that are free to shift at the breakpoints. The structural change is thus expressed entirely on the scale of the linear predictor.

This specification is a direct generalization of the linear partial change model. In particular, when $y_i \mid x_i, z_i \sim N(\mu_i, \sigma^2)$ and $g(\cdot)$ is the identity link, the linear predictor satisfies $\eta_i = \mu_i$ and the model reduces to the linear partial changepoint regression with $u_i \sim \mathrm{N}(0, \sigma^2)$.

2.2.3 Likelihood, Score, Information

2.2.3.1 Likelihood

Given n independent observations and a fixed partition $(\psi_1, \ldots, \psi_m)$, the likelihood factors over all observations as

$$\mathcal{L}(\beta, \delta, \phi \mid \psi_1, \ldots, \psi_m) = \prod_{j=1}^{m+1} \prod_{i \in R_j(\psi)} \exp\left(\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right),$$

where for $i \in R_j(\psi)$, the natural parameter θ_i is determined implicitly by $b'(\theta_i) = \mu_i = g^{-1}(\eta_i) = g^{-1}(x_i'\beta + z_i'\delta_j)$.

The total log-likelihood decomposes additively over regimes as

$$\ell(\beta, \delta, \phi \mid \psi_1, \ldots, \psi_m) = \sum_{j=1}^{m+1} \sum_{i \in R_j(\psi)} \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right] = \sum_{j=1}^{m+1} \ell_j(\beta, \delta_j, \phi \mid \psi),$$

where each regime log-likelihood is

$$\ell_j(\beta, \delta_j, \phi \mid \psi) = \sum_{i \in R_j(\psi)} \left[\frac{y_i \theta_i(\beta, \delta_j) - b(\theta_i(\beta, \delta_j))}{a(\phi)} + c(y_i, \phi) \right].$$

2.2.3.2 Score Equations and Working Quantities

The score function is the gradient of the log-likelihood function with respect to the parameter of interest. The derivative of the log-likelihood with respect to the linear predictor η_i is obtained by applying the chain rule

$$\frac{\partial \ell_i}{\partial \eta_i} = \frac{\partial \ell_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i}$$

Recalling that $\mu_i = g^{-1}(\eta_i)$ and $b'(\theta_i) = \mu_i$ and differentiating each component yields

$$\frac{\partial \ell_i}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)} = \frac{y_i - \mu_i}{a(\phi)}, \qquad \frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b''(\theta_i)} = \frac{1}{V(\mu_i)} \qquad \frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{g'(\mu_i)},$$

Combining these expressions, the score contribution of observation i with respect to η_i is

$$\frac{\partial \ell_i}{\partial \eta_i} = \frac{y_i - \mu_i}{a(\phi)} \frac{1}{V(\mu_i)} \frac{1}{g'(\mu_i)}$$

Score with respect to β Since $\partial\eta_i/\partial\beta = x_i$ for all i, the score contribution of observation $i \in R_j(\psi)$ with respect to β is given by the chain rule

$$\frac{\partial \ell_i}{\partial \beta} = \frac{(y_i - \mu_i)x_i}{a(\phi)V(\mu_i)g'(\mu_i)}$$

Summing over all observations the total score with respect to β is

$$s_\beta(\beta, \delta, \phi \mid \psi) = \sum_{j=1}^{m+1} \sum_{i \in R_j(\psi)} \frac{(y_i - \mu_i) \cdot x_i}{a(\phi)V(\mu_i)g'(\mu_i)}$$

Defining the diagonal matrices

$$G = \text{diag}(g'(\mu_1), \ldots, g'(\mu_n)), \quad \text{where } g'(\cdot) \text{ are the link derivatives}$$

$$W = \text{diag}(w_1, \ldots, w_n), \quad \text{with weights} \quad w_i = \frac{1}{a(\phi)V(\mu_i)(g'(\mu_i))^2},$$

the score with respect to β takes the compact matrix form

$$s_\beta(\beta, \delta, \phi \mid \psi) = X'WG(Y - \mu)$$

Score with respect to δ_j Since $\partial\eta_i/\partial\delta_j = z_i$ for $i \in R_j(\psi)$ and zero otherwise, the score with respect to the regime-specific coefficient δ_j involves only observations in regime $R_j(\psi)$

$$s_{\delta_j}(\beta, \delta_j, \phi \mid \psi) = \sum_{i \in R_j(\psi)} \frac{(y_i - \mu_i) \cdot z_i}{a(\phi)V(\mu_i)g'(\mu_i)}$$

For the regime-specific coefficients, the score with respect to δ_j is

$$s_{\delta_j}(\beta, \delta_j, \phi \mid \psi) = Z_j'W_jG_j(Y_j - \mu_j),$$

where Y_j, μ_j, W_j, G_j denote the restrictions of Y, μ, W and G to observations in regime $R_j(\psi)$.

Working residuals and compact score form These expressions simplify considerably by introducing the working residual and working response for observation i,

$$r_i^* = (y_i - \hat{\mu}_i)\, g'(\hat{\mu}_i), \qquad y_i^* = \hat{\eta}_i + r_i^*,$$

or in vector form, $R^* = G(Y - \hat{\mu})$ and $Y^* = \hat{\eta} + G(Y - \hat{\mu}) = \hat{\eta} + R^*$. Using these quantities, the score equations can be rewritten compactly as

$$s_\beta(\beta, \delta, \phi \mid \psi) = X'WG(Y - \mu) = X'WR^*, \qquad s_{\delta_j}(\beta, \delta_j, \phi \mid \psi) = Z_j'W_jG_j(Y_j - \mu_j) = Z_j'W_jR_j^*,$$

where R_j^* denotes the restriction of R^* to observations in $R_j(\psi)$.

Weighted form and hat matrix The score equations motivate a decomposition of the weighted working residuals. Define the weighted working response, weighted working residuals, and weighted design matrix, respectively, by

$$Y_w^* = W^{1/2}Y^*, \qquad R_w^* = W^{1/2}R^*, \qquad X_w^* = W^{1/2}X^*$$

The weighted hat matrix is defined as

$$H_w = W^{1/2}X^*(X^{*'}WX^*)^{-1}X^{*'}W^{1/2} = X_w^*\ \mathcal{I}^{-1}X_w^{*'}$$

The matrix H_w is symmetric and idempotent ($H_w^2 = H_w$) and projects the weighted working response Y_w^* onto the column space of X_w^* in the standard Euclidean inner product. Its complement $(I_n - H_w)$ is likewise symmetric and idempotent and projects onto the orthogonal complement, extracting the part of the weighted working response that is not explained by the design matrix.

Using the weighted quantities, the score equations can be re-written as

$$s_\beta(\beta, \delta, \phi \mid \psi) = X^{'}WR^* = X^{'}W^{1/2}W^{1/2}R^* = (W^{1/2}X)'R_w^* = X_w'R_w^*$$

$$s_{\delta_j}(\beta, \delta_j, \phi \mid \psi) = Z_j'W_jR_j^* = Z_j'W_j^{1/2}W_j^{1/2}R_j^* = (W_j^{1/2}Z_j)'(R_j)_w^* = (Z_j)_w'(R_j)_w^*$$

where $R_w^* = W^{1/2}R^* = (I_n - H_w)Y_w^*$.

The connection between H_w and the IRLS algorithm, which generates H_w as the projection matrix of the weighted least squares step at convergence, is established in the Appendix 7.

2.2.3.3 Fisher Information

The Fisher Information $\mathcal{I}$ with respect to a parameter of interest is given by the negative expected value of the second derivative of the log-likelihood function. We derive it here for each block of $(\beta, \delta_1, \ldots, \delta_{m+1})$.

Second derivative of the log-likelihood As previously derived,

$$\frac{\partial \ell_i}{\partial \eta_i} = \frac{y_i - \mu_i}{a(\phi)} \cdot h(\mu_i), \qquad h(\mu_i) = \frac{1}{V(\mu_i) g'(\mu_i)}$$

Since $\frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{g'(\mu_i)}$, the second derivative with respect to η_i is

$$\frac{\partial^2 \ell_i}{\partial \eta_i^2} = \frac{1}{a(\phi)} \left[-(y_i - \mu_i) \frac{\partial h}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} - h(\mu_i) \frac{\partial \mu_i}{\partial \eta_i} \right] = \frac{1}{a(\phi)} \left[-\frac{(y_i - \mu_i)}{g'(\mu_i)} \frac{\partial h}{\partial \mu_i} - \frac{h(\mu_i)}{g'(\mu_i)} \right]$$

Taking expectations and using $E(y_i - \mu_i) = 0$ provides

$$E\left[\frac{\partial^2 \ell_i}{\partial \eta_i^2} \right] = \frac{1}{a(\phi)} \cdot E\left[-\frac{(y_i - \mu_i)}{g'(\mu_i)} \frac{\partial h}{\partial \mu_i} - \frac{h(\mu_i)}{g'(\mu_i)} \right] = \frac{1}{a(\phi)} \cdot \left[-\frac{h(\mu_i)}{g'(\mu_i)} \right] = -\frac{1}{a(\phi) V(\mu_i) (g'(\mu_i))^2} = -w_i$$

Fisher Information for β Applying the chain rule,

$$\frac{\partial^2 \ell_i}{\partial \beta \ \partial \beta'} = \frac{\partial^2 \ell_i}{\partial \eta_i^2} x_i x_i',$$

so the Fisher information with respect to the stable coefficients is

$$\mathcal{I}_{\beta\beta} = -E\left[\frac{\partial^2 \ell}{\partial \beta \ \partial \beta'} \right] = \sum_{j=1}^{m+1} \sum_{i \in R_j(\psi)} w_i \ x_i \ x_i' = X' W X$$

Fisher Information for δ_j Analogously,

$$\frac{\partial^2 \ell_i}{\partial \delta_j \ \partial \delta_j'} = \frac{\partial^2 \ell_i}{\partial \eta_i^2} z_i z_i',$$

and since only observations in regime $R_j(\psi)$ contribute to δ_j,

$$\mathcal{I}_{\delta_j \delta_j} = -E\left[\frac{\partial^2 \ell}{\partial \delta_j \ \partial \delta_j'} \right] = \sum_{i \in R_j(\psi)} w_i \ z_i \ z_i' = Z_j' W_j Z_j$$

Cross-block: β and δ_j. For $i \in R_j(\psi)$,

$$\frac{\partial^2 \ell_i}{\partial \beta \ \partial \delta_j'} = \frac{\partial^2 \ell_i}{\partial \eta_i^2} x_i z_i', \quad i \in R_j(\psi),$$

while, for $i \notin R_j$ the log-likelihood does not depend on δ_j, so those terms vanish. Hence,

$$\mathcal{I}_{\beta \delta_j} = -E\left[\frac{\partial^2 \ell}{\partial \beta \ \partial \delta_j'} \right] = \sum_{i \in R_j(\psi)} w_i \ x_i \ z_i' = X_j' W_j Z_j$$

Cross-block: δ_j and δ_k, $j \neq k$. Since the regimes are disjoint, no observation belongs to both $R_j(\psi)$ and $R_k(\psi)$, so

$$\frac{\partial^2 \ell_i}{\partial \delta_j \; \partial \delta_k'} = 0 \quad \Rightarrow \quad \mathcal{I}_{\delta_j \delta_k} = 0 \qquad \text{for } j \neq k$$

Full Fisher information matrix Assembling these block, the full Fisher information matrix for $(\beta, \delta_1, \ldots, \delta_{m+1})$ is

$$\mathcal{I}(\beta, \delta_1, \ldots, \delta_{m+1}) = \begin{pmatrix} X'WX & X_1'W_1Z_1 & \cdots & X_{m+1}'W_{m+1}Z_{m+1} \\ Z_1'W_1X & Z_1'W_1Z_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ Z_{m+1}'W_{m+1}X & 0 & \cdots & Z_{m+1}'W_{m+1}Z_{m+1} \end{pmatrix} = X^{*'}WX^*,$$

where $X^* = (X \mid Z)$. The block-diagonal structure of the δ-blocks reflects the regime separation: each δ_j is informed solely by observations in $R_j(\psi)$.

3 Estimation of the Changepoints

Chapter 2 established the partial change model in both the linear and GLM settings, treating the breakpoints $\psi = (\psi_1, \ldots, \psi_m)$ as unknown parameters to be recovered from the data alongside the regression coefficients. The present chapter develops estimation procedures for both frameworks.

For a fixed set of breakpoints, the model reduces to a piecewise standard model in each regime and the regression coefficients can be estimated by standard methods, such as the Ordinary Least Squares (OLS) in the linear case, or IRLS in the GLM case. This yields an objective function depending only on ψ, which is optimized over all admissible partitions. The implementation, however, differs substantially between the two frameworks. In the linear model, the OLS objective is a sum of squared residuals admitting a closed-form solution for each candidate breakpoint partition, making the search computationally tractable. This is the Global Least Squares (GLS) approach of Bai and Perron (1998), developed in Section 3.1. In the GLM setting, the objective function is a log-likelihood and the regression coefficients for each candidate partition must be estimated iteratively via the IRLS algorithm, making an exhaustive search over all admissible partitions through a computationally demanding procedure. Muggeo (2003) proposes an alternative approach that avoids this search by treating the breakpoints as parameters of the linear predictor and estimating them jointly with the regression coefficients through an iterative linearization scheme. This is developed in Section 3.2.

Throughout this chapter, the number of breaks m is treated as known. When m is unknown, a sequential testing and estimation procedure applies, as discussed in Chapter 4.

3.1 Linear Model: Global Least Squares

Conditional on knowing the breakpoint locations, the linear partial change model reduces to $m + 1$ segments of standard linear regressions fitted on disjoint subsamples, so the regression coefficients can be estimated by OLS within each regime. The difficulty is that the breakpoints are unknown and must be estimated jointly with the regression coefficients. The estimation strategy addresses this by evaluating all admissible partitions of the sample, given a trimming condition, and selecting the partition that minimizes the

concentrated sum of squared residuals. This approach is commonly referred to as Global Least Squares (GLS) estimation.

3.1.1 OLS Estimation with Known Breakpoints

Given a fixed partition $0 = \psi_0 < \psi_1 < \cdots < \psi_m < \psi_{m+1} = n$, the OLS objective and the conditional OLS estimators of the regression coefficients are defined as

$$S(\beta, \delta \mid \psi) = \sum_{j=1}^{m+1} \sum_{i \in R_j(\psi)} \left(y_i - x_i'\beta - z_i'\delta_j\right)^2, \qquad \left(\hat{\beta}(\psi),\, \hat{\delta}(\psi)\right) = \underset{\beta,\delta}{\arg\min}\ S(\beta, \delta \mid \psi)$$

Since the regimes are disjoint and the matrix Z is block-diagonal, the minimization over $(\beta, \delta_1, \ldots, \delta_{m+1})$ reduces to a single regression on the full design matrix $X^* = (X \mid Z)$, with closed-form solution

$$\begin{pmatrix} \hat{\beta}(\psi) \\ \hat{\delta}(\psi) \end{pmatrix} = \left(X^{*\prime} X^*\right)^{-1} X^{*\prime} Y.$$

3.1.2 Trimming Conditions and Admissible Partitions

The estimator of the breakpoints $\psi_1, \ldots, \psi_m$ is obtained by searching over all admissible partitions of the sample, given a trimming condition. The trimming condition is imposed to ensure that each regime contains a sufficient number of observations for estimation and to prevent breakpoint estimates from being too close to the sample boundaries or to one another. For a trimming fraction $\varepsilon \in (0, 1)$, set $h = \lfloor \varepsilon n \rfloor$ and require

$$\psi_j - \psi_{j-1} \geq h, \qquad j = 1, \ldots, m+1$$

Denote by $\Psi_{\varepsilon,m}$ the set of all m-partitions satisfying this condition.

3.1.3 Global Least Squares Estimator

For each $\psi \in \Psi_{\varepsilon,m}$, substituting $(\hat{\beta}(\psi), \hat{\delta}(\psi))$ back into the OLS objective gives the concentrated sum of squared residuals $S(\psi) = S(\hat{\beta}(\psi), \hat{\delta}(\psi) \mid \psi)$, which depends on the partition only. The GLS estimator of the breakpoints is the partition that minimizes this concentrated objective

$$(\hat{\psi}_1, \ldots, \hat{\psi}_m) = \underset{(\psi_1, \ldots, \psi_m) \in \Psi_{\varepsilon,m}}{\arg\min} S(\psi_1, \ldots, \psi_m),$$

with corresponding coefficient estimates $\hat{\beta} = \hat{\beta}(\hat{\psi})$ and $\hat{\delta}_j = \hat{\delta}_j(\hat{\psi})$ for $j = 1, \ldots, m+1$.

The pseudo-algorithm for the GLS estimation is exposed in Appendix 7. The `R` package `strucchange` (Zeileis et al., 2002) implements this directly via the function `breakpoints()`, with `breakdates()` and `coef()` extracting the estimated breakpoints and the regime-specific coefficient estimates, respectively.

3.1.4 Large Sample Properties

Bai and Perron (1998) establish the large-sample properties of the GLS estimators. The asymptotic results rest on regularity assumptions that ensure identification of the breakpoints, well-behaved behavior of the least squares objective function, and the validity of the asymptotic theory for consistency and convergence rates.

Let m^0 denote the true number of breaks, $\lambda_j^0 = \psi_j^0/n$ the true break fractions, and $\hat{\lambda}_j = \hat{\psi}_j/n$ the estimated break fractions.

Consistency Under the maintained assumptions, the GLS estimators of the break fractions are consistent,

$$\hat{\lambda}_j \xrightarrow{p} \lambda_j^0 \quad \Longleftrightarrow \quad \frac{\hat{\psi}_j - \psi_j^0}{n} \xrightarrow{p} 0 \qquad j = 1, \ldots, m^0,$$

and the regression coefficient estimators are likewise consistent: $\hat{\beta} \xrightarrow{p} \beta^0$ and $\hat{\delta}_j \xrightarrow{p} \delta_j^0$ for $j = 1, \ldots, m^0 + 1$. Consistency of $\hat{\beta}$ and $\hat{\delta}_j$ follows since, given consistent breakpoint estimates $\hat{\psi}_j$, the regime membership of each observation is correctly classified with probability approaching one, and standard least squares arguments apply within each regime.

Rate of convergence The breakpoint estimators converge at a faster rate than the regression coefficients,

$$\hat{\psi}_j - \psi_j^0 = O_p(1), \qquad \hat{\beta} - \beta^0 = O_p(n^{-1/2}), \qquad \hat{\delta}_j - \delta_j^0 = O_p(n^{-1/2})$$

The estimation of $\hat{\psi}_j$ is super-consistent, meaning that the estimation error $\hat{\psi}_j - \psi_j^0$ is bounded in probability, remaining of order $O_p(1)$, rather than diverging with n.

3.2 GLM: Iterative Linearization

In the GLM partial change model, the breakpoints enter the log-likelihood nonlinearly through the regime allocation of each observation (Section 2.2). A Global Maximum Likelihood (GML) analogue of the GLS approach is in principle available, where, instead of the minimized residual sum of squares the maximized log-likelihood is used. For each candidate $\psi \in \Psi_{\varepsilon,m}$, the regression coefficients are estimated via the IRLS algorithm to convergence, yielding the concentrated log-likelihood $\ell(\psi)$ which must be maximized over $\Psi_{\varepsilon,m}$. However, each partition requires a full IRLS run to convergence, making the exhaustive search quickly computationally prohibitive, particularly when m is unknown or n is large. Muggeo (2003) proposes an alternative that avoids this search entirely. (Pseudo algorithm in the Appendix 7)

3.2.1 Estimation of the breakpoints via Iterative Linearization

The approach is developed for the continuous partial change model, requiring $z'_{\psi_j}(\delta_j - \delta_{j+1}) = 0$ at each breakpoint (Section 2.1.2). Under this continuity restriction, the regime-specific coefficients satisfy $\delta_{j+1} = \delta_j + \xi_j$ for a change-in-slope vector ξ_j at ψ_j, and the GLM partial change model reparametrizes as

$$\eta_i = x_i'\beta + z_i'\delta_1 + \sum_{j=1}^{m} \xi_j \,(z_i - \psi_j)_+, \qquad i = 1, \ldots, n,$$

where $(z_i - \psi_j)_+ = (z_i - \psi_j)\,\mathbf{1}(z_i > \psi_j)$ is the hinge function at ψ_j. The slope of z_i in regime 1 is δ_1, and in regime $j+1$ is $\delta_1 + \sum_{l=1}^{j} \xi_l$. The breakpoints enter the model nonlinearly through the hinge functions.

3.2.1.1 Taylor Linearization

Writing the linear predictor as a sum of hinge functions, as previously derived, decomposes η_i into a sum of a linear component $x_i'\beta + z_i'\delta_1$ with a non-linear one $\sum_{j=1}^{m} \xi_j \,(z_i - \psi_j)_+$. The hinge function $(z_i - \psi_j)_+$ is nonlinear and non-differentiable in ψ_j. Muggeo (2003) proposes approximating each hinge function in the linear predictor by a first-order Taylor expansion around current estimates $\psi_j^{(k)}$.

$$(z_i - \psi_j)_+ \approx (z_i - \psi_j^{(k)})_+ - \mathbf{1}(z_i > \psi_j^{(k)})\,(\psi_j - \psi_j^{(k)}),$$

where $\mathbf{1}(z_i > \psi_j^{(k)})$ is the first derivative of $(z_i - \psi_j^{(k)})_+$ with respect to ψ_j, evaluate at $\psi_j^{(k)}$.

Substituting into the model and defining the incremental correction $\gamma_j^{(k)} = \xi_j(\psi_j - \psi_j^{(k)})$, the segmented component becomes

$$\sum_{j=1}^{m} \xi_j \, (z_i - \psi_j)_+ \approx \sum_{j=1}^{m} \xi_j \, U_{ij}^{(k)} + \sum_{j=1}^{m} \gamma_j^{(k)} \, V_{ij}^{(k)},$$

where the $2m$ working covariates at iteration k are

$$U_{ij}^{(k)} = (z_i - \psi_j^{(k)})_+, \qquad V_{ij}^{(k)} = -\mathbf{1}(z_i > \psi_j^{(k)}), \qquad j = 1, \ldots, m$$

The resulting linearized predictor is

$$\eta_i \approx x_i'\beta + z_i'\delta_1 + \sum_{j=1}^{m} \xi_j \, U_{ij}^{(k)} + \sum_{j=1}^{m} \gamma_j^{(k)} \, V_{ij}^{(k)}$$

and it is linear in all parameters $(\beta, \delta_1, \xi_1, \ldots, \xi_m, \gamma_1, \ldots, \gamma_m)$ for fixed $\psi_j^{(k)}$ and can therefore be fitted as a standard GLM by IRLS. The procedure is iterative and at each k-th iteration until a stopping criterion, the breakpoint estimate is updated.

3.2.1.2 Breakpoint Update and Convergence

Given the ML estimates $(\hat{\xi}_j^{(k)}, \hat{\gamma}_j^{(k)})$ from the linearized model at iteration k, the definition $\gamma_j^{(k)} = \xi_j(\psi_j - \psi_j^{(k)})$ yields the breakpoint update

$$\psi_j^{(k+1)} = \psi_j^{(k)} + \frac{\hat{\gamma}_j^{(k)}}{\hat{\xi}_j^{(k)}}, \qquad j = 1, \ldots, m$$

At convergence $\hat{\gamma}_j^{(k)} \approx 0$ and the update becomes negligible. The stopping rule is

$$\max_{j=1,\ldots,m} |\psi_j^{(k+1)} - \psi_j^{(k)}| < \tau \qquad \text{or equivalently} \qquad \max_{j=1,\ldots,m} |\hat{\gamma}_j^{(k)}| < \tau,$$

for a prescribed small tolerance $\tau > 0$. The final regime-specific coefficient estimates are recovered as $\hat{\delta}_1 = \hat{\delta}_1^{(k)}$ and $\hat{\delta}_{j+1} = \hat{\delta}_1^{(k)} + \sum_{l=1}^{j} \hat{\xi}_l^{(k)}$.

3.2.1.3 Properties of the Estimator

Muggeo (2003) notes that the convergence of the iterative linearization scheme is not guaranteed in general. The Taylor expansion underlying the algorithm is a local first-order approximation, and the sequence $\{\psi_j^{(k)}\}$ may fail to converge if the initial values $\psi_j^{(0)}$ are far from the true breakpoints or if the hinge terms are poorly separated. However, Muggeo (2003) argues that the algorithm is exact when a true breakpoint exists and

the error variance $a(\phi)V(\mu_i)$ is sufficiently small. The linearization becomes exact at convergence in the sense that $\hat{\gamma}_j^{(k)} \to 0$ and $\{\psi_j^{(k)}\}$ stabilizes at the ML estimate. Through simulations, Muggeo (2003) showed that for Gaussian data with moderate variance and a single breakpoint the algorithm converged to the correct solution.

For the regression coefficients (β, δ), classical asymptotic theory holds exactly at each iteration, since each step reduces to fitting a standard GLM. Simulation confirms that these estimators are unbiased with empirical coverage at the nominal level throughout. The consistency of the breakpoint estimator $\hat{\psi}$, by contrast, remains conjectured from simulation evidence rather than formally established. This stands in contrast to the GLS estimator of Section 3.1, for which Bai and Perron (1998) establish super-consistency. The linearization approach sacrifices the formal guarantees in exchange for the substantial computational gain of avoiding the exhaustive partition search, and it remains the standard practical approach for GLM changepoint estimation.

4 Testing for the Existence of a Changepoint

Before applying the changepoint regression models of Chapter 2, it is necessary to test whether structural change is present at all. The null hypothesis of no structural change, meaning that the regression coefficients are stable throughout the sample, must be assessed against the alternative that at least one breakpoint exists at an unknown location. This chapter develops a unified treatment of such tests across both the linear and GLM settings introduced in Chapter 2, drawing on estimation procedures of Chapter 3.

Five distinct methodology families of tests in both the linear and GLM settings are covered: the F-based tests, the Classical Likelihood Ratio Test (LRT), the Empirical Likelihood Ratio Test (LRT), the Cumulative Sum (CUSUM) fluctuation tests, and the Score-based tests. Their corresponding pseudo-algorithms are exposed in Appendix 7.

4.1 Shared Framework

All explored tests share a common structure, regardless of the specific test statistic employed or the regression setting in which they operate. Throughout this chapter, the tests are derived for a single breakpoint model ($m = 1$), without loss of generality, since Bai and Perron (1998) propose an extension to multiple breakpoints via a sequential procedure discussed in Section 4.8.

4.1.1 Testing with a Known Breakpoint

Suppose the true breakpoint ψ_1 is known. The null hypothesis of no structural change implies that the regression coefficients are stable throughout the sample. Therefore, it corresponds to the equality of the regime-specific coefficients δ_j, with $j = 1, 2$

$$H_0 : \ \delta_1 = \delta_2 = \delta,$$

Under H_0 there is no structural change, and the null model M_0 reduces to single-regime standard regression model on the full sample. In the linear and GLM settings, M_0 becomes, respectively

$$M_0 : y_i = x_i'\beta + z_i'\delta + u_i, \qquad M_0 : \eta_i = x_i'\beta + z_i'\delta = g(\mu_i), \qquad i = 1, \ldots, n$$

The alternative hypothesis is that a single shift occurs at the breakpoint ψ_1, meaning that

$$H_1 : \ \delta_1 \neq \delta_2$$

Under H_1, the alternative model M_1 is the single-break partial change model with two regimes $R_1(\psi_1) = \{i : i \leq \psi_1\}$ and $R_2(\psi_1) = \{i : i > \psi_1\}$.

4.1.1.1 Nested Models

The alternative segmented model with one break can be reparametrized as

$$y_i = x_i'\beta + z_i'\delta_1 + z_i'\xi\,\mathbf{1}(i \in R_2(\psi_1)) + u_i, \qquad \eta_i = x_i'\beta + z_i'\delta_1 + z_i'\xi\,\mathbf{1}(i \in R_2(\psi_1)),$$

where δ_1 denotes a baseline coefficient vector, taken as the coefficient vector associated with the first regime, and ξ captures the incremental change in the coefficient vector occurring after the break ψ_1.

Under this parameterization, the null hypothesis of no structural change can be rewritten as a single linear restriction, $H_0 : \xi = 0$. Therefore, when ψ_1 is known, the null model is nested within the alternative and nested-model comparison tools, such as the F-test (in the linear case) or the Likelihood Ratio Test (LRT), apply directly and are standard.

4.1.2 Common Strategy for Unknown Breakpoints

When ψ_1 is unknown, most tests apply the same common strategy. They evaluate a per-candidate test statistic $T(\psi_1)$ for each admissible partition $\psi_1 \in \Psi_{\varepsilon,1}$ and define the overall test statistic, T, as the supremum of all the computed per-candidate statistics $T(\psi_1)$ over the admissible set. The corresponding breakpoint estimator is the candidate breakpoint attaining the supremum statistic.

$$T = \sup_{\psi_1 \in \Psi_{\varepsilon,1}} T(\psi_1), \qquad \hat{\psi}_1 = \arg\max_{\psi_1 \in \Psi_{\varepsilon,1}} T(\psi_1)$$

The admissible breakpoint set $\Psi_{\varepsilon,1}$ imposes the same trimming condition introduced in Section 3.1.2: $\Psi_{\varepsilon,1} = \{\psi_1 : h \leq \psi_1 \leq n - h\}$, with $h = \lfloor n\varepsilon \rfloor$, ensuring both regimes contain enough observations for estimation.

This supremum strategy is common to the sup-F, classical LRT, empirical LRT and the supremum score test. However, some tests in this chapter depart from this approach. The CUSUM-based tests instead of scanning over candidate partitions, monitor the cumulative

behavior of a residual or score process across the full sample. The average score test eliminates the unknown breakpoint before testing by averaging the hinge function over a grid of candidate values.

Although most tests produce a breakpoint estimate as a by-product of their construction, this estimate serves a different role from the dedicated procedures of Chapter 3. The estimation procedures of Chapter 3 require the number of breaks m to be known. However, in settings where m is unknown, the sequential testing procedure explored in Section 4.8 estimates $\hat{m}$ by recursively locating one break at a time through the estimates derived from the testing procedures. Once m^0 has been selected, the GLS estimator of Section 3.1 refines these rough locations into super-consistent estimates satisfying $\hat{\psi}_j - \psi_j^0 = O_p(1)$, a rate that the test-based estimator does not achieve in general. Additionally, in the GLM setting, the iterative linearization of Section 3.2 requires a rough initial breakpoint partition $\psi_j^{(0)}$ to start the Taylor expansion. The test-based estimates are, therefore, a natural candidate for this input. The test-based $\arg\max$ estimates arises as an internal step of the testing procedure, but not as an end, since estimation procedures must be applied. The exceptions are the sup-F in which the breakpoint estimator coincides with the GLS estimation process by definition, and the average score test which coincides with the iterative linearization algorithm, both when m is known.

4.1.3 Non-standard Null Distributions

A central difficulty introduced by this supremum construction concerns the asymptotic distribution of the supremum statistic T. For any fixed candidate ψ_1, standard asymptotic theory applies provided that all parameters are identified under H_0. However, in the structural change setting, the null model fits a single regression on the full sample and the breakpoint ψ_1 is unidentified under H_0. For every candidate $\psi_1 \in \Psi_{\varepsilon,1}$, the restricted null model M_0 is fixed and does not depend on ψ_1. Therefore, ψ_1 is a nuisance parameter present only under H_1, that is unidentified under H_0. This invalidates the standard χ^2 or F approximation and yields a nonstandard limiting distribution (Andrews, 1993).

A further complication is the dependence structure of the per-candidate statistics. For different candidate breakpoint locations, the per-candidate statistics $T(\psi_1)$ compare the same null model M_0 to alternative models that only differ in the partition and are, therefore, typically very similar and highly correlated. Consequently, T is not the maximum

of approximately independent variables, but rather the maximum of strongly correlated components, making the derivation of the limiting distribution complex.

Calibration strategies vary by family: the F-based and CUSUM-based tests use tabulated Brownian bridge critical values implemented into the `strucchange` package in `R`, the classical LRT and EL tests use bootstrap procedures, the supremum score test uses the conservative Davies analytical bound, and the average score test recovers a standard $\chi^2(q)$ distribution.

4.2 F-Statistic Based Tests

The F-statistic based tests are restricted to the linear homoskedastic setting. Andrews (1993) extended the classical Chow (1960) for a known breakpoint, to the case of an unknown break location by deriving the asymptotic distribution of the supremum F-statistic over all admissible candidates, applying the common strategy of Section 4.1.2.

Bai and Perron (1998, 2003a) extend this framework to the general case of multiple structural breaks at unknown locations, developing both efficient estimation of break dates via dynamic programming and a sequential testing procedure that consistently determines the number of breaks. Their framework remains arguably the most complete unified treatment of structural change in linear regression.

Andrews and Ploberger (1994) showed that, since no uniformly most powerful test exists in this setting, different aggregations of the same per-candidate F-sequence lead to tests with different power properties, and proposed the Exp-F and Mean-F alternatives. All three tests are built from the same underlying sequence of per-candidate F-statistics and share the same Brownian bridge limiting distribution, differing only in how they summarize that sequence.

4.2.1 Standard F-Test with a Known Breakpoint

Suppose that the breakpoint ψ_1 is known. As established in Section 4.1.1.1, M_0 is nested within the alternative model $M_1(\psi_1)$ and, therefore, a standard F- test applies to compare them.

Let RSS_0 denote the sum of squared residuals of the restricted model under H_0, estimated on the full sample, and let $\mathrm{RSS}_1(\psi_1)$ denote the sum of squared residuals of the unrestricted model with a break at ψ_1, estimated on the two regimes separately.

Since the dimension of x_i is p and the dimension of z_i is q, the restricted and unrestricted models, M_0 and M_1, contain, respectively, $k_0 = p + q$ and $k_1 = p + 2q$ parameters. Therefore, the null model imposes $r = k_1 - k_0 = q$ restrictions on the alternative. Thus, the F-statistic for a known break at ψ_1 is

$$F(\psi_1) = \frac{\left(\text{RSS}_0 - \text{RSS}_1(\psi_1)\right)/r}{\text{RSS}_1(\psi_1)/(n - k_1)},$$

which follows an $F(r,\, n - k_1)$ distribution under H_0 and a known ψ_1.

4.2.2 Supremum F Test

When ψ_1 is unknown, the common strategy of Section 4.1.2 is applied. For each admissible candidate breakpoint $\psi_1 \in \Psi_{\varepsilon,1}$ the F-statistic from Section 4.2.1 is computed to measure evidence against the null hypothesis of no structural change with that particular configuration of breakpoints and the overall test statistic is the supremum over the admissible set. This statistic captures the strongest evidence against the null hypothesis H_0 across all admissible break locations.

$$\sup F = \sup_{\psi_1 \in \Psi_{\varepsilon,1}} F(\psi_1) \qquad \hat{\psi}_1 = \arg\max_{\psi_1 \in \Psi_{\varepsilon,1}} F(\psi_1)$$

4.2.2.1 Non-standard Null Distribution

As established in Section 4.1.3, the asymptotic distribution of the $\sup F$ is nonstandard and depends on the trimming parameter ε and on the number of breaks m (Andrews, 1993, Bai and Perron, 1998, 2003b). Andrews (1993) proved that the F-statistic at a fixed candidate break fraction λ_1 is asymptotically a quadratic functional of a Brownian bridge, satisfying

$$F(\lfloor n\lambda_1 \rfloor) \xrightarrow{d} \frac{\|B^0(\lambda_1)\|^2}{\lambda_1(1 - \lambda_1)},$$

where $B^0(\lambda_1)$ is a standard Brownian bridge.

A standard Brownian motion, or Wiener process, $W(\lambda_1)$ is a continuous Gaussian process with $W(0) = 0$ and independent increments satisfying $W(t) - W(s) \sim N(0, t - s)$ for $0 \leq s \leq t \leq 1$. A Brownian bridge is a Brownian motion on $[0, 1]$, with $B^0(0) = B^0(1) = 0$ and covariance $\text{Cov}(B^0(s), B^0(t)) = s(1 - t)$, for $s \leq t$. It can be represented in terms of a standard Brownian motion as $B^0(\lambda_1) = W(\lambda_1) - \lambda_1 W(1)$, which makes the conditioning on $W(1) = 0$ explicit.

Taking the supremum over admissible fractions, the limiting distribution of the sup-F statistic gives

$$\sup F \xrightarrow{d} \sup_{\lambda_1 \in [\varepsilon, 1-\varepsilon]} \frac{\|B^0(\lambda_1)\|^2}{\lambda_1(1-\lambda_1)},$$

a nonstandard distribution that depends only on q (the number of regime-varying coefficients) and the trimming parameter ε, but not on β, δ, or σ^2.

Critical values can be obtained from simulations of the limiting Brownian bridge distribution or via bootstrap procedures that approximate the null distribution empirically. The latter is particularly useful when the sample size is small or the distributional assumptions are in doubt, since the asymptotic Brownian bridge approximation is known to perform poorly in finite samples. This universality makes it possible to tabulate critical values once for use across models. Bai and Perron (1998) and Bai and Perron (2003b) provide asymptotic critical values for the sup-F with $q = 1, \ldots, 10$ and $\varepsilon \in \{0.05, 0.10, 0.15, 0.25\}$. This test is implemented directly in the `strucchange` package in R, with the function `Fstats()` computing the sequence $\{F(\psi_1)\}_{\psi_1 \in \Psi_{\varepsilon,1}}$, and `sctest()` performing the sup-F test.

4.2.3 Exp-F and Mean-F Tests

Andrews and Ploberger (1994) show that no uniformly most powerful test exists when ψ_1 is unidentified, motivating two alternatives to the sup-F test that aggregate the same per-candidate sequence $\{F(\psi_1)\}_{\psi_1 \in \Psi_{\varepsilon,1}}$ differently:

$$\text{Exp-}F = \log\left(\frac{1}{|\Psi_{\varepsilon,1}|}\sum_{\psi_1 \in \Psi_{\varepsilon,1}} \exp\left(\frac{1}{2}F(\psi_1)\right)\right), \qquad \text{Mean-}F = \frac{1}{|\Psi_{\varepsilon,1}|}\sum_{\psi_1 \in \Psi_{\varepsilon,1}} F(\psi_1)$$

All three tests share the same Brownian bridge limit distribution, differing only in the functional applied to it, and are implemented in `strucchange` via `sctest(fs, type = "expF")` and `sctest(fs, type = "meanF")`.

The choice among them reflects assumptions about the expected break pattern. The sup-F test is most powerful against a single sharp break, but achieves this by concentrating all its power on the single strongest candidate, discarding the information contained in the rest of the sequence $\{F(\psi_1)\}_{\psi_1 \in \Psi_{\varepsilon,1}}$. The Exp-$F$ test retains near-optimal power against a single break while also weighting contributions from candidates near the true location more heavily than distant ones, through the exponential transformation of $F(\psi_1)$. The Mean-F test is the most conservative of the three, it averages over the entire admissible set which means that candidates far from the true breakpoint, where $F(\psi_1)$ is close to its

null-distribution mean, dilute the statistic, so the test responds poorly to a single sharp, localized break. Its advantage emerges when the break signal is diffuse, spread gradually across many observations rather than concentrated at one location, in which case more of the admissible set carries genuine signal and the average accumulates it effectively.

4.3 Classical Likelihood Ratio Test

The classical parametric Likelihood Ratio Test (LRT) for structural change was developed by Quandt (1958, 1960). It follows the common strategy of Section 4.1.2, for each admissible candidate break location, a likelihood ratio statistic is computed comparing the null model to the alternative model with a break at that location, an the overall test statistic is the supremum over all candidates. Unlike the sup-F test, it does not require homoskedasticity, estimating separate error variances σ_1^2 and σ_2^2 in each regime under H_1. The test is developed first for the linear regression setting and then extended to the GLM setting.

4.3.1 Classical LRT in Linear Regression

Assume the partial change model of Section 2.1 with a single breakpoint:

$$y_i = \begin{cases} x_i'\beta + z_i'\delta_1 + u_i, & i \in R_1(\psi_1), \\ x_i'\beta + z_i'\delta_2 + u_i, & i \in R_2(\psi_1), \end{cases}$$

where $u_i \sim \mathcal{N}(0, \sigma_j^2)$ independently within each regime $R_j(\psi_1)$.

4.3.1.1 LRT with Known Breakpoint

When ψ_1 is known, a standard LRT applies. For each candidate breakpoint $\psi_1 \in \Psi_{\varepsilon,1}$, the test compares the maximized likelihood of the null model without a break $L(\hat{M}_0)$, with the maximized likelihood of the alternative model with a break at that candidate location, $L(\hat{M}_1(\psi_1))$

$$\mathrm{LR}(\psi_1) = \frac{L(\hat{M}_0)}{L(\hat{M}_1(\psi_1))}$$

Null model M_0: Under H_0, the likelihood of the null model M_0 is that of a single regression on the full sample.

$$L_0(\beta, \delta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - x_i'\beta - z_i'\delta)^2 \right)$$

The log likelihood is

$$\ell_0(\beta, \delta, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{\text{RSS}_0(\beta, \delta)}{2\sigma^2}, \qquad \text{RSS}_0(\beta, \delta) = \sum_{i=1}^{n}(y_i - x_i'\beta - z_i'\delta)^2$$

Maximizing over (β, δ) yields the OLS estimators $(\hat{\beta}, \hat{\delta})$ and the concentrated residual sum of squares $\text{RSS}_0 = \text{RSS}_0(\hat{\beta}, \hat{\delta})$. Differentiating ℓ_0 with respect to σ^2 gives

$$\frac{\partial \ell_0}{\partial \sigma^2} = 0 \Rightarrow \hat{\sigma}^2 = \frac{\text{RSS}_0}{n}$$

Hence, the maximized likelihood under H_0 is

$$L(\hat{M}_0) = (2\pi)^{-n/2}(\hat{\sigma}^2)^{-n/2}\exp(-n/2)$$

Alterative model with fixed candidate, M_1 Under H_1, with a known candidate break, the likelihood of $M_1(\psi_1)$ factorizes across the two regimes $R_1(\psi_1)$ and $R_2(\psi_1)$, with n_1 and n_2 observations, respectively.

$$L_1(\beta, \delta_1, \delta_2, \sigma_1^2, \sigma_2^2; \psi_1) = (2\pi\sigma_1^2)^{-n_1/2}\exp\left(-\frac{RSS_1(\beta, \delta_1; \psi_1)}{2\sigma_1^2}\right)(2\pi\sigma_2^2)^{-n_2/2}\exp\left(-\frac{RSS_2(\beta, \delta_2; \psi_1)}{2\sigma_2^2}\right),$$

where

$$\text{RSS}_{(j)}(\beta, \delta_j; \psi_1) = \sum_{i \in R_j(\psi_1)}(y_i - x_i'\beta - z_i'\delta_j)^2, \qquad j = 1, 2$$

Maximization over $(\beta, \delta_1, \delta_2)$ yields OLS estimators $(\hat{\beta}(\psi_1), \hat{\delta}_1(\psi_1), \hat{\delta}_2(\psi_1))$, and the concentrated segment residual sums of squares:

$$\text{RSS}_j(\psi_1) = \text{RSS}_j(\hat{\beta}(\psi_1), \hat{\delta}_j(\psi_1); \psi_1), \quad j = 1, 2$$

The log-likelihood function under H_1 becomes:

$$\ell_1(\beta, \sigma^2) = -\frac{n_1}{2}\log(2\pi\sigma_1^2) - \frac{RSS_1(\psi_1)}{2\sigma_1^2} - \frac{n_2}{2}\log(2\pi\sigma_2^2) - \frac{RSS_2(\psi_1)}{2\sigma_2^2}$$

Differentiating ℓ_1 and setting it to zero, with respect to σ_1^2 and σ_2^2 yields, respectively

$$\hat{\sigma}_1^2(\psi_1) = \frac{\text{RSS}_1(\psi_1)}{n_1} \qquad \hat{\sigma}_2^2(\psi_1) = \frac{\text{RSS}_2(\psi_1)}{n_2}$$

Hence, the maximized likelihood under H_1 is

$$L(\hat{M}_1(\psi_1)) = (2\pi)^{-n/2}(\hat{\sigma}_1^2(\psi_1))^{-n_1/2}(\hat{\sigma}_2^2(\psi_1))^{-n_2/2}\exp\left(-\frac{n}{2}\right),$$

Likelihood ratio simplification For a given candidate ψ_1, the likelihood ratio takes the form

$$\text{LR}(\psi_1) = \frac{\hat{\sigma}_1(\psi_1)^{n_1}\hat{\sigma}_2(\psi_1)^{n_2}}{\hat{\sigma}^n}$$

Applying the monotone transformation $\lambda(\psi_1) = -2\log \text{LR}(\psi_1)$, the per-candidate statistic is

$$\lambda(\psi_1) = -2\log\left(\frac{\hat{\sigma}_1(\psi_1)^{n_1}\hat{\sigma}_2(\psi_1)^{n_2}}{\hat{\sigma}^n}\right) = n\log(\hat{\sigma}^2) - n_1\log(\hat{\sigma}_1^2(\psi_1)) - n_2\log(\hat{\sigma}_2^2(\psi_1)),$$

where $\hat{\sigma}_1(\psi_1)$ and $\hat{\sigma}_2(\psi_1)$ are the estimated error standard deviations from the two regime regressions, and $\hat{\sigma}$ is the estimator from the single full sample regression under H_0. If allowing a break at ψ_1 substantially reduces within segment residual dispersion, then the unrestricted likelihood is much bigger, so $\text{LR}(\psi_1)$ is small and $\lambda(\psi_1)$ is large, providing evidence of a break at ψ_1.

4.3.1.2 Supremum Statistic and Breakpoint Estimator

When ψ_1 is unknown, the common strategy of Section 4.1.2 is applied. For each admissible candidate $\psi_1 \in \Psi_{\varepsilon,1}$, $\lambda(\psi_1)$ is computed, and the overall supremum test statistic and associated breakpoint estimator are

$$\Lambda = \max_{\psi_1 \in \Psi_{\varepsilon,1}} \lambda(\psi_1), \qquad \hat{\psi}_1 = \arg\max_{\psi_1 \in \Psi_{\varepsilon,1}} \lambda(\psi_1)$$

4.3.1.3 Relationship Between the Sup-F and Classical LRT

In a standard Gaussian linear regression without breakpoints, the LRT statistic is a monotone transformation of the F-statistic for testing linear restrictions. Consequently, the two tests yield identical rejection decisions. In the linear changepoint setting, when ψ_1 is known and under equal error variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the F-statistic and the LRT statistic are equivalent. Therefore, under Gaussian homoskedastic errors, since $F(\psi_1)$ and $\lambda(\psi_1)$ are equivalent at every fixed ψ_1, the two suprema are attained at the same candidate $\hat{\psi}_1$ and select the same break location.

However, when variances differ across regimes, this equivalence breaks down, the LRT accommodates heteroskedasticity by estimating $\hat{\sigma}_1^2(\psi_1)$ and $\hat{\sigma}_2^2(\psi_1)$ separately, whereas the F-statistic assumes homoskedasticity across regimes.

4.3.1.4 Null Distribution and Bootstrap Calibration

The equality of the two tests under Gaussian homoskedastic errors implies that Λ shares the same Brownian bridge limiting distribution and, therefore, the tabulated critical values of the sup-F test apply. However, under heteroskedasticity the equivalence and underlying tables need not hold. Instead, a wild bootstrap procedure is used, since it adapts automatically to the actual error structure of the data and remains valid whether or not variances are equal across regimes.

The wild bootstrap procedure was introduced by Wu (1986) and developed by Liu (1988) and Mammen (1993) to accommodate heteroskedasticity errors of unknown form. The data-level (fixed-design) wild bootstrap perturbs each null-model residual by an independent mean-zero, unit-variance multiplier, holding the design (x_i, z_i) fixed. Following Davidson and Flachaire (2008), Rademacher multipliers are used. Bootstrap calibration of goodness-of-fit statistics in regression is reviewed in Stute et al. (1998) and González-Manteiga and Crujeiras (2013).

4.3.2 Classical LRT in GLMs

The classical parametric LRT extends naturally to the GLM partial change model. The structure is identical to the linear case, for each candidate ψ_1, the null model M_0 is fitted once on the full sample and the alternative model $M_1(\psi_1)$ is fitted separately, the log-likelihood ratio is computed, and the supremum is taken over $\Psi_{\varepsilon,1}$. The key difference is that the maximized log-likelihoods no longer have closed-form expressions in terms of residual sums of squares and must be evaluated numerically via IRLS.

This procedure assumes a specification of the link function $g(\cdot)$, the variance function $V(\mu_i)$, and the dispersion function $a(\phi)$. Misspecification of any of these components affects both the maximized log-likelihoods and the bootstrap procedure.

4.3.2.1 LRT with Known Breakpoint

Null model M_0 Under H_0, the null model reduces to a standard GLM on the full sample.

$$\eta_i = x_i'\beta + z_i'\delta, \qquad i = 1, \ldots, n$$

The parameters $(\hat{\beta}, \hat{\delta}, \hat{\phi})$ are estimated by the IRLS procedure on the full sample. The maximized log-likelihood is

$$\ell_0 = \sum_{i=1}^{n} \left[\frac{y_i \hat{\theta}_i - b(\hat{\theta}_i)}{a(\hat{\phi})} + c(y_i, \hat{\phi}) \right], \qquad \hat{\theta}_i = b'^{-1}\Big(g^{-1}(x_i'\hat{\beta} + z_i'\hat{\delta})\Big)$$

Alterative model with fixed candidate, M_1 Under H_1, for a fixed candidate ψ_1, the model allows regime-specific coefficients:

$$\eta_i = x_i'\beta + z_i'\delta_j, \qquad i \in R_j(\psi_1), \quad j = 1, 2$$

The estimators $(\hat{\beta}(\psi_1), \hat{\delta}_1(\psi_1), \hat{\delta}_2(\psi_1))$ are obtained by running IRLS on the expanded design matrix $X^* = (X \mid Z)$ with Z block-diagonal at ψ_1. The maximized log-likelihood decomposes into regime-specific contributions

$$\ell_1(\psi_1) = \ell_1(\hat{\beta}(\psi_1), \hat{\delta}_1(\psi_1), \hat{\phi}_1(\psi_1) \mid \psi_1) + \ell_2(\hat{\beta}(\psi_1), \hat{\delta}_2(\psi_1), \hat{\phi}_1(\psi_1) \mid \psi_1)$$

Therefore,

$$\ell_1(\psi_1) = \sum_{i \in R_1(\psi_1)} \left[\frac{y_i \hat{\theta}_{i1} - b(\hat{\theta}_{i1})}{a(\hat{\phi})} + c(y_i, \hat{\phi}) \right] + \sum_{i \in R_2(\psi_1)} \left[\frac{y_i \hat{\theta}_{i2} - b(\hat{\theta}_{i2})}{a(\hat{\phi})} + c(y_i, \hat{\phi}) \right],$$

where $\hat{\theta}_{i1} = b'^{-1}\Big(g^{-1}(x_i'\hat{\beta}(\psi_1) + z_i'\hat{\delta}_1(\psi_1))\Big)$ and $\hat{\theta}_{i2} = b'^{-1}\Big(g^{-1}(x_i'\hat{\beta}(\psi_1) + z_i'\hat{\delta}_2(\psi_1))\Big)$.

Likelihood ratio statistic For a fixed candidate ψ_1, the log-likelihood ratio statistic is

$$\lambda(\psi_1) = 2[\ell_1(\psi_1) - \ell_0]$$

A large value of $\lambda(\psi_1)$ indicates that the unrestricted model with a break at ψ_1 fits substantially better than the null model, providing evidence of structural change.

4.3.2.2 Supremum Statistic and Breakpoint Estimator

When ψ_1 is unknown, the common strategy of Section 4.1.2 is applied. The overall supremum test statistic and breakpoint estimator are

$$\Lambda = \max_{\psi_1 \in \Psi_{\varepsilon,1}} \lambda(\psi_1), \qquad \hat{\psi}_1 = \arg\max_{\psi_1 \in \Psi_{\varepsilon,1}} \lambda(\psi_1)$$

4.3.2.3 Null Distribution and Bootstrap Calibration

In the GLM case, the same nonstandard distribution issues of Section 4.1.3 apply, and critical values are obtained via the parametric bootstrap rather than the data-level wild bootstrap employed in the linear case.

The data-level wild bootstrap procedure constructs pseudo-responses by ading perturbed residuals to the null fitted values, of the form $\hat{\mu}_i^{(0)} + v_i^{(b)}\hat{u}_i$. In the linear case, this is valid since the response is continuous and the perturbed values remain on the real line.

However, for discrete responses (e.g., Bernoulli, Poisson), pseudo-responses of this form would not respect the support of the GLM family, falling outside of $\{0,1\}$ or $\mathbb{Z}_0$. A parametric bootstrap, originally proposed by Efron (1979), is therefore used instead, in which pseudo-responses are drawn directly from the fitted null distribution. Although a data-level wild bootstrap remains valid for continuous GLM families, the parametric bootstrap is adopted uniformly to give a single procedure that handles both discrete and continuous responses. Bootstrap calibration is reviewed in González-Manteiga and Crujeiras (2013).

4.4 Empirical Likelihood Ratio Test

The Empirical Likelihood (EL) ratio test of Liu and Qian (2009), building on the nonparametric likelihood framework of Owen (1991), provides an alternative to the classical parametric LRT that imposes no distributional assumption of the error terms. Rather than maximizing a parametric likelihood, the EL approach assigns probability masses to observed residuals and maximizes a nonparamteric likelihood subject to a mean-zero condition imposed under H_0. A key structural difference from the classical parametric LRT is that the EL procedure is built on the continuity condition at the breakpoint. Therefore, under H_1, attention is restricted to continuous break models satisfying $z'_{\psi_1}(\delta_1 - \delta_2) = 0$ with $\delta_1 \neq \delta_2$.

The test is developed first in the linear regression setting, then extended to the GLM setting, a proposed generalization of the framework, not explored by Liu and Qian (2009).

4.4.1 EL Ratio Test in Linear Regressions

Assume the previously defined model of Section 2.1, with a single changepoint ($m = 1$),

$$y_i = \begin{cases} x_i'\beta + z_i'\delta_1 + u_i, & i \in R_1(\psi_1) \\ x_i'\beta + z_i'\delta_2 + u_i, & i \in R_2(\psi_2) \end{cases}$$

No particular distribution is assumed for the error terms u_i, with the only additionally requirement being the continuity condition at the breakpoint. For each candidate break $\psi_1 \in \Psi_{\varepsilon,1}$, let $(\hat{\beta}(\psi_1), \hat{\delta}_1(\psi_1))$ and $(\hat{\beta}(\psi_1), \hat{\delta}_2(\psi_1))$ denote the OLS estimators obtained from fitting separate regressions on each regime. If the continuity condition fails, it is enforced by refitting via the hinge parametrization

$$y_i = x_i'\beta + z_i'\gamma + \xi\,(z_i - z_{\psi_1})_+ + u_i,$$

setting $\hat{\delta}_1(\psi_1) = \hat{\gamma}(\psi_1)$ and $\hat{\delta}_2(\psi_1) = \hat{\gamma}(\psi_1) + \hat{\xi}(\psi_1)$.

4.4.1.1 Null-Consistent Errors

Then the estimated errors are

$$\hat{u}_i(\psi_1) = \begin{cases} y_i - x_i'\hat{\beta}(\psi_1) - z_i'\hat{\delta}_1(\psi_1), & i \in R_1(\psi_1) \\ y_i - x_i'\hat{\beta}(\psi_1) - z_i'\hat{\delta}_2(\psi_1), & i \in R_2(\psi_1) \end{cases}$$

Under the null hypothesis, $H_0 : \delta_1 = \delta_2 = \delta$, the null-consistent estimated errors can be constructed by swapping the regime dependent coefficient estimators, $\hat{\delta}_1$ with $\hat{\delta}_2$, across regimes:

$$\tilde{u}_i(\psi_1) = \begin{cases} y_i - x_i'\hat{\beta}(\psi_1) - z_i'\hat{\delta}_2(\psi_1), & i \in R_1(\psi_1), \\ y_i - x_i'\hat{\beta}(\psi_1) - z_i'\hat{\delta}_1(\psi_1), & i \in R_2(\psi_1) \end{cases}$$

These null-consistent errors are constructed to satisfy the mean zero moment condition in the population under H_0:

$$E[\tilde{u}_i(\psi_1)] = 0, \quad \text{for all } \psi_1 \in \Psi_{\varepsilon,1}$$

4.4.1.2 Weight spaces Under the Null and Alternative Hypotheses

Following Owen (1991), the EL approach avoids specifying a parametric density for the error terms and instead assigns weights as probability masses to observed sample points. Here, the sample points are the constructed null consistent residuals $\tilde{u}_i(\psi_1)$.

For a fixed ψ_1, let $w_i(\psi_1)$ denote the mass probability at $\tilde{u}_i(\psi_1)$ subject to the following constraints, which enter the weight spaces under both hypotheses

$$w_i(\psi_1) \geq 0, \quad \sum_{i=1}^{n} w_i(\psi_1) = 1, \quad i = 1, \ldots, n$$

Under H_1, the unrestricted simplex of admissible weights is

$$\mathcal{W} = \left\{ w \in \mathbb{R}^n : w_i(\psi_1) \geq 0,\ \sum_{i=1}^{n} w_i(\psi_1) = 1 \right\},$$

Under H_0, an additional restriction, $E[\tilde{u}_i(\psi_1)] = 0$, is imposed. Since the probability masses $w_i(\psi_1)$ are placed on the null-consistent observed residuals $\tilde{u}_i(\psi_1)$, the condition can be written as

$$E[\tilde{u}_i(\psi_1)] = \sum_{i=1}^{n} w_i(\psi_1)\, \tilde{u}_i(\psi_1) = 0$$

Therefore, the restricted simplex of admissible weights under the null hypothesis H_0 is

$$\mathcal{W}_0(\psi_1) = \left\{ w \in \mathbb{R}^n : w_i(\psi_1) \geq 0, \quad \sum_{i=1}^{n} w_i(\psi_1) = 1, \quad \sum_{i=1}^{n} w_i(\psi_1)\, \tilde{u}_i(\psi_1) = 0 \right\}$$

4.4.1.3 Empirical Likelihood Ratio

Likelihood under alternative model M_1 Since the distribution assigns probability masses $w_i(\psi_1)$ to the observed null-consistent residuals $\tilde{u}_i(\psi_1)$, the likelihood of observing the sample $\{\tilde{u}_i(\psi_1)\}$, under this discrete distribution is $\prod_{i=1}^{n} w_i(\psi_1)$. Maximizing $\prod_{i=1}^{n} w_i(\psi_1)$ is equivalent to maximizing $\sum_{i=1}^{n} \log(w_i(\psi_1))$, under the alternative weight space $\mathcal{W}$. Since, this problem does not depend on $\tilde{u}_i(\psi_1)$ it can be solved as an unconstrained optimization problem in the admissible simplex. A standard Lagrangian argument (see Appendix 7) gives $w_i(\psi_1) = \frac{1}{n}$ for all i. Therefore, under H_1, the maximized empirical likelihood is

$$\sup_{H_1} \mathsf{EL} = \sup_{w \in \mathcal{W}} \left\{ \prod_{i=1}^{n} w_i(\psi_1) : w_i(\psi_1) \geq 0, \sum_{i=1}^{n} w_i(\psi_1) = 1 \right\} = \left(\frac{1}{n} \right)^n$$

Likelihood of the restricted model M_0 Under H_0 maximized empirical likelihood is

$$\sup_{H_0} \mathsf{EL} = \sup_{w \in \mathcal{W}_0(\psi_1)} \left\{ \prod_{i=1}^{n} w_i(\psi_1) : w_i(\psi_1) \geq 0, \sum_{i=1}^{n} w_i(\psi_1) = 1, \sum_{i=1}^{n} w_i(\psi_1) \tilde{u}_i(\psi_1) = 0 \right\}$$

Empirical likelihood ratio For each fixed $\psi_1 \in \Psi_{\varepsilon,1}$ the EL ratio is

$$\mathcal{R}(\psi_1) = n^n \sup_{w \in \mathcal{W}_0(\psi_1)} \left\{ \prod_{i=1}^{n} w_i(\psi_1) \mid \sum_{i=1}^{n} w_i(\psi_1)\, \tilde{u}_i(\psi_1) = 0,\ w_i(\psi_1) \geq 0,\ \sum_{i=1}^{n} w_i(\psi_1) = 1 \right\}$$

The optimization problem The goal is to maximize the objective

$$\ell = \log\left(n^n \prod_{i=1}^{n} w_i(\psi_1)\right) = n\log(n) + \sum_{i=1}^{n} \log(w_i(\psi_1)),$$

which is equivalent to maximizing $\sum_{i=1}^{n} \log(w_i(\psi_1))$, subject to the constraints involved in $\mathcal{W}_0$. Solving the constrained optimization problem via Lagrange multipliers (see Appendix 7) yields

$$w_i(\psi_1) = \frac{1}{n(1+\lambda(\psi_1)\,\tilde{u}_i(\psi_1))}, \qquad i = 1,\ldots,n,$$

where $\lambda(\psi_1)$, is the solution to

$$\sum_{i=1}^{n} \frac{\tilde{u}_i(\psi_1)}{1+\lambda(\psi_1)\,\tilde{u}_i(\psi_1)} = 0$$

Therefore, the optimized EL ratio is

$$\mathcal{R}(\psi_1) = \prod_{i=1}^{n} n w_i(\psi_1) = \prod_{i=1}^{n} \frac{1}{1+\lambda(\psi_1)\tilde{u}_i(\psi_1)}$$

4.4.1.4 Supremum Statistic and Breakpoint Estimator

For each candidate $\psi_1 \in \Psi_{\varepsilon,1}$, the quantity $-2\log\hat{\mathcal{R}}(\psi_1)$ measures the degree of incompatibility between the data and the null hypothesis of no structural change. Following the common strategy of Section 4.1.2, the overall test statistic and breakpoint estimator are

$$M_n = \sup_{\psi_1\in\Psi_{\varepsilon,1}} \left\{-2\log\hat{\mathcal{R}}(\psi_1)\right\}, \qquad \hat{\psi}_1 = \arg\max_{\psi_1\in\Psi_{\varepsilon,1}} \left\{-2\log\hat{\mathcal{R}}(\psi_1)\right\}$$

4.4.1.5 Null Distribution and Critical Values

Liu and Qian (2009) show, through empirical evidence, that the transformed statistic $Z_n = \sqrt{M_n}$ converges to a Gumbel extreme-value distribution under H_0. However, the location and scale parameters of this distribution are not universal, they depend on the sample size, the finite sample behavior of Z_n for a given n, the error distribution, and the trimming constraint. Similarly to what was proposed for the classical LRT, a data-level wild bootstrap procedure can be applied, since it makes no distributional assumption on Z_n and adapts automatically to the sample size, the trimming condition, and the observed design points.

4.4.2 EL Ratio Test in GLMs

The EL ratio test extends to the GLM setting by replacing the OLS residuals with residuals derived from the fitted GLM. The core structure of the test, such as constructing null-consistent residuals by swapping regime estimators, assigning probability masses subject to a mean-zero moment condition, and taking the supremum of $-2\log\hat{\mathcal{R}}(\psi_1)$, carries over without modification.

It should be noted that Liu and Qian (2009) do not consider the GLM setting. The procedure described here is a proposed extension of their linear framework. The theoretical properties of this extension, in particular the limiting distribution of M_n and the validity of the Gumbel approximation for $Z_n = \sqrt{M_n}$, require additional development beyond what is established in Liu and Qian (2009). A key advantage of the EL approach relative to the classical LRT is its robustness, since it requires only correct specification of the conditional mean $\mu_i = g^{-1}(\eta_i)$ and is therefore robust to misspecification of the variance function $V(\mu_i)$ and dispersion $a(\phi)$.

4.4.2.1 Null-Consistent Errors

For each candidate $\psi_1 \in \Psi_{\varepsilon,1}$, separate GLMs are fitted by IRLS on each regime, yielding ML estimators $(\hat{\beta}(\psi_1), \hat{\delta}_j(\psi_1))$ and fitted means $\hat{\mu}_i^{(j)}(\psi_1) = g^{-1}(x_i'\hat{\beta}(\psi_1) + z_i'\hat{\delta}_j(\psi_1))$ for $j = 1, 2$. The continuity condition is checked as in the linear case, and, if it fails, it is enforced by refitting via the hinge parametrization on the linear predictor. Three choices of null-consistent errors are available, each obtained by swapping the regime estimators across regimes, as in the linear construction.

Null-consistent raw errors:
$$\tilde{u}_i(\psi_1) = \begin{cases} y_i - \hat{\mu}_i^{(2)}(\psi_1), & i \in R_1(\psi_1), \\ y_i - \hat{\mu}_i^{(1)}(\psi_1), & i \in R_2(\psi_1), \end{cases}$$

Null-consistent working errors:
$$\tilde{u}_i^*(\psi_1) = \begin{cases} \left(y_i - \hat{\mu}_i^{(2)}(\psi_1)\right) g'(\hat{\mu}_i^{(2)}(\psi_1)), & i \in R_1(\psi_1), \\ \left(y_i - \hat{\mu}_i^{(1)}(\psi_1)\right) g'(\hat{\mu}_i^{(1)}(\psi_1)), & i \in R_2(\psi_1), \end{cases}$$

Null-consistent Pearson errors: $$\tilde{u}_i^P(\psi_1) = \begin{cases} \dfrac{y_i - \hat{\mu}_i^{(2)}(\psi_1)}{\sqrt{V(\hat{\mu}_i^{(2)}(\psi_1))}}, & i \in R_1(\psi_1), \\ \dfrac{y_i - \hat{\mu}_i^{(1)}(\psi_1)}{\sqrt{V(\hat{\mu}_i^{(1)}(\psi_1))}}, & i \in R_2(\psi_1), \end{cases}$$

Pearson errors introduce $V(\mu_i)$, which undermines the distributional robustness that motivates the EL approach. The test is therefore not implemented with the Pearson variant. Both null-consistent errors satisfy the asymptotic moment condition under H_0

$$E[\tilde{u}_i(\psi_1)] = 0, \qquad E[\tilde{u}_i^*(\psi_1)] = 0, \qquad \text{for all } \psi_1 \in \Psi_{\varepsilon,1},$$

since $\delta_1 = \delta_2 = \delta$, $\hat{\delta}_1(\psi_1) \xrightarrow{p} \delta^0$ and $\hat{\delta}_2(\psi_1) \xrightarrow{p} \delta^0$ imply both fitted means converge to $\mu_i^0 = g^{-1}(x_i'\beta^0 + z_i'\delta^0)$.

4.4.2.2 EL Construction

With the null-consistent errors, the EL construction proceeds identically the linear construction, with the null-consistent raw or working residuals replacing the linear null-consistent residuals throughout. The following section uses working residuals $\tilde{u}_i^*(\psi_1)$.

Following the computations of Section 4.4.1.3 optimal weights are

$$w_i(\psi_1) = \frac{1}{n\Big(1 + \lambda(\psi_1)\,\tilde{u}_i^*(\psi_1)\Big)}, \qquad \text{satisfying} \qquad \sum_{i=1}^{n} \frac{\tilde{u}_i^*(\psi_1)}{1 + \lambda(\psi_1)\tilde{u}_i^*(\psi_1)} = 0$$

With $\mathcal{R}(\psi_1) = \prod_{i=1}^n 1/(1 + \lambda(\psi_1)\tilde{u}_i^*(\psi_1))$, the overall test statistic and breakpoint estimators are

$$M_n = \sup_{\psi_1 \in \Psi_{\varepsilon,1}} \Big\{-2\log \hat{\mathcal{R}}(\psi_1)\Big\}, \qquad \hat{\psi}_1 = \arg\max_{\psi_1 \in \Psi_{\varepsilon,1}} \Big\{-2\log \hat{\mathcal{R}}(\psi_1)\Big\}$$

4.4.2.3 Multiplier Wild Bootstrap

Calibrating M_n via a wild bootstrap as in the linear case, would generate pseudo-responses that would not respect the support of the GLM family. This was the motivation for applying a parametric bootstrap in the GLM Classical LRT (Section 4.3.2.3). However, a parametric procedure would reintroduce exactly the distributional assumption that the EL test is designed to avoid. Observing that the test statistic M_n depends on the data solely through

the null-consistent residuals $\{\tilde{u}_i^*(\psi_1)\}_{i=1}^n$, re-generating a pseudo-bootstrap response and re-running the entire estimation procedure for every replicate is unnecessary.

A multiplier wild bootstrap is therefore natural, as it randomly perturbs residuals directly by mean-zero, unit-variance multipliers rather than perturbing the response and without need to reference any parametric family. For each replicate $\psi_1 \in \Psi_{\varepsilon,1}$, a single multiplier vector $v^{(b)}$ is drawn and the perturbed residuals $\tilde{u}_i^{*(b)}(\psi_1) = v_i^{(b)}\tilde{u}_i^*(\psi_1)$ are passed directly to the EL ratio, requiring no pseudo-response and no refitting. Reusing the same $v^{(b)}$ across all ψ_1 within a replicate preserves the joint dependence structure of the process $\psi_1 \mapsto -2\log\hat{\mathcal{R}}(\psi_1)$, which justifies bootstrap calibration of the supremum.

4.5 Comparison of Likelihood-based Tests

Distributional assumptions The classical LRT assumes a fully specified parametric distribution and misspecification of any component affects both the test statistic and the bootstrap procedure. The EL test, by contrast, requires no distributional assumption, apart from, in the GLM setting, a correct specification of the conditional mean $\mu_i = g^{-1}(\eta_i)$. This makes the EL approach robust to misspecification, which follows as a key advantage of this test.

Continuity restriction and Homoscedasticity The classical LRT places no restriction on the behavior of the regression function at the breakpoint, while the EL test is constructed for continuous segmented regression models, requiring $z'_{\psi_1}(\delta_1 - \delta_2) = 0$ under H_1. Both tests are robust to heteroskedasticity: the classical LRT since it estimates σ_1^2 and σ_2^2 separately under H_1, and, the EL test because it operates through moment conditions rather than a full likelihood.

Calibration through bootstrap procedures In the linear case, a wild bootstrap procedure is proposed, while in GLMs, a parametric and multiplier wild bootstrap were developed for the classical parametric LRT and for the EL test, respectively.

Power comparison When the distributional assumption holds, the classical LRT attains maximum power asymptotically. In the linear model, under correct specification, the two tests share the same large sample power. Liu and Qian (2009) show that this test gives a statistic that is asymptotically equivalent to the LRT statistic, by proving that the power cost

of using nonparametric weights rather than the true likelihood is asymptotically negligible when the model is correctly specified and as $n \to \infty$. In finite samples, the EL test is more conservative, trading a potential finite-sample power loss for a clear robustness advantage when the error distribution is uncertain.

Computational cost Both tests scan over admissible candidates and require fitting M_1 for each $\psi_1 \in \Psi_{\varepsilon,1}$. The classical LRT requires $1 + |\Psi_{\varepsilon,1}|$ fits in total: one for M_0, plus one of $M_1(\psi_1)$ for each candidate via a single OLS or IRLS run on the expanded design matrix $X^* = (X \mid Z)$. The EL test requires between $2|\Psi_{\varepsilon,1}|$ and $3|\Psi_{\varepsilon,1}|$ fits. For each candidate, two separate OLS or IRLS fits run on $R_1(\psi_1)$ and $R_2(\psi_1)$, independently, since a run on X^* would not produce identified $(\hat{\delta}_1(\psi_1), \hat{\delta}_2(\psi_1))$ coefficients, which are needed for the null-consistent residual construction. If the continuity condition fails, a third fit on the full sample is required to enforce continuity via the hinge parametrization. In the worst case, where continuity fails at every candidate, the total number of fits reaches $3|\Psi_{\varepsilon,1}|$. In the linear setting, the difference in fit counts is of little practical consequence, since OLS fits admit closed-form solutions and are computationally cheap. However, in the GLM setting the computational difference becomes meaningful since each fit requires running IRLS to convergence. For large admissible sets or slow-converging IRLS runs this difference can be substantial in practice.

4.6 CUSUM Fluctuation Tests

The Cumulative Sum-based tests rather than scanning over candidate partitions, monitor the cumulative behavior of a residual or score process across the full sample and detect structural change through departures of that process from its null mean of zero. No per-candidate statistic is computed and no admissible set is scanned, instead, the test statistic is the supremum of the process norm over the sample index.

Three CUSUM-based tests are covered. The recursive residual CUSUM is a linear-only construction that accumulates recursive residuals sequentially, and requires constant variance across regimes. The Score CUSUM and Score MOSUM are built from the cumulative and moving sums of score contributions evaluated at the null estimates, and applying without modification to any EDF-based GLM family, including, in particular, the linear changepoint regression model.

4.6.1 Recursive Residual CUSUM

Let $k = p + q$ denote the total number of regressors, $x_i^{*(0)} = (x_i', z_i')' \in \mathbb{R}^k$ the full null covariate vector, and $X^{*(0)} = (X \mid Z) \in \mathbb{R}^{n \times k}$ the null design matrix. Since under H_0 there are no regimes and z_i enters with a single coefficient vector δ, $X^{*(0)}$ is not block-diagonal. Let $\hat{\theta}^{(0)} = (\hat{\beta}^{(0)\prime}, \hat{\delta}^{(0)\prime})'$ denote the OLS estimators obtained by fitting $M_0 : y_i = x_i'\beta + z_i'\delta + u_i$ on the full sample.

The full-sample OLS residuals $\hat{u}_i^{(0)} = y_i - x_i^{*(0)\prime}\hat{\theta}^{(0)}$, for $i = 1, \ldots, n$, satisfy the k moment conditions imposed by the OLS estimation, so that $\sum_{i=1}^n x_i \hat{u}_i^{(0)} = 0$ and $\sum_{i=1}^n z_i \hat{u}_i^{(0)} = 0$ by construction. Since each residual depends on all n observations through the full-sample estimate $\hat{\theta}^{(0)}$, the partial sums of these residuals do not behave like cumulative sums of independent quantities under the null, and their use in constructing a CUSUM process introduces an unwanted cross-observation dependence. To avoid this problem Brown et al. (1975) proposed working instead with recursive/forward residuals (one-step-ahead prediction errors), which are serially uncorrelated under H_0.

The Recursive Residual Process Denote the null OLS estimator of θ, using the first $k = p + q$ observations, as $\hat{\theta}_k^{(0)}$

$$\hat{\theta}_k^{(0)} = \left(X_k^{*(0)\prime} X_k^{*(0)}\right)^{-1} X_k^{*(0)\prime} Y_k, \qquad \text{with} \qquad Y_k = (y_1, \ldots, y_k)$$

The recursive residuals w_i at observations $i = k+1, \ldots, n$ are the standardized one-step-ahead prediction errors of y_i given the observations $1, \ldots, i-1$:

$$w_i = \frac{y_i - x_i^{*(0)\prime}\hat{\theta}_{i-1}^{(0)}}{\sqrt{1 + x_i^{*(0)\prime}(X_{i-1}^{*(0)\prime} X_{i-1}^{*(0)})^{-1} x_i^{*(0)}}}, \qquad i = k+1, \ldots, n,$$

where $\hat{\theta}_{i-1}^{(0)} = (\hat{\beta}_{i-1}', \hat{\delta}_{i-1}')'$ is the null OLS estimator based on the first $i-1$ observations and $X_{i-1}^{*(0)} \in \mathbb{R}^{(i-1)\times k}$ is the corresponding submatrix formed by the first $i-1$ rows of the null design matrix. The index starts at $k+1$ since at least k observations are needed to form a valid prediction.

Under H_0, the recursive residuals $w_{k+1}, \ldots, w_n$ satisfy: (i) zero mean, $\mathrm{E}(w_i) = 0$; (ii) constant variance, $\mathrm{Var}(w_i) = \sigma^2$; (iii) serial uncorrelatedness, $\mathrm{Cov}(w_i, w_j) = 0$ for $i \neq j$; and (iv) gaussianity, $w_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ if $u_i \sim \mathcal{N}(0, \sigma^2)$. Properties (i)-(iii) hold without any distributional assumptions, requiring only independence and homoskedasticity.

Efficient Computation via the Sherman-Morrison Update Computing $(X_i^{*(0)\prime} X_i^{*(0)})^{-1}$ at each step would be computationally demanding and can be avoided through the Sherman-Morrison formula, which updates the inverse recursively and only requires the computation of the initial inverse, $(X_k^{*(0)\prime} X_k^{*(0)})^{-1}$

$$\left(X_i^{*(0)\prime} X_i^{*(0)}\right)^{-1} = \left(X_{i-1}^{*(0)\prime} X_{i-1}^{*(0)}\right)^{-1} - \frac{\left(X_{i-1}^{*(0)\prime} X_{i-1}^{*(0)}\right)^{-1} x_i^{*(0)} x_i^{*(0)\prime} \left(X_{i-1}^{*(0)\prime} X_{i-1}^{*(0)}\right)^{-1}}{1 + x_i^{*(0)\prime} \left(X_{i-1}^{*(0)\prime} X_{i-1}^{*(0)}\right)^{-1} x_i^{*(0)}},$$

The OLS estimator is correspondingly updated as

$$\hat{\theta}_i^{(0)} = \hat{\theta}_{i-1}^{(0)} + \left(X_i^{*(0)\prime} X_i^{*(0)}\right)^{-1} x_i^{*(0)} \left(y_i - x_i^{*(0)\prime} \hat{\theta}_{i-1}^{(0)}\right)$$

CUSUM Test Under H_0, the partial sum of recursive residuals, $S_i = \sum_{r=k+1}^{i} w_r$, has mean zero and variance $(i-k)\sigma^2$. Therefore, consider the standardized CUSUM process

$$C(i) = \frac{S_i}{\hat{\sigma}\sqrt{n-k}}, \qquad \hat{\sigma} = \left[\frac{1}{n-k} \sum_{i=k+1}^{n} w_i^2\right]^{1/2} \qquad i = k+1, \ldots, n,$$

where $\hat{\sigma}$ is the sample standard deviation of the recursive residuals.

In the pre-break regime $R_1(\psi_1)$, the recursive $\hat{\theta}_{i-1}$ estimates are trained entirely on observations from the first segment, θ_1^0, so predictions are accurate and w_i fluctuates randomly around zero. Once the break occurs at ψ_1, $\hat{\theta}_{i-1}$ has not yet absorbed the shift to θ_2^0, so predictions become systematically biased, and the w_i for i just after ψ_1 acquire a non-zero mean, causing S_i to drift away from zero and $C(i)$ to excurse. As more post-break observations accumulate, $\hat{\theta}_{i-1}^{(0)}$ gradually adjusts toward θ_2^0, the bias in individual w_i diminishes, and the increments to S_i become smaller. Eventually, once $\hat{\theta}_{i-1}^{(0)}$ has fully adapted, the w_i return to fluctuating around zero and S_i stops drifting, wandering randomly rather than systematically growing further. The process $C(i)$ therefore does not increase monotonically after the break, it rises during the adaptation period and then levels off, so the supremum of $|C(i)|$ is attained somewhere in the post-break segment during the transition.

The test statistic and breakpoint estimator are, respectively,

$$\text{CUSUM} = \sup_{i \in \{k+1, \ldots, n\}} |C(i)|, \qquad \hat{\psi}_1 = \arg \max_{i \in \{h, \ldots, n-h\}} |C(i)|,$$

where $h = \max(k+1, \lfloor \varepsilon n \rfloor)$ arises as a trimming condition to ensure enough observations for the initialization of the recursion and regime estimation.

The recursive residual CUSUM is a signal detection procedure, not an estimation one, therefore, the argmax breakpoint estimator is a byproduct not designed for precision, and the CUSUM-based breakpoint estimates are expected to perform worse than those derived from likelihood-based tests.

Asymptotic Distribution Under the Null Under H_0 and standard regularity conditions on the regressors and the error distribution, Brown et al. (1975) show that

$$\{C(\lfloor n\tau \rfloor) : \tau \in [0,1]\} \xrightarrow{d} \{B^0(\tau) : \tau \in [0,1]\} \quad \text{as } n \to \infty,$$

where $B^0(\tau)$ is a standard Brownian bridge, as in Section 4.2.2.1. The boundary conditions $B^0(0) = B^0(1) = 0$ reflect the fact that, under H_0, $C(k) = 0$ at initialization and $C(n) \approx 0$ asymptotically, since the recursive residuals do not satisfy a global moment condition. The supremum $\sup_\tau |B^0(\tau)|$ follows a Kolmogorov-Smirnov distribution:

$$P\left(\sup_{0\leq\tau\leq 1} |B^0(\tau)| > \text{CUSUM}\right) = 2\sum_{r=1}^{\infty}(-1)^{r-1}e^{-2r^2\text{CUSUM}^2}$$

In `R`, the recursive residual CUSUM test is implemented in the `strucchange` package. The function `efp` computes the standardized CUSUM process and `sctest` performs the formal test, returning the test statistic and the asymptotic p-value.

4.6.2 Score CUSUM

The Score CUSUM, introduced by Hjort and Koning (2002) and unified in the GLM context by Zeileis (2005), replaces the recursive residuals of the CUSUM process with individual score contributions evaluated at the null estimates. This eliminates the need for sequential initialization and yields a test that applies to any EDF-based GLM, in particular to the linear changepoint regression model, recovering the same $p+q$ vector of independent Brownian bridges limit in all cases.

4.6.2.1 Individual Score Contributions

Under H_0, the model is fitted on the full sample, yielding $\hat{\theta}^{(0)} = (\hat{\beta}^{(0)\prime}, \hat{\delta}^{(0)\prime})'$, null fitted means $\hat{\mu}_i^{(0)} = g^{-1}(x_i^{*(0)\prime}\hat{\theta}^{(0)})$, IRLS weights $w_i^{(0)}$, and null working residuals $r_i^{*(0)} = (y_i - \hat{\mu}_i^{(0)})g'(\hat{\mu}_i^{(0)})$, as defined in Section 2.2.3.2. The full sample likelihood with respect to

$\theta = (\beta', \delta')'$ can be written as

$$\ell(\theta, \phi) = \sum_{i=1}^{n} \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right] = \sum_{i=1}^{n} \ell_i(\theta, \phi)$$

Following Section 2.2.3.2, the individual score contribution with respect to θ, evaluated at the null estimates $\hat{\theta}^{(0)}$ and null fitted means $\hat{\mu}_i^{(0)} = g^{-1}(x_i^{*(0)\prime} \hat{\theta}^{(0)})$ is

$$s_i(\hat{\theta}^{(0)}, \hat{\phi}^{(0)}) = \frac{\partial \ell_i(\theta, \phi)}{\partial \theta}\bigg|_{\hat{\theta}^{(0)}, \hat{\phi}^{(0)}} = \frac{y_i - \hat{\mu}_i^{(0)}}{a(\hat{\phi}^{(0)}) V(\hat{\mu}_i^{(0)}) g'(\hat{\mu}_i^{(0)})} x_i^{*(0)} \in \mathbb{R}^{p+q},$$

a vector with p elements corresponding to the derivative of ℓ_i with respect to β and q elements corresponding to the derivative with respect to δ.

Using the i-th IRLS weight and the i-th null working residual as defined in Section 2.2.3.2, the individual score contribution can be written as

$$s_i(\hat{\theta}^{(0)}, \hat{\phi}^{(0)}) = w_i^{(0)} r_i^{*(0)} x_i^{*(0)}, \quad w_i^{(0)} = \frac{1}{a(\hat{\phi}^{(0)}) V(\hat{\mu}_i^{(0)}) \left(g'(\hat{\mu}_i^{(0)}) \right)^2}, \quad r_i^{*(0)} = (y_i - \hat{\mu}_i^{(0)}) g'(\hat{\mu}_i^{(0)})$$

Under H_0, the score contributions satisfy $E(s_i(\hat{\theta}^{(0)}, \hat{\phi}^{(0)})) = 0$, the analogue of the zero-mean property of OLS residuals, and the IRLS first-order conditions force $\sum_{i=1}^{n} s_i(\hat{\theta}^{(0)}, \hat{\phi}^{(0)}) = 0$ exactly. The null Fisher information matrix, computed from the null fit, is $\hat{I}^{(0)} = X^{*(0)\prime} W^{(0)} X^{*(0)}$, where $W^{(0)} = \mathrm{diag}(w_1^{(0)}, \ldots, w_n^{(0)})$.

Under the Gaussian identity-link (the linear regression case), the IRLS weights reduce to $w_i^{(0)} = \hat{\sigma}^{-2}$ for all i, the working residuals reduce to $r_i^{*(0)} = y_i - \hat{\mu}_i^{(0)} = \hat{u}_i^{(0)}$, and the score contributions become $s_i = \hat{\sigma}^{-2} \hat{u}_i^{(0)} x_i^{*(0)}$, so that the Score CUSUM reduces to the covariate-weighted analogue of the OLS residual cumulative sum. The null Fisher information reduces to $\hat{I}^{(0)} = \hat{\sigma}^{-2} X^{*(0)\prime} X^{*(0)}$.

Score CUSUM Process The Score CUSUM process is the normalized cumulative sum of score contributions up to observation i

$$C^{\mathrm{score}}(i) = \frac{1}{\sqrt{n}} (\hat{I}^{(0)})^{-1/2} \sum_{l=1}^{i} s_l(\hat{\theta}^{(0)}, \hat{\phi}^{(0)}) = \frac{1}{\sqrt{n}} (\hat{I}^{(0)})^{-1/2} \sum_{l=1}^{i} w_l^{(0)} r_l^{*(0)} x_l^{*(0)}, \qquad i = 1, \ldots, n$$

Since $C^{\mathrm{score}}(i)$ is a vector in $\mathbb{R}^{p+q}$, the process simultaneously monitors for instability across all $p + q$ parameters. If only δ changes at the breakpoint while β remains stable, the bottom q components are expected to drift while the top p components remain near zero.

4.6.2.2 Asymptotic Distribution Under the Null

Under H_0 and standard GLM regularity condition, Hjort and Koning (2002) establish that

$$\left\{C^{\text{score}}(\lfloor n\tau \rfloor) : \tau \in [0,1]\right\} \xrightarrow{d} \left\{B^0(\tau) : \tau \in [0,1]\right\}, \qquad n \to \infty,$$

where $B^0(\tau) = (B_1^0(\tau), \ldots, B_{p+q}^0(\tau))'$ is the vector of $p+q$ independent standard Brownian bridges. This result holds for any EDF-based GLM without modification.

The standardized CUSUM process starts at $C^{\text{score}}(0) = 0$, since there are no score contributions accumulated yet, and returns to $C^{\text{score}}(n) = 0$ exactly, because the IRLS first order conditions force the total score to be zero at $\hat{\theta}^{(0)}$. This contrasts with the recursive residual CUSUM, where $C(n) \approx 0$ holds only asymptotically.

4.6.2.3 Test Statistic

Within the Hjort and Koning (2002) framework, both the supremum and the Cramér-von Mises functional are proposed as test statistics:

- **Supremum Statistic**: $\sup_{i=1,\ldots,n} ||C^{\text{score}}(i)||_2$, with null limit $\sup_\tau \|B^0(\tau)\|_2$.
- **CvM Statistic**: $\frac{1}{n}\sum_{i=1}^n C^{\text{score}}(i)' C^{\text{score}}(i)$, with null limit $\int_0^1 \|B^0(\tau)\|^2 d\tau$.

Since $\|C^{\text{score}}(i)\|_2$ already aggregates across the k components of the vector process, the Cramér-von Mises functional provides a natural complement that integrates the squared norm over the full sample rather than taking its maximum. Zeileis (2006) shows that both null limit distributions depend on $p+q$, and that critical values can be obtained from simulated limit distributions.

If H_0 is rejected, the breakpoint is estimated as

$$\hat{\psi}_1 = \arg \max_{i \in \{h,\ldots,n-h\}} \|C^{\text{score}}(i)\|_2, \qquad h = \lfloor \varepsilon n \rfloor.$$

In `R`, the Score CUSUM test is implemented in the `strucchange` package (Zeileis, 2006) by setting `type = "Score-CUSUM"` in the `efp` function. The formal test is performed via `sctest`, with `functional = "CvM"` and `functional = "maxBB"` implementing the Cramér-von Mises and supremum statistics, respectively.

4.6.3 Score MOSUM

The Score MOSUM, introduced by Chu et al. (1995) in the linear setting and extended to GLM by Zeileis (2005), replaces the cumulative sum of score contributions with a moving sum over a rolling window of fixed bandwidth $h = \lfloor bn \rfloor$, with $b \in (0,1)$ being the bandwidth. This test applies without modification to any EDF-based GLM family.

The Score MOSUM process is

$$M^{\text{score}}(i) = \frac{1}{\sqrt{h}} \left(\hat{I}^{(0)} \right)^{-1/2} \sum_{l=i-h+1}^{i} s_l(\hat{\theta}^{(0)}, \hat{\phi}^{(0)}), \qquad i = h+1, \ldots, n$$

At each index i, the process sums score contributions over the window $\{i-h+1, \ldots, i\}$ of h consecutive observations. As i advances, the window shifts forward, dropping the oldest contribution and adding the newest. The factor $\frac{1}{\sqrt{h}}$ replaces $\frac{1}{\sqrt{n}}$ in the Score CUSUM to reflect the fact that only h contributions are accumulated at each evaluation point.

Under H_0 and standard GLM regularity conditions (Zeileis, 2005), the Score MOSUM process satisfies

$$\{M^{\text{score}}(\lfloor n\tau \rfloor) : \tau \in [b,1]\} \xrightarrow{d} \left\{ \frac{W(\tau) - W(\tau - b)}{\sqrt{b}} : \tau \in [b,1] \right\} \qquad n \to \infty,$$

where $W(\tau)$ is a standard $p+q$-dimensional Wiener process, unlike the vector of $p+q$ independent standard Brownian bridges in the Score CUSUM limit.

Test Statistic and Breakpoint Estimator The test statistic and breakpoint estimator are, respectively, given by

$$\text{Score MOSUM} = \sup_{i \in \{h+1,\ldots,n\}} ||M^{\text{score}}(i)||_2, \qquad \hat{\psi}_1 = \left(\arg\max_{i \in \{h+1,\ldots,n\}} \|M^{\text{score}}(i)\|_2 \right) - \left\lfloor \frac{h}{2} \right\rfloor,$$

where the breakpoint is estimated at the center of the window achieving the supremum.

The asymptotic distribution of the supremum $\sup_{\tau \in [b,1]} |M^{\text{score}}(\lfloor n\tau \rfloor)|$, depends on both b and $p+q$, and critical values can be obtained by simulation.

Bandwidth Choice The bandwidth b is a tuning parameter, similar to the trimming fraction ε in the supremum-type tests in the sense that it controls the trade-off between sensitivity (power against local alternatives) and stability (control of spurious rejections). A smaller bandwidth makes the process more sensitive to sharp, localized breaks but also

increases variability under H_0, raising the risk of spurious rejections and favoring power at the cost of size control. A larger b smooths the process, reducing variability under H_0 but diluting the signal from a localized break by averaging it with surrounding in-regime residuals, favoring size control at the cost of potential power loss. Chu et al. (1995) recommend $b = 0.15$ as a default choice (also a default in the `strucchange` `R` package), however, this value should be modified if there are not enough observations for OLS estimation within each window, $h = \lfloor bn \rfloor < p + q$.

In `R`, the Score MOSUM test is implemented in the `strucchange` package (Zeileis, 2006), by setting `type = "Score-MOSUM"` in the `efp` function, with the bandwidth b controlled by the argument `h` and the test performed via `sctest`.

4.6.4 Comparison of CUSUM Fluctuation Tests

Recursive residual CUSUM versus Score-CUSUM Both are cumulative tests sharing the same Brownian bridge limit, but they differ in construction, residual input and finite-sample behavior. The recursive residual CUSUM uses recursive residuals w_i, which are genuine out-of-sample prediction errors, serially uncorrelated under H_0 by construction. The Score CUSUM instead uses full-sample score contributions, which satisfy the moment condition $\sum_{i=1}^{n} s_i(\hat{\theta}^{(0)}, \hat{\phi}^{(0)}) = 0$, introducing a finite-sample linear dependence among contributions, that vanishes asymptotically. Regarding the null distribution boundary, the Score CUSUM satisfies $C^{\text{score}}(n) = 0$ exactly by the IRLS first order conditions, whereas the recursive residual CUSUM satisfies $C(n) \approx 0$ only asymptotically, so the Score CUSUM more closely mimics the Brownian bridge boundary in finite samples.

The main practical disadvantages of the recursive residual CUSUM are that it requires sequential Sherman-Morrison updating, discards the first $p + q$ observations to initialize the recursion, and is restricted to the linear setting with no GLM extension. The Score CUSUM requires only a single full-sample fit, uses all n score contributions, is simpler to compute, better suited to small samples with large $p + q$, and extends without modification to any EDF-based GLM family.

Regarding power, the recursive residual CUSUM loses power for early breaks because too few pre-break observations are available for the recursive estimates to stabilize during initialization, possibly making the early recursive residuals noisy and diluting the signal. For late breaks, the recursive estimates are well calibrated to the pre-break regime but

too few post-break observations remain to accumulate evidence. Power is therefore maximized for middle breaks and is expected to deteriorate as the break moves toward either boundary. The Score CUSUM is robust to break location, since it uses the full-sample null fit rather than a recursively updated one, therefore, the signal accumulates regardless of whether the break is early, middle, or late.

Cumulative versus Moving Window The Score CUSUM accumulates residuals globally from the start of the sample to each evaluation point i. Under a break at ψ_1, post-break score contributions acquire a systematic mean displacement that accumulates over the entire post-break segment, building up a pronounced drift in the process. This makes the Score CUSUM most powerful for single breaks at any location, but vulnerable when two breaks of opposite sign and similar magnitude occur: the drift induced by the first break can be partially or fully canceled by the drift from the second, so the supremum may not be large enough to reject the null hypothesis of no break.

By contrast, the Score MOSUM evaluates only a local window of h consecutive observations at each point. Since each window is assessed independently, the test detects each shift as the window passes through it and is largely unaffected by such cancellation. The Score MOSUM is therefore more robust in multiple-break settings with opposite-sign shifts. The practical disadvantage of the Score MOSUM test is the choice of tuning parameter b, on which the critical values depend, whereas the Score CUSUM critical values require no tuning choice.

Remark 4.1. As mentioned, the cumulative CUSUM-based argmax breakpoint estimators are byproducts of a detection procedure not designed for estimation and the supremum tends to be attained well beyond the true break location. The Score MOSUM partially mitigates this via the window-centering shift $\lfloor h/2 \rfloor$. However, in either case, the CUSUM-based breakpoint estimates are expected to perform worse than those derived from likelihood-based tests.

4.7 Score-based Tests

The likelihood-based tests require fitting the alternative model $M_1(\psi_1)$ for every admissible candidate $\psi_1 \in \Psi_{\varepsilon,1}$, which in the GLM setting is computationally demanding, especially as n grows, since it means running IRLS to convergence at each step. Score-based tests offer a computationally attractive alternative, since they require fitting only M_0 once and evaluating the score of the unrestricted model (M_1) at the restricted null estimates,

without any additional IRLS iterations, reducing the total number of required IRLS fits, regardless of the size of the admissible set. Two score-based tests are developed here: the supremum score test of Davies (1977, 1987), which scans over candidates and addresses the resulting nonstandard null distribution via a conservative analytical bound, and the average score test of Muggeo (2016), which eliminates the unknown breakpoint before testing by averaging the hinge function over a grid of candidates, recovering a standard $\chi^2(q)$ null distribution. When applied in the linear setting, the score-based tests do not require homoskedasticity across regimes, since the IRLS weights $w_i^{(0)}$ absorb observation-specific variance and the dispersion σ^2 cancels in the normalised statistic.

4.7.1 Score Test in a Standard GLM

Before developing the changepoint score tests, it is useful to recall the standard score test in a GLM without breakpoints. The score test is based on the intuition that if $H_0 : \delta = 0$ is true, the gradient of the log-likelihood evaluated at the restricted IRLS estimates $(\hat{\beta}^{(0)}, \delta = 0)$ should be close to zero. If H_0 is false, the restricted estimator is far from the unconstrained maximum and the gradient will be large. The score test statistic is

$$S = s(\hat{\beta}^{(0)}, 0)' \mathcal{I}(\hat{\beta}^{(0)}, 0)^{-1} s(\hat{\beta}^{(0)}, 0),$$

where $s(\hat{\beta}^{(0)}, 0)$ and $\mathcal{I}(\hat{\beta}^{(0)}, 0)$ are the score vector and Fisher information matrix evaluated at the restricted estimates. Under H_0 and standard regularity conditions,

$$S \xrightarrow{d} \chi^2(q),$$

where $q = \dim(\delta)$. The key practical advantage is that only the restricted model needs to be fitted, and, therefore, no IRLS iterations under H_1 are required.

4.7.2 Supremum Score Test

Consider the GLM partial changepoint model with a single changepoint ψ_1. Using the hinge reparameterization, the linear predictor is written as:

$$\eta_i = x_i'\beta + z_i'\delta + \xi(z_i - \psi_1)_+, \quad i = 1, \ldots, n,$$

where $(z_i - \psi_1)_+ = (z_i - \psi_1)\mathbf{1}(z_i > \psi_1)$ is the hinge function at ψ_1, δ the slope in regime 1 and $\delta_1 + \xi$ the slope in regime 2. The null and alternative hypotheses of no structural change correspond, respectively, to $H_0 : \xi = 0$ and $H_1 : \xi \neq 0$.

4.7.2.1 Test Statistic

Null Model, M_0 Under H_0, the model reduces to a standard GLM on the full sample. The restricted IRLS estimates from the full sample $(\hat{\beta}^{(0)}, \hat{\delta}^{(0)})$, the diagonal matrices of Section 2.2.3.2 $W^{(0)}$ and $G^{(0)}$ and the fitted means $\hat{\mu}^{(0)} = g^{-1}(X\hat{\beta}_0 + Z\hat{\delta}_0)$ are computed once on the full sample with null design matrix $X^{*(0)} = (X \mid Z)$ and are fully determined, since they do not depend on the candidate breakpoint ψ_1.

Alternative Model, $M_1(\psi_1)$ The alternative model appends the hinge covariate $U(\psi_1) = ((z_1 - \psi_1)_+, \ldots, (z_n - \psi_1)_+)' \in \mathbb{R}^{n \times q}$, to the design matrix, which becomes $X^* = (X \mid Z \mid U(\psi_1)) \in \mathbb{R}^{n \times (p+q(m+2))}$.

Example

Suppose $n = 20$, that there are two constant regressors one for the intercept and x_i and two regime-varying regressors z_{i1} and z_{i2}. Consider one breakpoint at $z_{ij} = 12$.

$$X^* = \left(\begin{array}{cc|cccc|cc} 1 & x_1 & z_{1,1} & z_{1,2} & 0 & 0 & (z_{1,1}-\psi_1)_+ & (z_{1,2}-\psi_1)_+ \\ 1 & x_2 & z_{2,1} & z_{2,2} & 0 & 0 & (z_{2,1}-\psi_1)_+ & (z_{2,2}-\psi_1)_+ \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{12} & z_{12,1} & z_{12,2} & 0 & 0 & (z_{12,1}-\psi_1)_+ & (z_{12,2}-\psi_1)_+ \\ 1 & x_{13} & 0 & 0 & z_{13,1} & z_{13,2} & (z_{13,1}-\psi_1)_+ & (z_{13,2}-\psi_1)_+ \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{20} & 0 & 0 & z_{20,1} & z_{20,2} & (z_{20,1}-\psi_1)_+ & (z_{20,2}-\psi_1)_+ \end{array}\right) \in \mathbb{R}^{20\times 8}$$

Score of the $M_1(\psi_1)$ **with respect to** ξ For each candidate breakpoint ψ_1, applying the chain rule with $\partial\eta_i/\partial\xi = (z_i - \psi_1)_+$, the derivative of the log-likelihood of the alternative model with respect to ξ yields

$$\frac{\partial \ell_i}{\partial \xi} = \frac{\partial \ell_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \xi} = \frac{y_i - \mu_i}{a(\phi)\, V(\mu_i)\, g'(\mu_i)} \cdot (z_i - \psi_1)_+$$

Therefore, the score of $M_1(\psi_1)$ with respect to ξ, evaluated at the previously computed null estimates $(\hat{\beta}^{(0)}, \hat{\delta}^{(0)}, \hat{\phi}^{(0)})$, null-consistent diagonal matrices and fitted means is

$$s_\xi(\hat{\beta}^{(0)}, \hat{\delta}^{(0)}, \hat{\phi}^{(0)} \mid \psi_1) = \sum_{i=1}^{n} (z_i - \psi_1)_+ \cdot \frac{y_i - \hat{\mu}_i^{(0)}}{a(\hat{\phi}^{(0)}) V(\hat{\mu}_i^{(0)}) g'(\hat{\mu}_i^{(0)})}$$

Equivalently, in matrix form, $s_\xi(\psi_1) = U(\psi_1)' W^{(0)} R^{*(0)}$, where $R^{*(0)} = G^{(0)}(Y - \hat{\mu}^{(0)})$ are the null working residuals computed once from the single full-sample IRLS fit. For each candidate $\psi_1 \in \Psi_{\varepsilon,1}$, the vector $U(\psi_1)$ changes, but the null quantities do not, so no additional IRLS iterations are required for any candidate partition.

Partial Information for $\xi(\psi_1)$ As previously computed, $E(\frac{\partial^2 \ell_i}{\partial \eta_i^2}) = -w_i$, applying the chain rule

$$\frac{\partial^2 \ell_i}{\partial \xi \ \partial \xi'} = \frac{\partial^2 \ell_i}{\partial \eta_i^2}(z_i - \psi_1)_+(z_i - \psi_1)'_+,$$

the Fisher information with respect to ξ is

$$\mathcal{I}_{\xi\xi} = -E\left[\frac{\partial^2 \ell}{\partial \xi \ \partial \xi'}\right] = \sum_{i=1}^{n} w_i(z_i - \psi_1)_+(z_i - \psi_1)'_+ = U(\psi_1)' W U(\psi_1)$$

Equivalently, defining $\theta = (\beta, \delta)$

$$\frac{\partial^2 \ell_i}{\partial \xi \ \partial \theta'} = \frac{\partial^2 \ell_i}{\partial \eta_i^2}(z_i - \psi_1)_+ x_i^{*'} \qquad \Rightarrow \qquad \mathcal{I}_{\xi\theta} = \sum_{i=1}^{n} w_i(z_i - \psi_1)_+ x_i^{*'} = U(\psi_1)' W X^*,$$

$$\frac{\partial^2 \ell_i}{\partial \theta \ \partial \theta'} = \frac{\partial^2 \ell_i}{\partial \eta_i^2} x_i^{*'} x_i^{*'} \qquad \Rightarrow \qquad \mathcal{I}_{\theta\theta} = \sum_{i=1}^{n} x_i^{*'} x_i^{*'} = X^{*'} W X^*$$

The full Fisher information matrix in $M_1(\psi_1)$, computed at the null estimates, is

$$\mathcal{I}_{full} = \begin{pmatrix} \mathcal{I}_{\xi\xi} & \mathcal{I}_{\xi\theta} \\ \mathcal{I}_{\theta\xi} & \mathcal{I}_{\theta\theta} \end{pmatrix} = \begin{pmatrix} U(\psi_1)' W^{(0)} U(\psi_1) & U(\psi_1)' W^{(0)} X^{*(0)} \\ X^{*(0)'} W^{(0)} U(\psi_1) & X^{*(0)'} W^{(0)} X^{*(0)} \end{pmatrix}$$

Since the coefficients $\theta^{(0)} = (\beta^{(0)}, \delta^{(0)})$ from the null model are estimated through the IRLS under H_0, the score test for ξ should account for the fact that $\theta = (\beta, \delta)$ is also unknown and estimated. The partial information captures the information about ξ after partialling out the uncertainty from having estimated θ. This is achieved through the Schur complement $\mathcal{I}_{\xi\xi \cdot \theta} = \mathcal{I}_{\xi\xi} - \mathcal{I}_{\xi\theta}\mathcal{I}_{\theta\theta}^{-1}\mathcal{I}_{\theta\xi}$

$$\mathcal{I}_{\xi\xi \cdot \theta} = U(\psi_1)' W^{(0)} U(\psi_1) - U(\psi_1)' W^{(0)} X^{*(0)} (X^{*(0)'} W^{(0)} X^{*(0)})^{-1} X^{*(0)'} W^{(0)} U(\psi_1)$$

Defining the weighted hinge vector $U_w(\psi_1) = W^{(0)1/2} U(\psi_1)$ and the weighted hat matrix with the null estimates $H_w^{(0)} = W^{(0)1/2} X^{*(0)} (X^{*(0)'} W^{(0)} X^{*(0)})^{-1} X^{*(0)'} W^{(0)1/2}$, the partial information simplifies to

$$\mathcal{I}_{\xi\xi \cdot \theta^{(0)}}(\psi_1) = U_w(\psi_1)'(I_n - H_w^{(0)}) U_w(\psi_1)$$

Score test statistic for a fixed candidate The score test statistic for a fixed candidate ψ_1 is then

$$T(\psi_1) = s_{\tilde{\xi}}(\psi_1)' \mathcal{I}_{\tilde{\xi}\tilde{\xi}\cdot\theta^{(0)}}(\psi_1)^{-1} s_{\tilde{\xi}}(\psi_1)$$

Under H_0 and for a fixed known ψ_1, standard regularity conditions give $T(\psi_1) \xrightarrow{d} \chi^2(q)$, where q is the dimension of ξ.

Supremum score test statistic and changepoint estimate Davies (1977, 1987) suggests taking the supremum, over an admissible set of candidate breakpoints, of the score test statistic. If H_0 is rejected, the breakpoint is estimated as the candidate location achieving the supremum of the score statistic. Thus,

$$\mathcal{T} = \sup_{\psi_1 \in \Psi_{\varepsilon,1}} T(\psi_1), \qquad \hat{\psi}_1 = \arg\max_{\psi_1 \in \Psi_{\varepsilon,1}} T(\psi_1).$$

4.7.2.2 Non-standard Distribution and Davies Upper Bound

As explored in Section 4.1.3, the null distribution is nonstandard and critical values could be obtained via a bootstrap procedure. However Davies (1977, 1987) derives an analytical upper bound on the true p-value of the observed supremum $\mathcal{T}_{\text{obs}}$, that requires a correction that explicitly accounts for the maximization over ψ_1.

Combining the Bonferroni-type decomposition of the event $\{\sup T \geq u\}$ with the Rice (1944) formula for the expected number of level crossings of a smooth random process, for any fixed reference point $\psi_1^* \in \Psi_{\varepsilon,1}$,

$$P\left(\sup_{\psi_1 \in \Psi_{\varepsilon,1}} T(\psi_1) \geq u\right) \leq P(T(\psi_1^*) \geq u) + E[N_u^+],$$

where N_u^+ is the number of upcrossings of the level u by the process $T(\psi_1)$ on $\Psi_{\varepsilon,1}$. For a $\chi^2(q)$ process, Davies (1987) computes the expected upcrossing rate explicitly and obtains

$$E[N_u^+] = g(u,q) \cdot V, \qquad g(u,q) = \frac{u^{(q-1)/2} \exp(-u/2)}{2^{q/2} \Gamma(q/2)}$$

Here $g(u,q)$ is a density-type correction derived from the $\chi^2(q)$ density, while V is a process specific variability parameter that summarizes the total fluctuation of the underlying Gaussian score process over $\Psi_{\varepsilon,1}$ and is typically intractable. Using a plug-in estimator $\hat{V}$ constructed directly from the observed values of $T(\psi_1)$ and setting $u = \mathcal{T}_{\text{obs}}$ yields the

bound:

$$p = P\left(\sup_{\psi_1 \in \Psi_{\varepsilon,1}} T(\psi_1) \geq \mathcal{T}_{\text{obs}} \middle| H_0 \right) \leq p_0 + \hat{V} \cdot g(\mathcal{T}_{\text{obs}}, q),$$

where

- $p_0 = P(\chi^2(q) \geq \mathcal{T}_{\text{obs}})$ is the naive $\chi^2(q)$ p-value that ignores the maximization and would be valid if ψ_1 were identified.
- $\hat{V} \cdot g(\mathcal{T}_{\text{obs}}, q) \geq 0$ is an explicit penalty for having maximized over $\psi_1 \in \Psi_{\varepsilon,1}$. With

$$\hat{V} = \sum_{i=1}^{K-1} \left| \sqrt{T(\psi_1^{(i+1)})} - \sqrt{T(\psi_1^{(i)})} \right| \quad \text{and} \quad g(\mathcal{T}_{\text{obs}}, q) = \frac{\mathcal{T}_{\text{obs}}^{(q-1)/2} \exp(-\mathcal{T}_{\text{obs}}/2)}{2^{q/2}\ \Gamma(q/2)}$$

The Davies bound is conservative by construction, it controls the type-I error at the nominal level and $\hat{p}$ is an analytical upper bound on the true tail probability under H_0. This means that whenever we reject H_0 because $\hat{p} < \alpha$, the true p-value is also below α, therefore type I error is controlled. However the true p-value can be below α while the inflated $\hat{p}$ can be above α, which leads to a failure in rejection. This potential power loss comes as a cost of the test's conservativeness.

In R, the Supremum score test is implemented in the `segmented` package (Muggeo et al., 2008) via `davies.test()`.

4.7.2.3 Comparison with Classical Likelihood Ratio Test

The score principle also clarifies the relationship between the linear and GLM test families. In the linear Gaussian model, the score and likelihood ratio statistics are asymptotically equivalent, so the supremum score test coincides with the classical LRT, and in the homoskedastic setting with the sup-F test. However, in the linear framework, the supremum score test offers little practical advantage, since OLS fits admit closed-form solutions, evaluating $1 + |\Psi_{\varepsilon,1}|$ of them instead of 1 carries negligible computational cost relative to the GLM setting, where the IRLS fits to convergence are computationally burdensome. The supremum score test is therefore of interest primarily in the GLM setting, where the equivalence breaks down and the advantages over the classical Likelihood Ratio Test (LRT) are concrete.

In the GLM setting, the two tests share the same scanning strategy, both evaluate a per-candidate statistic over $\Psi_{\varepsilon,1}$ and take the supremum, and both produce the same

form of breakpoint estimator, $\hat{\psi}_1 = \arg\max_{\psi_1 \in \Psi_{\varepsilon,1}} T(\psi_1)$ or $\lambda(\psi_1)$ respectively. Both also have nonstandard null distributions requiring calibration. They differ in three fundamental respects.

First, the per-candidate statistic and the associated computational cost differ. The classical LRT uses $\lambda(\psi_1) = 2[\ell_1(\psi_1) - \ell_0]$, requiring a full IRLS fit of $M_1(\psi_1)$ for each candidate, giving $1 + |\Psi_{\varepsilon,1}|$ IRLS fits in total. The score test uses $T(\psi_1) = s_{\tilde{\xi}}(\psi_1)' \mathcal{I}_{\tilde{\xi}\tilde{\xi}\cdot\theta^{(0)}}(\psi_1)^{-1} s_{\tilde{\xi}}(\psi_1)$, evaluated at the null estimates via a matrix product, a single IRLS fit of the null model is required. This follows as the biggest advantage of the supremum score test relative to the classical LRT in the GLM setting. Secondly, while $\lambda(\psi_1)$ and $T(\psi_1)$ are asymptotically equivalent for each fixed ψ_1 under standard GLM regularity conditions, this pointwise equivalence does not extend to the suprema. The processes $\{\lambda(\psi_1)\}_{\psi_1 \in \Psi_{\varepsilon,1}}$ and $\{T(\psi_1)\}_{\psi_1 \in \Psi_{\varepsilon,1}}$ may have different covariance structures across candidates, so $\sup \lambda$ and $\sup T$ are not asymptotically equivalent in general and can lead to different rejection decisions. Thirdly, the calibration strategies differ. The classical LRT uses a parametric bootstrap to approximate the null distribution of $\sup \lambda$, while the supremum score test uses the conservative Davies analytical bound. The bootstrap is exact in principle but computationally expensive, the Davies bound is cheap but conservative, potentially inflating the p-value and reducing power.

4.7.3 Average Score Test

The Davies bound provides valid inference for the supremum score statistic without requiring any additional IRLS fits, but it does so at the cost of conservatism. Muggeo (2016) proposes a different strategy that sidesteps the supremum problem entirely, not implementing within the common strategy for unknown breakpoints established in Section 4.1.2. Rather than scanning the score over the admissible set and correcting for the maximum, the unknown breakpoint is eliminated before any test is constructed, by replacing the unknown hinge $(z_i - \psi_1)_+$ with its average over a grid of candidate values. This produces a single fixed covariate that aggregates threshold information across the grid, reducing the test to a standard one-coefficient score test in an augmented GLM with an asymptotic ξ^2 null distribution that requires no correction, no simulation, and no bootstrap. This test applies to any EDF-based GLM, in particular to the linear changepoint regression model.

4.7.3.1 Average Hinge

Muggeo (2016) considers K candidate breakpoints, $\{\psi_1, \ldots, \psi_K\}$ over an admissible candidate set $\Psi_{\varepsilon,1}$ and defines the averaged hinge for each observation i as a vector of dimension q (the dimension as z_i)

$$A_i = \frac{1}{K}\sum_{k=1}^{K}(z_i - \psi_k)_+, \qquad i = 1, \ldots, n,$$

which replaces the hinge term in η_i for that i-th observation, as a fully determined quantity that does not depend on any unknown parameter.

$$\eta_i = x_i'\beta + z_i'\delta + A_i'\xi \qquad i = 1, \ldots, n$$

$A = (A_1', \ldots, A_n')' \in \mathbb{R}^{n\times q}$ is a known covariate, which can be appended to the design matrix, $X^* = (X \mid Z \mid A) \in \mathbb{R}^{p+q(m+2)}$.

Example

Suppose $n = 20$, that there are two constant regressors for the intercept and x_i, and two regime-varying regressors z_{i1} and z_{i2}. Consider one breakpoint at $z_{ij} = 12$.

$$X^* = \left(\begin{array}{cc|cccc|cc}
1 & x_1 & z_{1,1} & z_{1,2} & 0 & 0 & \frac{1}{K}\sum_{k=1}^{K}(z_{1,1}-\psi_k)_+ & \frac{1}{K}\sum_{k=1}^{K}(z_{1,2}-\psi_k)_+ \\
1 & x_2 & z_{2,1} & z_{2,2} & 0 & 0 & \frac{1}{K}\sum_{k=1}^{K}(z_{2,1}-\psi_k)_+ & \frac{1}{K}\sum_{k=1}^{K}(z_{2,2}-\psi_k)_+ \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & x_{12} & z_{12,1} & z_{12,2} & 0 & 0 & \frac{1}{K}\sum_{k=1}^{K}(z_{12,1}-\psi_k)_+ & \frac{1}{K}\sum_{k=1}^{K}(z_{12,2}-\psi_k)_+ \\
1 & x_{13} & 0 & 0 & z_{13,1} & z_{13,2} & \frac{1}{K}\sum_{k=1}^{K}(z_{13,1}-\psi_k)_+ & \frac{1}{K}\sum_{k=1}^{K}(z_{13,2}-\psi_k)_+ \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & x_{20} & 0 & 0 & z_{20,1} & z_{20,2} & \frac{1}{K}\sum_{k=1}^{K}(z_{20,1}-\psi_k)_+ & \frac{1}{K}\sum_{k=1}^{K}(z_{20,2}-\psi_k)_+
\end{array}\right) \in \mathbb{R}^{20\times 8}$$

4.7.3.2 Test Statistic

Since the average hinge A_i is fully defined, M_1 reduces to a standard GLM with the augmented predictor and, testing $H_0 : \xi = 0$ is a standard test on q regression coefficients.

Null model Under H_0, M_0 has no regimes and reduces to a full sample GLM with design matrix $X^{*(0)} = (X \mid Z)$. Denote $(\hat\beta^{(0)}, \hat\delta^{(0)}, \hat\phi^{(0)})$ as the IRLS estimates of (β, δ, ϕ), under H_0, with fitted means $\hat\mu_i^{(0)}$, null IRLS weight matrix $W^{(0)}$, null working residuals $r_i^{*(0)}$ and null weighted hat matrix $H_w^{(0)}$.

Alternative model Following the chain rule with $\partial\eta_i/\partial\xi = A_i$,

$$\frac{\partial \ell_i}{\partial \xi} = \frac{\partial \ell_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \xi} = \frac{y_i - \mu_i}{a(\phi)V(\mu_i)g'(\mu_i)} \cdot A_i$$

Therefore, the score of the alterative model with respect to ξ, evaluated at the previously computed null quantities, yields

$$s_\xi(\hat{\beta}^{(0)}, \hat{\delta}^{(0)}, \hat{\phi}^{(0)}) = \sum_{i=1}^{n} \frac{y_i - \hat{\mu}_i^{(0)}}{a(\hat{\phi}^{(0)})V(\hat{\mu}_i^{(0)})g'(\hat{\mu}_i^{(0)})} \cdot A_i$$

In matrix form, $s_\xi(\hat{\beta}^{(0)}, \hat{\delta}^{(0)}, \hat{\phi}^{(0)}) = A'W^{(0)}R^{*(0)}$.

Partial Fisher Information Since, $E\left(\frac{\partial^2 \ell_i}{\partial \eta_i^2}\right) = -w_i$, applying the chain rule

$$\frac{\partial^2 \ell_i}{\partial \xi \partial \xi'} = \frac{\partial^2 \ell_i}{\partial \eta_i^2} A_i A_i',$$

The Fisher information with respect to ξ is

$$\mathcal{I}_{\xi\xi} = -E\left[\frac{\partial^2 \ell_i}{\partial \xi \partial \xi'}\right] = \sum_{i=1}^{n} w_i A_i A_i' = A'WA$$

Applying the chain rule through $E\left[\frac{\partial^2 \ell_i}{\partial \eta_i^2}\right] = -w_i$ and defining $\theta = (\beta, \delta)$ yields

$$\frac{\partial^2 \ell_i}{\partial \xi \partial \theta'} = \frac{\partial^2 \ell_i}{\partial \eta_i^2} A_i x_i^{*'} \qquad \Rightarrow \qquad \mathcal{I}_{\xi\theta} = A'WX^*$$

$$\frac{\partial^2 \ell_i}{\partial \theta \partial \theta'} = \frac{\partial^2 \ell_i}{\partial \eta_i^2} x_i^* x_i^{*'} \qquad \Rightarrow \qquad \mathcal{I}_{\xi\theta} = X^{*'}WX^*$$

Thus, the full Fisher information matrix for the $M_1(\psi_1)$, computed at the null estimates is

$$\mathcal{I}_{full} = \begin{pmatrix} I_{\xi\xi} & I_{\xi\theta^{(0)}} \\ I_{\theta^{(0)}\xi} & I_{\theta^{(0)}\theta^{(0)}} \end{pmatrix} = \begin{pmatrix} A'W^{(0)}A & A'W^{(0)}X^{*(0)} \\ X^{*(0)'}W^{(0)}A & X^{*(0)'}W^{(0)}X^{*(0)} \end{pmatrix},$$

The partial Fisher information for ξ after removing the contribution of the null model parameters $\theta^{(0)} = (\beta^{(0)}, \delta^{(0)})$, through the Shur complement is

$$\mathcal{I}_{\xi\xi\cdot\theta^{(0)}} = A'W^{(0)}A - A'W^{(0)}X^{*(0)}(X^{*(0)'}W^{(0)}X^{*(0)})^{-1}X^{*(0)'}W^{(0)}A$$

Defining the weighted version of A as a weighted matrix $A_w = W^{(0)1/2}A$. The partial Fisher information can be rewritten as $\mathcal{I}_{\xi\xi\cdot\theta^{(0)}} = A_w'(I_n - H_w^{(0)})A_w$.

Average score statistic Therefore, applying the formula of the score test

$$s_0 = s_\xi' \mathcal{I}^{-1}_{\xi\xi\cdot\theta^{(0)}} s_\xi = (A'W^{(0)}R^{*(0)})' (A_w'(I_n - H_w^{(0)})A_w)^{-1} (A'W^{(0)}R^{*(0)})$$

Equivalently in the weighted space, using $R_w^{*(0)} = W^{(0)1/2}R^{*(0)}$, the statistic becomes

$$s_0 = (A_w'R_w^{*(0)})' (A_w'(I_n - H_w^{(0)})A_w)^{-1} (A_w'R_w^{*(0)})$$

Under H_0 and standard regularity conditions $s_0 \xrightarrow{d} \chi^2(q)$.

Note that the average score test does not produce a breakpoint estimator, unlike the supremum score test. In R, the average score test is implemented in the `segmented` package (Muggeo et al., 2008) via `pscore.test()`.

4.7.4 Comparison of the Score-Based Tests

The supremum and average score tests share the same null model fit, the same score-based construction, and the same single IRLS fit requirement. However, they differ in how they handle the unknown breakpoint ψ_1 and the inferential problem it creates.

Test statistic magnitude By construction, the supremum statistic $\mathcal{T} = \sup_{\psi_1 \in \Psi_{\varepsilon,1}} T(\psi_1)$ is always at least as large as the average score one, since taking the supremum over candidates ψ_k can only exceed or equal an average. The supremum score test is more sensitive to a signal concentrated at a single candidate, while the average score test accumulates evidence spread across the grid.

Null distribution and inference The supremum score test has a nonstandard null distribution, requiring the conservative Davies analytical bound, which inflates the p-value relative to the true tail probability. The average score test sidesteps this problem entirely, recovering a standard $\chi^2(q)$ null distribution that requires no correction, no simulation, and no bootstrap. The validity of this standard distribution rests on the score contributions being asymptotically uncorrelated across well-separated candidates, which holds when the grid $\{\psi_1, \ldots, \psi_K\}$ is sufficiently spread over $\Psi_{\varepsilon,1}$.

Power The potential cost of averaging is signal dilution and power loss. If the true breakpoint is close to one candidate ψ_k, that candidate produces a large hinge term $(z_i - \psi_k)_+$, while the remaining candidates contribute small near zero amounts, dragging the average down. Therefore averaging across all K candidates dilutes this concentrated signal that the supremum would have captured directly. However, Muggeo (2016) shows through simulation that in most realistic scenarios the average-based the average-based test achieves comparable or better power than the supremum-based test, because the potential power loss coming from the conservatism of the Davies bound more than offsets the loss from averaging. The supremum score test can, however, outperform the average score test when the break signal is strongly concentrated at a single candidate and $\hat{V}$ is small, so that the bound is tight and the penalty for maximization is small.

Computational cost Both tests require only a single IRLS fit of M_0, being computationally equivalent at the level of IRLS fits, however, the average score test has a slight edge since it avoids the per-candidate loop required for the supremum test, which evaluates $T(\psi_1)$ at each of the K candidates.

Implementation Both tests are implemented in the `segmented` package (Muggeo et al., 2008), which has undergone substantial development since its original release. The estimation methodology is rooted in Muggeo (2003), with implementation examples explored in Muggeo et al. (2008). The supremum score test is available via `davies.test()` and the average score test via `pscore.test()`.

4.8 Sequential Extensions For Multiple Breakpoints

Throughout this chapter, all tests have been developed for the single breakpoint case ($m = 1$). The extension to $m \geq 2$, follows the sequential framework of Bai and Perron (1998), a sequential estimation and testing procedure in which only a single break needs to be estimated and tested at each step, applied to the full sample and to successive subsamples. This section develops that framework, establishes its theoretical properties, and discusses which tests can be embedded within it and under what conditions.

1. Apply a single-break test to the full sample of n observations. If H_0 (no break) is rejected, estimate the dominant breakpoint $\hat{\psi}_1$ as the $\arg\max$ of the per-candidate test statistic over $\Psi_{\varepsilon,1}$.

2. Split the sample at $\hat{\psi}_1$ into regimes $R_1(\hat{\psi}_1)$ and $R_2(\hat{\psi}_1)$.
3. Apply the same single-break test independently to each subsample $R_j(\hat{\psi}_j)$, $j = 1, 2$, using the same trimming condition ε applied relative to the subsample size.
4. If H_0 is rejected in subsample R_j, estimate the dominant break within that subsample and split further.
5. Continue recursively until no further breaks are detected in any subsample.

The total number of detected breakpoints, $\hat{m}$ corresponds to the number of performed splits.

F-statistic based tests The sequential procedure is fully developed and theoretically justified for the sup-F test by Bai (1997) and Bai and Perron (1998), with its implementation in R via the `dosequa()` function in the `mbreaks` package (Nguyen et al., 2024). It is shown that when estimating a single-break model in the presence of multiple breaks, the break fraction estimate converges to one of the true break fractions, the dominant one, in the sense that accounting for it yields the greatest reduction in the sum of squared residuals. Sequential estimation therefore yields consistent estimators for all breakpoints, $\hat{\psi}_j^{\text{seq}} \xrightarrow{p} \psi_j^0$, $j = 1, \ldots, m^0$, at the same super-consistent rate as the GLS estimator, satisfying $\hat{\psi}_j^{\text{seq}} - \psi_j^0 = O_p(1)$. In the case of two equally dominant breaks, the estimate converges with probability $1/2$ to either break.

However, when the number of breaks m is determined by the sequential testing rule itself rather than assumed known, consistency of the breakpoint estimators holds only conditionally on correct selection of m^0. It is not in general established that $\Pr(\hat{m} = m^0) \to 1$, since the stopping rule relies on nonstandard test statistics and is influenced by finite-sample effects and multiple testing considerations.

Connection with other tests This sequential framework applies across all developed single-break tests, however the established theoretical properties for the sup-F have not been formally established and should be treated as heuristic.

4.9 Overall Comparison of Tests for Detecting Structural Change

The five test families developed in this chapter share the goal of testing H_0 of parameter stability against the alternative with at least one structural break at an unknown location.

They differ, however, in their regression setting, the type of break they are designed to detect, their distributional assumptions, their computational cost, and the validity of their null distributions. This section synthesizes those differences across families, focusing on considerations that are only visible when all five families are viewed together.

Regression setting The F-statistic based tests are restricted to the linear homoskedastic regression setting. Their construction relies on the closed-form OLS objective function and the resulting Brownian bridge limit, neither of which carries over to the GLM setting. The remaining four families all extend to the GLM setting, though with different degrees of theoretical support. The classical parametric LRT, the score-based tests, and the CUSUM-based tests, apart for the recursive residual CUSUM have been developed and justified in the GLM setting. The EL test is a proposed extension of Liu and Qian (2009) to the GLM setting, whose theoretical properties require additional development.

Type of structural change The tests are not equally sensitive to all forms of structural change. The sup-F test, the classical LRT, the EL test, and the supremum score test are all built around the supremum-over-candidates strategy of Section 4.1.2, concentrating power at the single candidate that best explains the data. They are therefore most powerful against sharp, well-localized breaks.

The CUSUM-based tests differ fundamentally from the supremum family in their aggregation strategy. The cumulative tests (recursive residual CUSUM and score CUSUM) accumulate residuals or score contributions from the start of the sample to each evaluation point. This global accumulation makes them powerful, in comparison to the moving-window test, against a single persistent break at any location, but reduces their power when two breaks of opposite sign are present, since the drifts can cancel in the cumulative sum. The score MOSUM, by evaluating on a local rolling window, detects each shift independently as the window passes through it, making the test more robust in a multiple-break setting.

Distributional assumptions The tests span a spectrum of distributional requirements, from fully parametric to nonparametric. In the GLM setting, the classical parametric LRT requires full specification of the EDF, including the link function $g(\cdot)$, variance function $V(\mu_i)$, and dispersion function $a(\phi)$. Misspecification of any component corrupts both the test statistic and the parametric bootstrap used to calibrate it. The score-based tests and

score CUSUM-based tests require correct specification of $g(\cdot)$ and $V(\mu_i)$, since both enter the score contributions and the Fisher information matrix, but do not require the full EDF to be specified. The EL test is a non-parametric approach that requires only correct specification of the conditional mean $\mu_i = g^{-1}(\eta_i)$, making it robust to misspecification of the variance function and dispersion. It, however, is the only test in which continuity at the break is imposed. The sup-F and recursive residual CUSUM tests both apply only in the linear homoskedastic setting. However, the sup-F test additionally relies on the full classical LINE assumptions (linearity, independence, normality, and equal variance), which are needed to deliver the exact finite-sample F-distribution of the per-candidate statistic $F(\psi_1)$. The recursive residual CUSUM test, by contrast, requires only linearity, independence, and equal variance (LIE), without normality, since its Brownian bridge limit is a purely asymptotic result, obtained by applying a functional central limit theorem to the partial-sum process of the recursive residuals, which establishes convergence to Brownian motion under independence and finite variance alone, regardless of the underlying error distribution.

Computational cost The number of IRLS fits required to evaluate the test statistic varies substantially across families and is the primary practical differentiator in the GLM setting. The score-based tests, require only a single IRLS fit of the null model, regardless of the size of the admissible set. The candidate breakpoints enter solely through the hinge covariate $U(\psi_1)$ or the average hinge A, which are evaluated by matrix products at the null estimates without any additional IRLS iterations. The score CUSUM-based tests in the GLM setting similarly require only a single null IRLS fit. In contrast, the classical parametric LRT requires $1 + |\Psi_{\varepsilon,1}|$ fits in total: one for M_0 and one per ψ_1 candidate, and the EL test is the most expensive, requiring two to three IRLS fits per candidate and thus $2|\Psi_{\varepsilon,1}|$ to $3|\Psi_{\varepsilon,1}|$ fits in total. For large admissible sets or for GLMs where IRLS convergence is slow, such as logistic regression near separation or Gamma regression with small dispersion, this difference can be substantial in practice.

Null distribution and calibration All tests that follow the supremum over candidates strategy produce a nonstandard null distribution, however the calibration strategies adopted across families differ. The F-based and the recursive residual CUSUM test share the same Brownian bridge limit, and asymptotic critical values are tabulated, avoiding any

simulation. In the score CUSUM and score MOSUM tests, the statistic converges, respectively, to a vector of Brownian bridges and a vector of Wiener processes. The classical parametric LRT and EL tests rely on bootstrap procedures — a wild bootstrap in the linear case and a parametric or multiplier wild bootstrap in the GLM case, which are asymptotically valid but computationally expensive. The supremum score test uses the Davies analytical upper bound on the true p-value that is cheap to compute but conservative by construction, inflating the p-value relative to the true tail probability. The average score test is the only test in this chapter that produces an exact asymptotic $\chi^2(q)$ null distribution requiring no correction, no simulation, and no bootstrap, at the cost of signal dilution from averaging the hinge over candidates rather than taking its supremum.

Table 2 summarizes the properties of all tests across these dimensions.

Table 2: Overall comparison of tests for detecting structural change.

Test	Reg. setting	Strategy	Cont. cond.	Homosk. (linear)	Distributional assumption	IRLS fits	Null distribution
Sup-F	Linear	Sup over candidates	No	Yes	Gaussian errors	–	Brownian bridge
LRT	GLM	Sup over candidates	No	No	Full EDF	$1 + \lvert\Psi_{\varepsilon,1}\rvert$	Bootstrap
EL	GLM	Sup over candidates	Yes	No	Mean only – robust to misspecification	$2\lvert\Psi_{\varepsilon,1}\rvert$–$3\lvert\Psi_{\varepsilon,1}\rvert$	Bootstrap
RR-CUSUM	Linear	Cumulative	No	Yes	Independent errors	–	Brownian bridge
Sc.CUSUM	GLM	Cumulative	No	No	Link + variance	1	Vector of Brownian bridges
Sc.MOSUM	GLM	Rolling window – advantage under multiple breaks	No	No	Link + variance	1	Vector of Wiener processes
Sup.Sc	GLM	Sup over candidates	No	No	Link + variance	1	Davies conservative upper bound
Av.Sc	GLM	Average over candidates	No	No	Link + variance	1	$\chi^2(q)$

5 Projection-Based Goodness-of-Fit Testing

Before drawing inferential conclusions from a fitted regression model, it is essential to assess whether it provides an adequate description of the data. Undetected misspecification may lead to biased parameter estimates, incorrect standard errors, and ultimately erroneous conclusions. This is the role of Goodness-of-Fit (GoF) testing.

This chapter develops a unified projection-based GoF testing framework that applies to any parametric regression model, standard or segmented, linear or GLM. The framework is rooted in the consistent and omnibus test of Escanciano (2006) for parametric regression with continuous responses, which is adapted to logistic regression by Liu et al. (2024) and extended here to segmented regression models. The central contribution is to show that the same test statistic, with the same computational form and the same bootstrap procedure, applies uniformly once the design matrix and projection dimension are adjusted for the model at hand.

5.1 The Goodness-of-Fit Problem

Let $(Y, X')'$ be a random vector in a $(d+1)$-dimensional Euclidean space, where Y is a real-valued response and $X \in \mathbb{R}^d$ is a vector of explanatory variables. Under $\mathrm{E}|Y| < \infty$ the regression function $m(x) = \mathrm{E}[Y \mid X = x]$ is well defined, and under $\mathrm{E}|Y|^2 < \infty$ it is the mean-square optimal predictor of Y given X. Parametric modeling postulates a family $\{f(\cdot, \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$ for m, so that $Y = f(X, \theta) + \varepsilon(\theta)$, with $\varepsilon(\theta) := Y - f(X, \theta)$, where $\varepsilon(\theta)$ represents the random disturbance of the model. The model is correctly specified if and only if there exists $\theta_0 \in \Theta$ such that the parametric family contains the true regression function $m(X) = f(X, \theta_0)$ almost surely (a.s.). In that case, $E(\varepsilon(\theta_0) \mid X) = 0$ a.s. The null hypothesis of the model being correctly specified is, therefore

$$H_0 : \ E(\varepsilon(\theta) \mid X) = 0 \text{ a.s. for some } \theta \in \Theta,$$

and the alternative is $H_1 : \ P(E(Y \mid X) \neq f(X, \theta)) > 0 \ \text{ for all } \ \theta \in \Theta$.

The test is omnibus, designed to detect a broad class of alternatives. A consequence of this generality is that rejection of H_0 provides evidence of misspecification without indicating its nature or direction.

5.2 The Integrated Testing Approach

The integrated (cumulative) approach replaces the conditional moment restriction $E(\varepsilon(\theta_0) \mid X) = 0$ by an infinite parametric family of unconditional restrictions:

$$H_0 : E(\varepsilon(\theta) \mid X) = 0 \text{ a.s.} \quad \Leftrightarrow \quad E\Big[\varepsilon(\theta)\, w(X, x)\Big] = 0, \quad \forall x \in \Pi, \text{ for some } \theta \in \Theta,$$

for a suitable parametric family $\{w(\cdot, x) : x \in \Pi\}$, over an index set Π.

The empirical counterpart of the moment condition in $E(\varepsilon(\theta)\, w(X, x))$ is the Residual Marked Empirical Process (RMEP), $R_{n,w}(x, \hat{\theta}) = n^{-1/2} \sum_{i=1}^{n} \hat{\varepsilon}_i\, w(X_i, x)$, where $\hat{\varepsilon}_i = Y_i - f(X_i, \hat{\theta})$ are the residuals from the fitted model. Under H_0, $R_{n,w}$ should fluctuate around zero, and any reasonable test rejects for large values of a norm of $R_{n,w}$, typically the Cramér-von Mises norm $\int_{\Pi} \left| R_{n,w} \right|^2 \Psi(dx)$ or the Kolmogorov-Smirnov norm $\sup_{x \in \Pi} \left| R_{n,w} \right|$.

The reference tests in the literature differ in their choice of w, and different choices deliver tests with different power properties. Bierens (1982) uses exponential weights $w(X, x) = \exp(ix'X)$, which enter the data only through the one-dimensional projections $x'X_i$, making the test robust to the dimension of X, at the cost of an arbitrary user-chosen integrating measure. Stute (1997) uses indicator weights $w(X, x) = 1(X \leq x)$, which removes this arbitrary choice by integrating against the empirical distribution of $\{X_i\}$, but depends on X as a d-dimensional object and therefore loses power as the dimension grows due to the curse of dimensionality. Both tests are consistent and asymptotically admissible (Bierens and Ploberger, 1997), and neither dominates the other uniformly. Escanciano (2006) combines the dimension-robustness of Bierens (1982) with the parameter-free integration of Stute (1997) through the projection-indicator family $w(X, x) = 1(\beta' X \leq u)$, $x = (\beta, u)$, which enters the data through one-dimensional projections while integrating against the empirical distribution of the projected regressors, requiring no user-chosen input.

5.3 The Escanciano Projection-Based Test

Escanciano (2006) proposed a consistent integrated test that is simple to compute, free of any user-chosen tuning parameter (no bandwidth, no integrating measure μ), robust to high-order dependence such as conditional heterocedasticity, suited to covariates of moderate or high finite dimension, and able to detect local alternatives converging to H_0 at the parametric rate $n^{-1/2}$.

5.3.1 Reduction to One-Dimensional Projections

A direct assessment of H_0 would require nonparametric estimation of the full multivariate conditional expectation $E(\varepsilon(\theta_0) \mid X)$, which may suffer from the curse of dimensionality when d (the dimension of X) is moderate or high. The following lemma, proposed by Jones (1987), reduces the multivariate conditional moment restriction to a family of univariate ones indexed by the direction β in the unit sphere $\mathbb{S}^d$.

Lemma 5.1. *For any* $\beta \in \mathbb{S}^d = \{\beta \in \mathbb{R}^d : \|\beta\| = 1\}$,

$$E(\varepsilon(\theta_0) \mid X) = 0 \text{ a.s. } \Leftrightarrow \; E(\varepsilon(\theta_0) \mid \beta' X) = 0 \text{ a.s.}$$

Since $\beta' X$ ranges over all linear combinations of X, such that $\beta' X \in \mathbb{R}$, this lemma allows us to construct consistent tests for H_0 based on one-dimensional projections. Escanciano (2006) further showed that each univariate conditional restriction is equivalent to an unconditional one involving indicator functions.

Lemma 5.2. *For any* $\beta \in \mathbb{S}_d$,

$$E(\varepsilon(\theta_0) \mid \beta' X) = 0 \text{ a.s.} \quad \Longleftrightarrow \quad E\big[\varepsilon(\theta_0)\, 1(\beta' X \le u)\big] = 0 \quad \text{for all } u \in \mathbb{R}.$$

Combining both of the lemmas H_0 becomes

$$H_0 : \; E\Big[\varepsilon(\theta)\, 1(\beta' X \le u)\Big] = 0 \quad \text{a.e. } (\beta, u) \in \Pi = \mathbb{S}^d \times \mathbb{R}, \quad \text{for some } \theta \in \Theta \subset \mathbb{R}^p.$$

This reformulation eliminates the need to estimate conditional expectations, replacing them with unconditional moment conditions that can be estimated directly from the data.

5.3.2 Test Statistic and Computational Form

The empirical analogue of the moment condition $E(\varepsilon(\theta)\, 1(\beta' X \le u))$ is the Residual Marked Process Based on Projections (RMPP), $R_n(\beta, u) = n^{-1/2} \sum_{i=1}^n \hat{\varepsilon}_i\, 1(\beta' X_i \le u)$, where $(\beta, u) \in \Pi = \mathbb{S}^d \times \mathbb{R}$. Under H_0, $R_n(\beta, u)$ should be close to zero for all (β, u), and systematic departures provide evidence against H_0. Let $F_{n,\beta}$ denote the empirical distribution function of the projected regressors $\{\beta' X_i\}_{i=1}^n$, and let $d\beta$ the uniform density on the unit sphere $\mathbb{S}^d$. The projected Cramér-von Mises statistic is

$$\mathrm{T}_{\mathrm{proj}} = \int_{-\infty}^{\infty} \int_{\mathbb{S}_d} \Big(R_n(\beta, u)\Big)^2 d\beta\, F_{n,\beta}(du)$$

The statistic aggregates evidence against H_0 across every projection direction simultaneously, and large values indicate misspecification. Therefore, the test rejects H_0 for large values of $\mathrm{T}_{\mathrm{proj}}$. The integrating measure is canonical since it is the product of the empirical distribution of the projected regressors and the uniform measure on the sphere, and it requires no user chosen input.

Escanciano (2006) avoids the direct computationally demanding evaluation of the double integral by deriving a closed-form expression that reduces $\mathrm{T}_{\mathrm{proj}}$ to a quadratic form. Defining the symmetric $n \times n$ matrix $A = (A_{ij})$ with entries $A_{ij} = \sum_{l=1}^{n} A_{ijl}$, where

$$A_{ijl} = C_d \left| \pi - \arccos\left(\frac{(X_i - X_l)'(X_j - X_l)}{\|X_i - X_l\| \, \|X_j - X_l\|} \right) \right|, \qquad C_d = \frac{\pi^{d/2-1}}{\Gamma(d/2+1)},$$

with the convention $A_{ijl} = 0$ when $X_i = X_l$ or $X_j = X_l$. The statistic reduces to the quadratic form

$$T_{\mathrm{proj}} = \frac{1}{n^2} \hat{\varepsilon}' A \, \hat{\varepsilon}, \qquad \hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)'$$

The matrix A depends only on the covariates $\{X_i\}$ and on the dimension d, it does not involve the response, the fitted parameters, or the choice of parametric family. The same quadratic form applies when using Pearson or working residuals in place of raw residuals. This statistic avoids both shortcomings identified in Section 5.2: there is no arbitrary integrating measure μ, and the data enter only through one-dimensional projections, so the curse of dimensionality and related loss of power does not apply.

5.3.3 Bootstrap procedures

The asymptotic null distribution of T_{proj} is non-standard and depends on the Data Generating Process (DGP), so critical values are obtained by bootstrap procedures. For linear regression models, Escanciano (2006) proposed a data-level wild bootstrap procedure. For the GLM case, as mentioned in Section 4.4.2.3, adding resampled residuals to fitted values may produce pseudo bootstrap responses outside of the admissible support of the GLM family, so a parametric bootstrap is used instead (Liu et al., 2024).

5.4 PBT Goodness-of-Fit Test for Segmented Regression Models

The aim of the present chapter is to extend the PBT of Section 5.3 to the segmented regression setting. We assume that the existence of a breakpoint has been established by one of the procedures in Chapter 4 and that the breakpoint has been estimated following

Chapter 3. The construction below is developed for a single breakpoint ($m = 1$) and the extension to $m \geq 2$ is discussed in Section 5.4.4.

5.4.1 Segmented Regression Model and GoF Testing

Let $\{(Y_i, x_i, z_i)\}_{i=1}^n$ be the data, where Y_i is the response, $x_i \in \mathbb{R}^p$ the vector of regressors with stable predictors, and $z_i \in \mathbb{R}^q$ the regime-specific regressors. The segmented regression model with one breakpoint ψ_1 specifies $E(Y_i \mid x_i, z_i) = f(x_i'\beta + z_i'\delta_j)$, with $i \in R_j(\psi_1)$ and $j = 1, 2$, where f is the inverse-link function. The full parameter vector is $\theta = (\beta', \delta_1', \delta_2', \psi_1)' \in \Theta \subset \mathbb{R}^{p+2q+1}$. After estimation, the raw ($\hat{\varepsilon}_i(\theta)$), working ($\hat{\varepsilon}_i^*(\theta)$) or Pearson ($\varepsilon_i^P(\theta)$) residuals can be computed. The following section is expressed in terms of raw residuals but the procedure is identical across all residual types. The null hypothesis is

$$H_0 : E(\varepsilon_i(\theta) \mid x_i, z_i) = 0 \quad \text{a.s.} \quad \text{for some } \theta \in \Theta,$$

which is precisely the conditional moment restriction from the null hypothesis of Section 5.1, so the PBT applies directly. A rejection of H_0 is evidence that the segmented specification is inadequate. Crucially, this is a different question from the one tackled in Chapter 4; there the question was whether the breakpoint existed, here, the aim is to evaluated the GoF of the final segmented model.

5.4.2 Projection Approach, Statistic and Computational Form

The null hypothesis can be expressed, using the full design matrix $X^* = (X \mid Z)$ as

$$E(\varepsilon(X^*, \theta) \mid X^*) = 0 \quad \text{a.s for some } \theta \in \Theta$$

The chain of equivalences established in Section 5.3.1 applies with the projection direction $W \in \mathbb{S}^{p+2q}$. The empirical analogue is the RMPP $R_n(W, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\varepsilon}_i \, 1(X_i^{*\prime} W \leq u)$ with $(W, u) \in \mathbb{S}^{p+2q} \times \mathbb{R}$.

Following Section 5.3.2, the projection-based statistic for the segmented model is

$$T_{\text{seg}} = \int_{-\infty}^{\infty} \int_{\mathbb{S}^{p+2q}} \Big(R_n(W, u) \Big)^2 \, dW \, F_{n,W}(du),$$

where $F_{n,W}$ is the empirical distribution function of the projected covariates $\{X_i^{*\prime} W\}_{i=1}^n$ and dW is the uniform measure on $\mathbb{S}^{p+2q}$. By the closed-form derivation of Escanciano (2006),

this reduces to the quadratic form

$$T_{\text{seg}} = \frac{1}{n^2}\hat{\varepsilon}'\, A\, \hat{\varepsilon}, \qquad \hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)',$$

where $A = (A_{is})_{n\times n}$ with $A_{is} = \sum_{l=1}^{n} A_{isl}$ and

$$A_{isl} = C_{p+2q}\left|\pi - \arccos\left(\frac{(X_i^* - X_l^*)'(X_s^* - X_l^*)}{\|X_i^* - X_l^*\|\,\|X_s^* - X_l^*\|}\right)\right|, \qquad C_{p+2q} = \frac{\pi^{\frac{p+2q}{2}-1}}{\Gamma(\frac{p+2q}{2}+1)},$$

with $A_{isl} = 0$ when $X_i^* = X_l^*$ or $X_s^* = X_l^*$.

The only quantities that change relative to Section 5.3.2 are the projection dimension, now $p + 2q$, and the residual vector $\hat{\varepsilon}$, which is now computed from the segmented model and depends on $\hat{\psi}_1$ through the regime assignment as well as on $\hat{\beta}$, $\hat{\delta}_1$, $\hat{\delta}_2$. As before, A depends only on X^* and on the dimension $p + 2q$. The procedure preserves every theoretical and practical advantage of the original PBT: it is omnibus, free of tuning parameters, computationally cheap, consistent, and detects alternatives at the parametric rate $n^{-1/2}$.

5.4.3 Bootstrap Procedures

As in Section 5.3.3, the asymptotic null distribution of T_{seg} is non-standard and critical values are obtained by bootstrap. For linear regression models, a data-level wild bootstrap is used, while in the GLM case, a parametric bootstrap is used. (See Appendix 7 for the pseudo-algorithm and bootstrap procedures).

5.4.4 Extension to Multiple Breakpoints $m \geq 2$

The extension to multiple breakpoints $m \geq 2$ is straightforward, and the construction is practically unchanged. The parameter vector is enlarged to $\theta = (\beta', \delta_1', \ldots, \delta_{m+1}', \psi_1, \ldots, \psi_m)' \in \mathbb{R}^{p+(m+1)q+m}$, the conditional mean $\hat{\mu}_i$ is computed from the corresponding regime $R_j(\psi)$ for $j = 1, \ldots, m+1$, and full design matrix $X^* = (X \mid Z)$. The matrix A, the projection dimension $p + (m+1)q$, the closed-form expression for T_{seg}, and the bootstrap procedures of Section 5.4.3 all carry over to this case. The residual vector $\hat{\varepsilon}$ changes, since it is now computed from the multi-breakpoint segmented fit.

Remark 5.3. Once the breakpoint has been located at $\hat{\psi}_1$, an alternative strategy is to refit independent GLM inside each segment and apply the PBT separately to each. This per-segment approach is not equivalent to the whole-model test, even with a multiplicity correction such as Bonferroni. The whole-model null H_0 imposes that the stable coefficients

β are common across regimes, while the per-segment nulls impose no such restriction, so $H_0 \subsetneq H_0^{(1)} \cap H_0^{(2)}$. Moreover, the whole-model statistic T_{seg} contains a cross-segment term $2\hat{\varepsilon}^{(1)\prime} A^{[12]} \hat{\varepsilon}^{(2)}$ that no per-segment procedure can recover. The two approaches are therefore complements: the whole-model PBT is the primary specification test, and the per-segment statistics serve as a localization device if the whole-model test rejects.

6 Simulation

This chapter evaluates the finite-sample performance of the detection and estimation procedures presented in Chapters 4 and 3 through a controlled Monte Carlo experiment. All settings use $n = 200$ observations and a nominal significance level $\alpha = 0.05$. Due to runtime constraints, the number of Monte Carlo replicates N_{sim} and bootstrap draws B differ between linear and GLM cases, with $N_{\text{sim}} = 1000 = B$ in the linear case and $N_{\text{sim}} = 500 = B$ for GLMs. A single covariate $z_i \overset{\text{iid}}{\sim} \text{Uniform}(0, 1)$ is considered, with both intercept and slope allowed to shift at ψ_1.

Three changepoint regression families are studied (Linear, Logistic and Poisson), under two break locations: break at the sample midpoint, with index $k^0 = 100$, and boundary break near the beginning of the trimmed set, $k^0 = 50$, since the first 30 observations are excluded from the admissible set ($\varepsilon = 0.15$). Three signal scenarios regulate the strength of the structural change: Scenario A (strong evidence for the existence of a break), Scenario B (moderate evidence), and Scenario C (no break). To ensure applicability of every studied test, the changepoint model is constructed to be continuous at the breakpoint ψ_1^0, and, in the linear case, data is generated with homocedasticity across regimes, $\sigma = 0.2$. In the logistic model, coefficients are chosen so that the two regimes remain essentially non-separable, while, in the Poisson model, they are chosen as to avoid degenerate (very low or high) rates.

In the linear scenarios, the same coefficients are used regardless of break location, so the effect of moving the break towards the boundary can be interpreted directly, with the underlying signal held fixed. This is not possible in the logistic and Poisson models, since in scenario B, coefficients had to be recalibrated separately for the middle and boundary designs to avoid quasi-separation and degenerate count rates. Consequently, any power or estimation performance difference observed between break locations in these two families reflects not only the change in location, but also this adjustment in signal strength, and should not be read as isolating a pure boundary effect. The parameters and regression functions for each regime, under scenarios A and B with middle (M) and boundary (B) break, are reported below, together with the scenario C (no-break) parameters. The regression function is displayed as $y(z)$ in the linear case, as the success probability $\pi(z)$

in the logistic case, and as the mean count $\mu(z)$ in the Poisson case, each shown as a function of the covariate z on either side of the changepoint ψ_1^0.

	α_0	δ_1	δ_2
A (M)	1	-7	7
B (M)	1	-0.5	0.5
A (B)	1	-7	7
B (B)	1	-0.5	0.5
C	1	-7	–

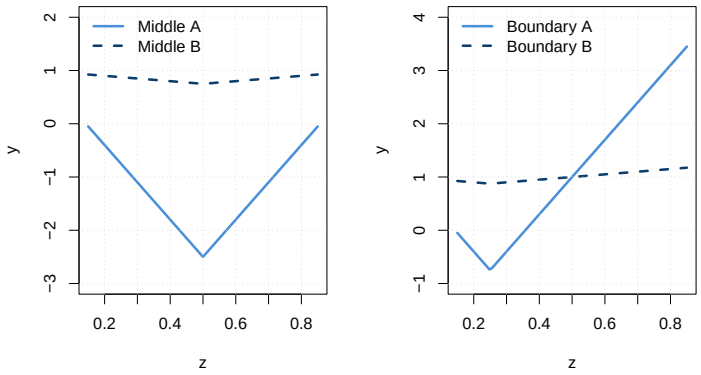

Figure 2: Linear Data Generating Processes.

	α_0	δ_1	δ_2
A (M)	1.5	-4	4
B (M)	1	-2.5	2.5
A (B)	2.5	-12	3
B (B)	1.5	-8	1.5
C	1.5	-3	–

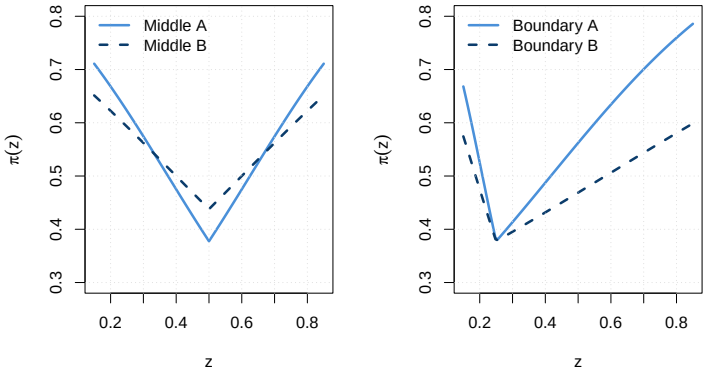

Figure 3: Logistic Data Generating Processes.

	α_0	δ_1	δ_2
A (M)	2.5	-1	1
B (M)	2.5	-0.45	0.45
A (B)	3	-2.5	0.8
B (B)	2.5	-0.6	0.7
C	2.5	-0.8	–

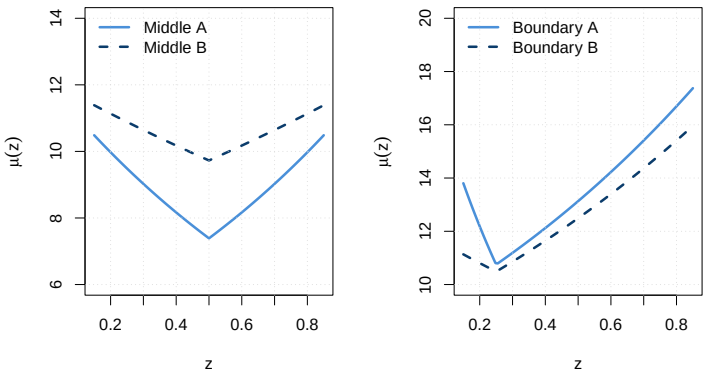

Figure 4: Poisson Data Generating Processes.

6.1 Test Simulation: Empirical Power and Empirical Size

6.1.1 Empirical Power

The null hypothesis, H_0, states that no breakpoint exists. To conduct the empirical power simulation, data is generated under H_1, where a break is present at the true breakpoint ψ_1^0, under Scenarios A and B. The empirical power is computed as the proportion of simulation replicates in which each test correctly rejects H_0, at the nominal significance level $\alpha = 0.05$. detecting an effective break.

Linear Middle Break, $k^0 = 100$

Scenario	LRT (cont.)	EL	Rec-CUSUM	Sc. CUSUM	Sc. MOSUM	Av. Sc.
A	1.000	1.000	1.000	1.000	1.000	1.000
B	0.998	0.212	0.710	0.991	0.954	0.997

Table 3: Empirical power, linear model, middle break $k^0 = 100$.

Linear Boundary Break, $k^0 = 50$

Scenario	LRT (cont.)	EL	Rec-CUSUM	Sc. CUSUM	Sc. MOSUM	Av. Sc.
A	1.000	1.000	1.000	1.000	1.000	1.000
B	0.777	0.195	0.569	0.785	0.593	0.724

Table 4: Empirical power, linear model, boundary break $k^0 = 50$.

In the linear setting, the LRT enforcing continuity, the EL test, the CUSUM-fluctuation tests and the average score test are considered. Under Scenario A all tests report a maximum empirical power at both break locations, as the slope change is sufficiently large for uniform detection, even when the break falls close to the sample boundary. In Scenario B, as expected the empirical power diminishes across all presented tests. At the middle break, it is highest for the LRT, while the score CUSUM and average score tests approach almost the same value. The score MOSUM reports a slightly lower value, which was expected given that the score CUSUM is most powerful for single breaks whereas the MOSUM foregoes global accumulation in favor of robustness to multiple breaks. The recursive residual

CUSUM reports the lowest empirical power among the CUSUM-fluctuation tests. The EL test, even though it behaved as expected in the strong evidence case, reports markedly low empirical power, reflecting its conservative nature, trading a potential power loss for a clear robustness advantage when the error distribution is uncertain. In the boundary break setting, there are only 20 pre-break observations in the admissible set, and, there is a visible and theoretically expected heterogeneous reduction in empirical power across all tests under Scenario B. The LRT, score CUSUM and average score decline moderately and still report the highest empirical powers, reflecting the reduced effective sample size in the shorter first regime rather than any structural disadvantage of the tests themselves. The recursive residual CUSUM and the score MOSUM report lower values, while the EL test remains low across both break locations, confirming that its power limitation and distributional robustness trade-off is not mitigated by moving the break towards the boundary.

Logistic Middle Break, $k^0 = 100$

Scenario	LRT (cont.)	EL	Sc. CUSUM	Sc. MOSUM	Sup. Sc.	Av. Sc.
A	0.942	0.760	0.916	0.806	0.882	0.946
B	0.636	0.408	0.606	0.472	0.518	0.680

Table 5: Empirical power, logistic model, middle break $k^0 = 100$.

Logistic Boundary Break, $k^0 = 50$

Scenario	LRT (cont.)	EL	Sc. CUSUM	Sc. MOSUM	Sup. Sc.	Av. Sc.
A	0.958	0.926	0.956	0.862	0.380	0.902
B	0.800	0.688	0.778	0.564	0.070	0.706

Table 6: Empirical power, logistic model, boundary break $k^0 = 50$.

In the logistic setting, the LRT enforcing continuity, the EL test, the score CUSUM and MOSUM tests, and the score tests are considered. Under Scenario A with an middle break all tests achieve high empirical power, though none reach the maximum. The EL test continues to present the lowest value, consistent with its conservatism, and the score

MOSUM remains the lowest among the CUSUM-fluctuation tests, for the same reasons explored previously. Under Scenario B with an middle break, power falls across all tests, and the ordering of procedures differs from the linear case. The average score test reports the highest power, followed closely by the LRT, however the difference between the two is small and the reversal relative to the linear setting may be attributable to the average score test recovering a standard χ^2 null distribution that requires no correction or bootstrap approach. The EL test reports higher power than in the linear moderate-signal case, while all other tests report lower power, this might suggest an advantage in the logistic setting, perhaps because the working residuals carry a stronger signal about the break, or it may simply reflect differences in signal parametrization between the two DGP rather than a structural property of the test across families.

At the boundary, under scenario A, all tests report high empirical power with the exception of the supremum score test. Under scenario B, power drops across all tests when compared to scenario A, as expected, and the relative disadvantage of the supremum score test persists. This is likely attributable to the conservatism of the Davies analytical bound used to compute its p-value, indicating that as the break approximates to the trimming boundary, the conservative nature of the bound becomes more apparent. This is further supported by the empirical size results explored ahead. Compared to the middle break, power under scenario B is higher at the boundary for every test except the supremum score test, which appears to contradict the expectation that detection is harder near the boundary, however, as discussed above, this comparison is not valid, since the DGP parameters were adjusted relative to the middle break case and may have encoded a stronger signal than those used at the middle, which is sufficient to explain the higher power without any genuine location effect.

Poisson Middle Break, $k^0 = 100$

Scenario	LRT (cont.)	EL	Sc. CUSUM	Sc. MOSUM	Sup. Sc.	Av. Sc.
A	1.000	0.994	1.000	0.992	1.000	1.000
B	0.782	0.622	0.708	0.556	0.674	0.816

Table 7: Empirical power, Poisson model, middle break $k^0 = 100$.

Poisson Boundary Break, $k^0 = 50$

Scenario	LRT (cont.)	EL	Sc. CUSUM	Sc. MOSUM	Sup. Sc.	Av. Sc.
A	1.000	1.000	1.000	1.000	1.000	1.000
B	0.738	0.768	0.714	0.462	0.576	0.630

Table 8: Empirical power, Poisson model, boundary break $k^0 = 50$.

Under scenario A, all tests achieve high empirical power at both break locations (1.000 across all tests in the boundary break case) indicating that the strong signal is sufficient for uniform detection regardless of break location. Under scenario B, a drop in empirical power is observed, consistent with all previous patterns. At the middle break, the average score test reports the highest power, followed closely by the LRT and the score CUSUM tests. The supremum score test and the EL test report lower values, possibly due to the Davies analytical bound and the conservative nature of the test, respectively. The score MOSUM test reports the lowest value, reflecting the power trade-off associated with its local window evaluation.

At the boundary break under scenario B, the ordering changes, notably, the EL test reports the highest empirical power, surpassing even the LRT, which is unexpected given its conservative behavior across all previous settings, this may reflect a particular feature of the Poisson working residuals at this break location, or it may be a simulation artefact. The score MOSUM and supremum score test again report the lowest values, consistent with their observed behavior in the middle break case.

6.1.2 Empirical Size

For the empirical size simulation, data is generated under H_0, that is, under Scenario C where there is no breakpoint. The empirical size is computed as the proportion of simulation replicates in which each test falsely rejects H_0, at the nominal significance level $\alpha = 0.05$. A well-calibrated test should report empirical size close to the nominal level, with values substantially above 0.05 indicating size distortion, and values substantially below indicating conservatism.

Model	LRT (cont.)	EL	Rec-CUSUM	Sc. CUSUM	Sc. MOSUM	Sup. Sc.	Av. Sc.
Linear	0.043	0.045	0.045	0.045	0.062	0.036	0.043
Logistic	0.044	0.056	–	0.056	0.104	0.002	0.046
Poisson	0.054	0.044	–	0.054	0.074	0.002	0.060

Table 9: Empirical size under Scenario C, nominal level $\alpha = 0.05$.

In the linear setting, all tests control the size well, with empirical rejection frequencies close to the nominal level. The score MOSUM is marginally oversized and the supremum score test is slightly conservative, but all tests report values close to 0.05. In the logistic setting, the LRT, the EL, the Score CUSUM and the average score tests are well calibrated. In the Poisson setting, the LRT, the EL, and the score CUSUM tests are well calibrated. However, in the both settings the score MOSUM presents a higher empirical size, suggesting anti-conservatism. This implies that its power figures in Scenarios A and B may be somewhat inflated relative to a perfectly calibrated test, as part of the observed rejections under the alternative may reflect excess rejection under the null rather than genuine detection of the break. The same is observed for the average score test in the Poisson setting, where this test is slightly anti-conservative. In both the logistic and Poisson settings, the supremum score test reports a near-zero empirical size, indicating severe conservatism, meaning that the test almost never rejects under H_0 in these settings. This suggests that the Davies analytical bound used to compute the p-value is excessively conservative in the logistic and Poisson cases, and as a consequence the power figures reported for this test should be interpreted with caution, which is consistent with its extremely low power results in the logistic boundary break settings.

6.2 Simulation Study for the Estimation Process

Here, the finite-sample performance of the two breakpoint estimators (from the GLS and the Iterative Linearization procedures explored in Chapter 3) is evaluated, across scenarios and break locations. As discussed previously, the Iterative Linearization method requires a starting value, which in practice could be obtained from one of the detection tests. However, initializing the algorithm from a test-based estimate would conflate the imprecision of the detection test in locating the break, and the behavior of the estimation algorithm given that starting point. Under such an initialization, any observed bias or variability in results could not be cleanly attributed to the estimation procedure itself. To avoid

this, all replicates are initialized at the true breakpoint ψ_1^0, which is fixed and known in the simulation context. This ensures that the reported Bias, SD, and RMSE reflect solely the properties of the estimation algorithm - the quality of the linearization approximation, the convergence behavior, and the sensitivity to signal strength - making this a clean evaluation of the estimator rather than a joint evaluation of detection and estimation. As a consequence of this initialization strategy, the mean number of iterations to convergence is expected to be small across all settings.

For each simulation replicate, the estimated breakpoint $\hat{k}$ is recorded, and the following summary statistics are computed across the N_{sim} replicates: the mean estimate $\bar{\hat{k}}$; the bias, defined as $\bar{\hat{k}} - k^0$; the Standard Deviation (SD) of $\hat{k}$ across replicates; the Root Mean Squared Error (RMSE), $\sqrt{\text{Bias}^2 + \text{SD}^2}$; the empirical coverage of the 95% confidence interval for k^0; and, for the iterative linearization estimator, the mean number of iterations to convergence. Scenario C is omitted from this simulation study, as under the null hypothesis of no break the breakpoint is not identified.

Linear Middle Break, $k^0 = 100$

Scenario	k^0	Mean $\hat{k}$	Bias	SD	RMSE	Coverage	Mean Iter.
A - GLS	100	99.708	-0.292	4.169	4.177	0.969	–
A - IL	100	100.022	0.022	1.391	1.391	0.981	2.080
B - GLS	100	99.294	-0.706	30.698	30.690	0.961	–
B - IL	100	100.050	0.050	12.341	12.335	0.916	2.320

Table 10: Breakpoint estimation (GLS and Iterative Linearization), linear model, middle break $k^0 = 100$.

Linear Boundary Break, $k^0 = 50$

Scenario	k^0	Mean $\hat{k}$	Bias	SD	RMSE	Coverage	Mean Iter.
A - GLS	50	49.637	-0.363	4.095	4.107	0.969	–
A - IL	50	49.962	-0.038	1.531	1.530	0.970	2.160
B - GLS	50	65.759	15.759	29.943	33.782	0.905	–
B - IL	50	51.050	1.050	14.759	14.781	0.906	2.290

Table 11: Breakpoint estimation (GLS and Iterative Linearization), linear model, boundary break $k^0 = 50$.

In the linear setting, both the GLS and the iterative linearization estimators can be compared directly, since both procedures apply. Under scenario A at the middle break, all mean estimates are extremely close to the true breakpoint value, as expected given the strong evidence of a break. Across both scenarios and estimation approaches, bias is negligible relative to SD, so variability rather than bias drives the estimation error. The iterative linearization estimator reports lower absolute bias, SD and RMSE than the GLS procedure across both scenarios, suggesting better finite-sample precision, and converges rapidly in all cases. Under scenario B, the SD increases considerably for both estimators, relative to scenario A, reflecting the weaker evidence of a break. Regarding coverage, the GLS procedure achieves coverage close to the nominal level, for both scenarios, while the iterative linearization procedure reports higher coverage under scenario A, suggesting a slightly wide confidence interval, and lower coverage under scenario B, indicating some undercoverage.

At the boundary break under scenario A, the mean estimated values remain close to the true breakpoint for both estimators, bias is still negligible, and variability still drives the estimation error. Under scenario B, the iterative linearization estimator maintains relatively small bias, with the mean estimated value remaining close to the true breakpoint, though the bias is slightly higher than in the middle break case, as expected. The most notable result is the GLS estimator under scenario B, which exhibits a large positive bias and a mean estimate furthest from the true breakpoint of all settings considered, indicating a systematic tendency to overestimate the break location in this case. The iterative linearization estimator exhibits a much smaller bias, with also smaller SD, and RMSE than the GLS estimator, suggesting greater robustness to the boundary break under moderate

signal. The GLS coverage drops below the nominal level, which may reflect that the large bias displaces the interval away from the true value despite the wide spread. The iterative linearization coverage also drops below 0.05, suggesting undercoverage.

Logistic Middle Break, $k^0 = 100$

Scenario	k^0	Mean $\hat{k}$	Bias	SD	RMSE	Coverage	Mean Iter.	Conv. Rate
A	100	100.304	0.304	15.716	15.704	0.892	1.300	1.000
B	100	99.986	-0.014	23.127	23.104	0.874	1.340	1.000

Table 12: Breakpoint estimation (IL), logistic model, middle break $k^0 = 100$.

Logistic Boundary Break, $k^0 = 50$

Scenario	k^0	Mean $\hat{k}$	Bias	SD	RMSE	Coverage	Mean Iter.	Conv. Rate
A	50	50.820	0.820	11.762	11.779	0.884	1.370	1.000
B	50	50.848	0.848	14.965	14.974	0.918	1.280	1.000

Table 13: Breakpoint estimation (IL), logistic model, boundary break $k^0 = 50$.

In the logistic setting, only the iterative linearization estimator is considered, and convergence is complete across all replicates in both scenarios and break locations. Across all settings bias is negligible relative to SD, confirming that variability rather than bias drives the estimation error. Coverage is below the nominal level in all settings, suggesting some undercoverage, in scenario B with a boundary break, it is closer to 0.95, likely due to a larger SD producing a wider confidence interval.

A counterintuitive pattern is observed in the middle break scenarios, where bias is smaller under scenario B than under scenario A, despite the weaker signal, however, this may be a simulation artefact or reflect particular features of this DGP. Overall, the iterative linearization estimator displays well-behaved estimation in the logistic setting, with negligible bias and rapid convergence across both break locations and scenarios. However it is important to note that the logistic and linear results are not directly comparable, since the signal parametrization differs across families, in particular, the scenario B coefficients in the logistic DGP may represent a stronger effective signal than their linear counterparts, which could partially explain the relatively well-behaved estimation results observed here.

Poisson Middle Break, $k^0 = 100$

Scenario	k^0	Mean $\hat{k}$	Bias	SD	RMSE	Coverage	Mean Iter.	Conv. Rate
A	100	98.846	-1.154	10.083	10.139	0.932	1.280	1.000
B	100	98.456	-1.544	21.224	21.259	0.900	1.400	1.000

Table 14: Breakpoint estimation (IL), Poisson model, middle break $k^0 = 100$.

Poisson Boundary Break, $k^0 = 50$

Scenario	k^0	Mean $\hat{k}$	Bias	SD	RMSE	Coverage	Mean Iter.	Conv. Rate
A	50	50.592	0.592	5.937	5.961	0.956	1.200	1.000
B	50	51.570	1.570	18.694	18.741	0.890	1.390	1.000

Table 15: Breakpoint estimation (IL), Poisson model, boundary break $k^0 = 50$.

In the Poisson setting, only the iterative linearization estimator is again considered, and convergence is complete across all replicates in both scenarios and break locations. Across all settings, bias is negligible relative to SD, confirming that variability rather than bias drives the estimation error, and mean estimated values remain close to the true breakpoint throughout. Under both break locations, the bias, SD and RMSE increase from scenario A to scenario B, as expected given the weaker signal. Coverage is below the nominal level in most settings, suggesting some undercoverage, with the exception of scenario A at the boundary break, which reports coverage very close to the nominal level. As with the logistic setting, the Poisson and linear results are not directly comparable due to differences in signal parametrization across families.

7 Conclusions

This thesis set out to provide a unified treatment of detection, estimation and goodness-of-fit testing for structural change across the linear and GLM regression settings, and to compare the resulting procedures on equal footing rather than in isolation.

Chapter 2 laid the common ground by formalizing the partial change model in both settings and deriving the likelihood, score, and information objects needed throughout the thesis. Chapter 3 showed that, while breakpoint estimation is a solved problem in the linear model through Global Least Squares, the GLM setting requires a genuine methodological choice: an exhaustive Global Maximum Likelihood search is theoretically available but computationally prohibitive, motivating the adoption of the Iterative Linearization estimation process under a continuity restriction. Chapter 4 assembled five families of tests for the existence of a break, extended the Empirical Likelihood ratio test of Liu and Qian (2009) to the GLM setting, and closed with a systematic comparison along a regression setting, sensitivity to break type, distributional assumptions, computational cost, and calibration strategy. Chapter 5 extended Goodness-of-Fit (GoF) testing to the changepoint regression model via an adaptation of Escanciano (2006) projection-based approach, providing a tool to assess the adequacy of a fitted changepoint model once the existence of a break has already been established.

The simulation study of Chapter 6 evaluated the finite-sample performance of the detection and estimation procedures, across three scenarios with different signal for the strength of the structural change. Regarding the tests for the existence of a breakpoint a comparison of empirical power and empirical size, substantiated several of the theoretical distinctions drawn in Chapter 4. On the estimation side, the Iterative Linearization procedure and the GLS were compared in the linear setting, while the Iterative Linearization procedure was applied for the logistic and Poisson changepoint regression, across the two scenarios.

Taken together, these results show that no single test or estimator dominates uniformly across the linear and GLM settings: the appropriate choice depends on the regression family, the anticipated number of breaks, break type and location, the plausibility of the underlying distributional assumptions, and the computational budget available. The Empirical Likelihood test traded power for robustness to error misspecification, providing a

non-parametric alternative the classicallrt. The CUSUM-fluctuation tests separated along their sensitivity to a single persistent break versus multiple or boundary breaks, which, together with the score-based tests, provide a computationally attractive alternative. The Escanciano PBT, operating as a goodness-of-fit diagnostic rather than a primary test, offered a complementary check on model adequacy once a break had been identified. The comparative framework and simulation evidence developed across Chapters 2–6 are intended to make that choice an informed one, grounded in how each procedure actually behaves under the conditions a practitioner is likely to face, rather than a default inherited from the linear case. The theoretical gaps and open questions that surfaced along the way, from the untested behavior of several test families under multiple breaks to the unresolved asymptotics of the EL test in GLMs, are taken up in the following Section.

7.1 Future Work

Several extensions follow naturally from the results and limitations identified throughout this thesis.

- **Extension to two or more breakpoints.** The sequential procedure of Section4.8 is theoretically justified for the sup-F test (Bai, 1997, Bai and Perron, 1998), but for the remaining four families its validity was only argued heuristically. A natural next step is to formally establish the sequential framework's properties for the classical LRT, the EL test, the CUSUM-fluctuation tests, and the score-based tests, and to evaluate all five families through simulation under $m \geq 2$ true breaks.

- **Simulation with two break scenarios.** Based on the discussion of Section 4.6.4, the Score MOSUM is expected to outperform the remaining CUSUM-fluctuation tests in this setting, since its local-window construction should be less exposed to the cancellation of opposite-signed cumulative drifts than the Recursive Residual and Score CUSUM.

- **Theoretical development of the EL test in GLMs.** The EL test is a genuine extension of Liu and Qian (2009), who did not consider the GLM setting. Its asymptotic properties in the GLM case, in particular the validity of the Gumbel approximation for $Z_n = \sqrt{M_n}$, were not formally established and warrant dedicated theoretical treatment, alongside a formal consistency proof under the alternative.

- **Extension of the EL test to discontinuous alternatives.** The EL test is currently restricted to the continuous alternative, and is inconsistent against jump alternatives,

unlike every other family developed in Chapter 3. Relaxing this restriction would require reformulating the moment conditions and weight spaces to accommodate a discontinuity in the conditional mean at the breakpoint, and represents an open and non-trivial extension of the empirical likelihood construction itself.

- **Estimation in the discontinuous GLM model.** The Iterative Linearization procedure is only valid under the continuity restriction. In the discontinuous case, the natural alternative is the Global Maximum Likelihood (GML) analogue of the GLS, but this requires a full IRLS fit for every admissible partition and quickly becomes computationally prohibitive as n grows or m is unknown. A worthwhile direction is to investigate how to reduce this exhaustive search.

- **Sensitivity to sample size.** All simulations in Chapter 6 were conducted at a single, fixed sample size. Since Section 4.7.2.3 establishes that the classical LRT and the EL test share the same large-sample power under correct specification, with the EL test's power deficit attributable to finite-sample conservatism, an explicit sample-size sweep would be informative. In this case, the two tests are expected to converge in power as n increases, and all tests should gain power as the effective regime sizes grow. This would also clarify whether the Poisson size distortions observed in Poisson boundary break case are a finite-sample artifact that shrinks with n, or a more persistent feature of the tests in that family.

- **Extending the simulation study within each model class.** A natural extension of the simulation study in Chapter 6 is to broaden it to additional DGPs within the same class. This is particularly relevant for the logistic and Poisson settings, where it is important to understand how the tests behave under conditions which are known to be problematic for each family, such as, (quasi-)separation in the logistic case, low or high mean counts in the Poisson setting. Such an extension would test the robustness of the detection procedures under conditions that are common in applied logistic and Poisson regression.

- **Application to real data.** All results in this thesis are simulation-based. Applying the full pipeline to real datasets from the application areas explored in Chapter 1 (genomics, financial returns, or clinical time series) would test the practical relevance of the comparative guidance under realistic, possibly misspecified, conditions.

References

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, pages 821–856. [Cited on pages 2, 25, 26, and 27.]

Andrews, D. W. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3):683–734. [Cited on page 2.]

Andrews, D. W., Lee, I., and Ploberger, W. (1996). Optimal changepoint tests for normal linear regression. *Journal of Econometrics*, 70(1):9–38. [Cited on page 2.]

Andrews, D. W. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, pages 1383–1414. [Cited on pages 2, 26, and 28.]

Aston, J. A. and Kirch, C. (2012). Evaluating stationarity via change-point alternatives with applications to fmri data. [Cited on page 1.]

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, 79(4):551–563. [Cited on pages 2, 60, and 83.]

Bai, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics*, 157(1):78–92. [Cited on page 1.]

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78. [Cited on pages 2, 17, 19, 22, 23, 26, 27, 28, 59, 60, and 83.]

Bai, J. and Perron, P. (2003a). Computation and analysis of multiple structural change models. *Journal of applied econometrics*, 18(1):1–22. [Cited on pages 2 and 26.]

Bai, J. and Perron, P. (2003b). Critical values for multiple structural change tests. *The Econometrics Journal*, 6(1):72–78. [Cited on pages 27 and 28.]

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, 20(1):105–134. [Cited on page 65.]

Bierens, H. J. and Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica: Journal of the Econometric Society*, pages 1129–1151. [Cited on page 65.]

Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 37(2):149–163. [Cited on pages 2, 42, and 44.]

Cho, H. and Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 77(2):475–507. [Cited on page 1.]

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, pages 591–605. [Cited on pages 2 and 26.]

Chu, C.-S. J., Hornik, K., and Kaun, C.-M. (1995). Mosum tests for parameter constancy. *Biometrika*, 82(3):603–617. [Cited on pages 47 and 48.]

Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, 146(1):162–169. [Cited on page 32.]

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64(2):247–254. [Cited on pages 50 and 53.]

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74(1):33–43. [Cited on pages 50 and 53.]

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26. [Cited on page 34.]

Escanciano, J. C. (2006). A consistent diagnostic test for regression models usingprojections. *Econometric Theory*, 22(6):1030–1051. [Cited on pages 64, 65, 66, 67, 68, and 82.]

González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *Test*, 22(3):361–411. [Cited on pages 32 and 34.]

Hawkins, D. M. (1977). Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, 72(357):180–186. [Cited on page 2.]

Hawkins, D. M. (2001). Fitting multiple change-point models to data. *Computational Statistics & Data Analysis*, 37(3):323–341. [Cited on page 3.]

Hjort, N. L. and Koning, A. (2002). Tests for constancy of model parameters over time. *Journal of Nonparametric Statistics*, 14(1-2):113–132. [Cited on pages 44 and 46.]

Jones, L. K. (1987). On a conjecture of huber concerning the convergence of projection pursuit regression. *The annals of statistics*, pages 880–882. [Cited on page 66.]

Kejriwal, M. and Perron, P. (2008). The limit distribution of the estimates in cointegrated regression models with multiple structural changes. *Journal of Econometrics*, 146(1):59–73. [Cited on page 3.]

Kim, H.-J. and Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression. *Biometrika*, 76(3):409–423. [Cited on page 2.]

Lavielle, M. (2005). Using penalized contrasts for the change-point problem. *Signal processing*, 85(8):1501–1510. [Cited on page 3.]

Leonardi, F. and Bühlmann, P. (2016). Computationally efficient change point detection for high-dimensional regression. *arXiv preprint arXiv:1601.03704*. [Cited on page 3.]

Liu, B., Zhou, C., Zhang, X., and Liu, Y. (2020). A unified data-adaptive framework for high dimensional change point detection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(4):933–963. [Cited on page 1.]

Liu, H., Li, X., Chen, F., Härdle, W., and Liang, H. (2024). A comprehensive comparison of goodness-of-fit tests for logistic regression models. *Statistics and Computing*, 34(5):175. [Cited on pages 64 and 67.]

Liu, R. Y. (1988). Bootstrap procedures under some non-iid models. *The annals of statistics*, pages 1696–1708. [Cited on page 32.]

Liu, Z. and Qian, L. (2009). Changepoint estimation in a segmented linear regression via empirical likelihood. *Communications in Statistics-Simulation and Computation*, 39(1):85–100. [Cited on pages 34, 37, 38, 40, 61, 82, and 83.]

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The annals of statistics*, 21(1):255–285. [Cited on page 32.]

Muggeo, V. M. (2003). Estimating regression models with unknown break-points. *Statistics in medicine*, 22(19):3055–3071. [Cited on pages 4, 17, 20, 21, 22, and 59.]

Muggeo, V. M. (2016). Testing with a nuisance parameter present only under the alternative: a score-based approach with application to segmented modelling. *Journal of Statistical Computation and Simulation*, 86(15):3059–3067. [Cited on pages 50, 55, 56, and 59.]

Muggeo, V. M. et al. (2008). Segmented: an r package to fit regression models with broken-line relationships. *R news*, 8(1):20–25. [Cited on pages 54, 58, and 59.]

Nelder, J. A. and Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 135(3):370–384. [Cited on page 10.]

Nguyen, L., Yamamoto, Y., and Perron, P. (2024). *mbreaks: Estimation and Inference for Structural Breaks in Linear Regression Models*. R package version 1.0.1. [Cited on page 60.]

Owen, A. (1991). Empirical likelihood for linear models. *The Annals of Statistics*, pages 1725–1747. [Cited on pages 34 and 35.]

Perron, P. and Zhu, X. (2005). Structural breaks with deterministic and stochastic trends. *Journal of Econometrics*, 129(1-2):65–119. [Cited on page 3.]

Qu, Z. and Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75(2):459–502. [Cited on page 3.]

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the american statistical association*, 53(284):873–880. [Cited on pages 2 and 29.]

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American statistical Association*, 55(290):324–330. [Cited on pages 2 and 29.]

Rice, S. O. (1944). Mathematical analysis of random noise. *The Bell System Technical Journal*, 23(3):282–332. [Cited on page 53.]

Spokoiny, V. (2009). Multiscale local change point detection with applications to value-at-risk. [Cited on page 1.]

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, pages 613–641. [Cited on page 65.]

Stute, W., Manteiga, W. G., and Quindimil, M. P. (1998). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, 93(441):141–149. [Cited on page 32.]

Wang, Y., Wu, C., Ji, Z., Wang, B., and Liang, Y. (2011). Non-parametric change-point method for differential gene expression detection. *PloS one*, 6(5):e20060. [Cited on page 1.]

Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *the Annals of Statistics*, 14(4):1261–1295. [Cited on page 32.]

Zeileis, A., . H. K. (2007). Generalized m□fluctuation tests for parameter instability. *Statistica Neerlandica*, 61(4). [Cited on page 3.]

Zeileis, A. (2005). A unified approach to structural change tests based on ml scores, f statistics, and ols residuals. *Econometric Reviews*, 24(4):445–466. [Cited on pages 44 and 47.]

Zeileis, A. (2006). Implementing a class of structural change tests: An econometric computing approach. *Computational Statistics & Data Analysis*, 50(11):2987–3008. [Cited on pages 46, 48, 102, and 103.]

Zeileis, A., Kleiber, C., Krämer, W., and Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational statistics & data analysis*, 44(1-2):109–123. [Cited on page 3.]

Zeileis, A., Leisch, F., Hornik, K., and Kleiber, C. (2002). strucchange: An r package for testing for structural change in linear regression models. *Journal of statistical software*, 7:1–38. [Cited on page 19.]

Zeileis, A., Shah, A., and Patnaik, I. (2010). Testing, monitoring, and dating structural changes in exchange rate regimes. *Computational Statistics & Data Analysis*, 54(6):1696–1706. [Cited on page 3.]

Zhou, H., Zhu, H., and Wang, X. (2025). Change-point detection for multivariate nonparametric regression with deep neural networks. *Computational Statistics & Data Analysis*, page 108334. [Cited on page 1.]

Zhou, M. (2026). *emplik: Empirical Likelihood Ratio for Censored/Truncated Data*. R package version 1.3-3. [Cited on page 98.]

A. Appendix

A.1. Derivations

A.1.1. Chapter 3 - Estimation

A.1.1.1. IRLS Estimation with Known Breaks

When the breakpoints $\psi_1, \ldots, \psi_m$ are known and fixed, the GLM partial change model reduces to a standard GLM with design matrix $X^* = (X \mid Z)$ fully determined by the partition. The only unknown parameters are the regression coefficients $(\beta, \delta_1, \ldots, \delta_{m+1})$ and the dispersion parameter ϕ, which can estimated by solving the score equations

$$s_\beta(\beta, \delta, \phi \mid \psi) = 0 \qquad \text{and} \qquad s_{\delta_j}(\beta, \delta_j, \phi \mid \psi) = 0, \quad j = 1, \ldots, m+1,$$

derived in Section 2.2.3.2. Since these equations are nonlinear in (β, δ_j), they do not admit a closed-form solution in general.

Two equivalent iterative methods are available: the Fisher scoring algorithm and the IRLS algorithm. The two methods are equivalent, in the sense that they produce identical updates at each iteration. We explore the IRLS algorithm to estimate the parameters β, δ_j for $j = 1, \ldots, m+1$.

Estimation of β The Fisher scoring update at iteration k is

$$\beta^{(k+1)} = \beta^{(k)} + \mathcal{I}(\beta^{(k)})^{-1} s_\beta(\beta^{(k)}, \delta^{(k)}, \phi \mid \psi) = \beta^{(k)} + (X' W^{(k)} X)^{-1} X' W^{(k)} G^{(k)} (Y - \hat{\mu}^{(k)})$$

Writing $\beta^{(k)} = (X' W^{(k)} X)^{-1} (X' W^{(k)} X) \beta^{(k)}$ and collecting,

$$\beta^{(k+1)} = (X' W^{(k)} X)^{-1} X' W^{(k)} \left[X\beta^{(k)} + G^{(k)} (Y - \hat{\mu}^{(k)}) \right] = (X' W^{(k)} X)^{-1} X' W^{(k)} Y^{*(k)},$$

a weighted least squares regression of the working response $Y^{*(k)}$ on X with weights $W^{(k)}$.

Estimation of δ_j Analogously, since s_{δ_j} and $\mathcal{I}_{\delta_j \delta_j}$ involve only observations in $R_j(\psi)$,

$$\delta_j^{(k+1)} = \delta_j^{(k)} + \mathcal{I}(\delta_j^{(k)})^{-1} s_{\delta_j}(\beta^{(k)}, \delta_j^{(k)}, \phi \mid \psi) = (Z_j' W_j^{(k)} Z_j)^{-1} Z_j' W_j^{(k)} Y_j^{*(k)},$$

a weighted least squares regression of the working response $Y_j^{*(k)}$ on Z_j with weights $W_j^{(k)}$, restricted to regime $R_j(\psi)$.

Joint update Since β enters in every regime, while δ_j only enters in regime j, the joint IRLS update across all parameters can be written compactly using the full design matrix $X^* = (X \mid Z)$ as

$$\begin{pmatrix} \beta^{(k+1)} \\ \delta^{(k+1)} \end{pmatrix} = (X^{*'} W^{(k)} X^*)^{-1} X^{*'} W^{(k)} Y^{*(k)},$$

where the block-diagonal structure of Z ensures that the δ_j updates remain regime-separated. This is precisely the weighted least squares analogue of the OLS closed-form solution, with the weight matrix $W^{(k)}$ and working response $Y^{*(k)}$ replacing the identity weights and observed response of the linear setting. The algorithm iterates until convergence, yielding the ML estimators $(\hat{\beta}(\psi), \hat{\delta}_1(\psi), \ldots, \hat{\delta}_{m+1}(\psi))$ for the fixed partition ψ.

The weighted hat matrix H_w introduced in Section 2.2.3.2 arises naturally. At convergence the IRLS update is a weighted least least squares regression of the working response Y^* on the design matrix X^* with weight matrix W. Therefore, at convergence the hat matrix is a projection matrix of this weighted least squares step:

$$H_w = W^{1/2} X^* \left(X^{*\prime} W X^* \right)^{-1} X^{*\prime} W^{1/2} = X_w^* \mathcal{I}^{-1} X_w^{*\prime}$$

with weighted working residuals $R_w^* = (I_n - H_w) Y_w^*$.

The full algorithm proceeds as follows.

1. Input: Data $\{(y_i, x_i, z_i)\}_{i=1}^n$, known partition $(\psi_1, \ldots, \psi_m)$, link function $g(\cdot)$, variance function $V(\cdot)$, and convergence tolerance $\tau > 0$.
2. Construct the full design matrix $X^* = (X \mid Z)$, where Z is the $n \times (m+1)q$ block-diagonal matrix determined by the known partition $(\psi_1, \ldots, \psi_m)$.
3. Initialize $\beta^{(0)}$ and $\delta_j^{(0)}$ for $j = 1, \ldots, m+1$.
4. For $k = 0, 1, 2, \ldots$:
   (a) For each $i \in R_j(\psi)$, compute the fitted linear predictor and fitted mean:

$$\hat{\eta}_i^{(k)} = x_i' \beta^{(k)} + z_i' \delta_j^{(k)}, \qquad \hat{\mu}_i^{(k)} = g^{-1}\left(\hat{\eta}_i^{(k)} \right)$$

   (b) Compute the diagonal weight matrix $W^{(k)} = \mathrm{diag}(w_1^{(k)}, \ldots, w_n^{(k)})$ with weights

$$w_i^{(k)} = \frac{1}{a(\phi)\, V\left(\hat{\mu}_i^{(k)} \right) \left[g'\left(\hat{\mu}_i^{(k)} \right) \right]^2}$$

(c) Compute the working residuals and working response:

$$r_i^{*(k)} = \left(y_i - \hat{\mu}_i^{(k)}\right) g'\left(\hat{\mu}_i^{(k)}\right), \qquad y_i^{*(k)} = \hat{\eta}_i^{(k)} + r_i^{*(k)},$$

or in matrix form, $R^{*(k)} = G^{(k)}(Y - \hat{\mu}^{(k)})$ and $Y^{*(k)} = \hat{\eta}^{(k)} + R^{*(k)}$.

(d) Update all parameters jointly

$$\begin{pmatrix} \beta^{(k+1)} \\ \delta^{(k+1)} \end{pmatrix} = \left(X^{*\prime} W^{(k)} X^*\right)^{-1} X^{*\prime} W^{(k)} Y^{*(k)}$$

which decomposes into the regime-separated updates

$$\beta^{(k+1)} = (X' W^{(k)} X)^{-1} X' W^{(k)} Y^{*(k)}, \qquad \delta_j^{(k+1)} = (Z_j' W_j^{(k)} Z_j)^{-1} Z_j' W_j^{(k)} Y_j^{*(k)}$$

(e) Stop if $\|\beta^{(k+1)} - \beta^{(k)}\| + \sum_{j=1}^{m+1} \|\delta_j^{(k+1)} - \delta_j^{(k)}\| < \tau$; otherwise return to step (a).

5. Set $\hat{\beta}(\psi) = \beta^{(k+1)}$ and $\hat{\delta}_j(\psi) = \delta_j^{(k+1)}$. Estimate the dispersion parameter by the Pearson estimator:

$$\hat{\phi} = \frac{1}{n - p - (m+1)q} \sum_{j=1}^{m+1} \sum_{i \in R_j(\psi)} \frac{\left(y_i - \hat{\mu}_i\right)^2}{V(\hat{\mu}_i)}$$

A.1.1.2. Pseudo-algorithm - Global Least Squares

The GLS computational procedure for $m = 1$ and $X = 0$ is outlined as an example.

1. For a given candidate break ψ_1, define a function that estimates the regime-specific coefficients by OLS and computes the corresponding sum of squared residuals.

   With $X = 0$, the estimators are

$$\hat{\delta}_1(\psi_1) = \arg\min_{\delta_1} \sum_{i \in R_1(\psi_1)} (y_i - \delta_1 z_i)^2, \qquad \hat{\delta}_2(\psi_1) = \arg\min_{\delta_2} \sum_{i \in R_2(\psi_1)} (y_i - \delta_2 z_i)^2,$$

   and the concentrated objective function is

$$S(\psi_1) = \sum_{i \in R_1(\psi_1)} \left(y_i - \hat{\delta}_1(\psi_1)\, z_i\right)^2 + \sum_{i \in R_2(\psi_1)} \left(y_i - \hat{\delta}_2(\psi_1)\, z_i\right)^2$$

2. Construct the admissible candidate breakpoint set by imposing a trimming condition. With $h = \lfloor \varepsilon n \rfloor$, require $h \leq \psi_1 \leq n - h$. For example, if $\varepsilon = 0.20$ and $n = 20$, then $h = 4$ and $\Psi_{\varepsilon,1} = \{4, 5, \ldots, 16\}$.

3. Evaluate $S(\psi_1)$ for every $\psi_1 \in \Psi_{\varepsilon,1}$. Apply the function from step 1 to every admissible point on $\Psi_{\varepsilon,1}$ and store the results.

4. Set $\hat{\psi}_1 = \arg\min_{\psi_1 \in \Psi_{\varepsilon,1}} S(\psi_1)$ and retain the corresponding coefficient estimates $\hat{\delta}_1(\hat{\psi}_1)$ and $\hat{\delta}_2(\hat{\psi}_1)$.

A.1.1.3. Pseudo-algorithm - Iterative Linearization

1. Input: Data $\{(y_i, x_i, z_i)\}_{i=1}^n$, link function $g(\cdot)$, variance function $V(\cdot)$, number of breaks m, initial breakpoint estimates $\psi_j^{(0)}$ for $j = 1, \ldots, m$, and convergence tolerance $\tau > 0$.

2. For $k = 0, 1, \ldots$:

   (a) For $j = 1, \ldots, m$, compute the $2m$ working covariates

   $$U_{ij}^{(k)} = (z_i - \psi_j^{(k)})_+, \qquad V_{ij}^{(k)} = -\mathbf{1}(z_i > \psi_j^{(k)}), \qquad i = 1, \ldots, n$$

   (b) Fit the linearized model, with linear predictor

   $$\eta_i = x_i'\beta + z_i'\delta_1 + \sum_{j=1}^{m} \xi_j U_{ij}^{(k)} + \sum_{j=1}^{m} \gamma_j^{(k)} V_{ij}^{(k)},$$

   by IRLS, obtaining estimates $\hat{\beta}^{(k)}$, $\hat{\delta}_1^{(k)}$, $\hat{\xi}_j^{(k)}$ and $\hat{\gamma}_j^{(k)}$ for $j = 1, \ldots, m$.

   (c) Update breakpoints. For $j = 1, \ldots, m$:

   $$\psi_j^{(k+1)} = \psi_j^{(k)} + \frac{\hat{\gamma}_j^{(k)}}{\hat{\xi}_j^{(k)}}$$

   (d) Check convergence. Stop if $\max_{j=1,\ldots,m} |\psi_j^{(k+1)} - \psi_j^{(k)}| < \tau$, otherwise return to step (a).

3. Output: Set $\hat{\psi}_j = \psi_j^{(k+1)}$ for $j = 1, \ldots, m$, and recover the regime-specific coefficient estimators: $\hat{\delta}_1 = \hat{\delta}_1^{(k)}$ and $\hat{\delta}_{j+1} = \hat{\delta}_1^{(k)} + \sum_{l=1}^{j} \hat{\xi}_l^{(k)}$ for $j = 1, \ldots, m$.

A.1.2. Chapter 4 - Testing for the existence of a break

A.1.2.1. Empirical Likelihood Ratio (Linear)

Likelihood under the alternative model M_1 As referenced in Section 4.4.1.3, to obtain the likelihood under the alternative model M_1, we maximize $\sum_{i=1}^{n} \log(w_i(\psi_1))$ subject to

simple constraints imposed in $\mathcal{W}$. The Lagrangian is

$$\mathcal{L}(w, \lambda) = \sum_{i=1}^{n} \log(w_i(\psi_1)) + \lambda \left(1 - \sum_{i=1}^{n} w_i(\psi_1)\right)$$

Note that this expression already satisfies $w_i(\psi_1) > 0$.

Differentiating with respect to a specific weight $w_k(\psi_1)$, for a fixed k:

$$\frac{\partial \mathcal{L}}{\partial w_k(\psi_1)} = \frac{1}{w_k(\psi_1)} - \lambda = 0 \quad \Longleftrightarrow \quad w_k(\psi_1) = \frac{1}{\lambda}$$

Therefore, for each index of the weights,

$$w_k(\psi_1) = \frac{1}{\lambda}, \quad \text{for all } k$$

Applying the constraint $\sum_{i=1}^{n} w_i(\psi_1) = 1$,

$$\sum_{i=1}^{n} w_i(\psi_1) = \sum_{i=1}^{n} \frac{1}{\lambda} = \frac{n}{\lambda} = 1 \quad \Longleftrightarrow \quad \lambda = n \quad \Longleftrightarrow \quad w_i(\psi_1) = \frac{1}{n}, \text{ for all } i$$

Optimization Problem We can construct the Lagrangian using the multipliers μ and λ

$$\mathcal{L}(w, \mu, \lambda) = \sum_{i=1}^{n} \log w_i(\psi_1) - \mu \left(\sum_{i=1}^{n} w_i(\psi_1) - 1\right) - \lambda \left(\sum_{i=1}^{n} w_i(\psi_1)\tilde{u}_i(\psi_1)\right)$$

Differentiating with respect to $w_k(\psi_1)$ for a fixed k and setting the result to zero we get

$$\frac{\partial \mathcal{L}}{\partial w_k(\psi_1)} = \frac{1}{w_k(\psi_1)} - \mu - \lambda \tilde{u_k}(\psi_1) = 0 \quad \Rightarrow \quad w_k(\psi_1) = \frac{1}{\mu + \lambda \tilde{u_k}(\psi_1)}$$

Since

$$w_k(\psi_1) = \frac{1}{\mu + \lambda \tilde{u_k}(\psi_1)} \quad \Leftrightarrow \quad \mu + \lambda \tilde{u_k}(\psi_1) = \frac{1}{w_k \psi_1)} \quad \Leftrightarrow \quad w_k(\psi_1)\mu + \lambda w_k(\psi_1)\tilde{u_k}(\psi_1) = 1 \quad \text{for all } k,$$

summing over all indices:

$$\sum_{i=1}^{n} w_i(\psi_1)\mu + \sum_{i=1}^{n} \lambda w_i(\psi_1)\tilde{u_i}(\psi_1) = \sum_{i=1}^{n} 1 \quad \Leftrightarrow \quad \mu \sum_{i=1}^{n} w_i(\psi_1) + \lambda \sum_{i=1}^{n} w_i(\psi_1)\tilde{u_i}(\psi_1) = n$$

Applying the constraints $\sum_{i=1}^{n} w_i(\psi_1) = 1$ and $\sum_{i=1}^{n} w_i(\psi_1)\tilde{u_i}(\psi_1) = 0$:

$$\mu = n$$

Substituting back into the expression for the individual weights:

$$w_i(\psi_1) = \frac{1}{n + \lambda \tilde{u_j}(\psi_1)} = \frac{1}{n(1 + (\lambda/n)\tilde{u_i}(\psi_1))}$$

Renaming λ/n as λ, the standard EL normalization gives

$$w_i(\psi_1) = \frac{1}{n(1+\lambda(\psi_1)\tilde{u}_i(\psi_1))}, \qquad i = 1, \ldots, n$$

The multiplier $\lambda(\psi_1)$, involved in the expression of $w_i(\psi_1)$ is determined by the moment constraint, $\sum_{i=1}^{n} w_i(\psi_1)\,\tilde{u}_i(\psi_1) = 0$,

$$\sum_{i=1}^{n} \frac{1}{n(1+\lambda(\psi_1)\,\tilde{u}_i(\psi_1))}\tilde{u}_i(\psi_1) = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{n} \frac{\tilde{u}_i(\psi_1)}{1+\lambda(\psi_1)\,\tilde{u}_i(\psi_1)} = 0$$

A.2. Testing for the Existence of a Break: Pseudo-Algorithms

A.2.1. Sup-F Test

Given data $\{(y_i, x_i, z_i)\}_{i=1}^{n}$, trimming parameter ε and significance level α.

1. Compute $h = \lfloor \varepsilon n \rfloor$ and define $\Psi_{\varepsilon,m} = \{\psi_1 : h \leq \psi_1 \leq n - h\}$.
2. Fit M_0 by OLS on the full sample and compute $RSS_0 = \sum_{i=1}^{n} \hat{u}_{0,i}^2$
3. For each $\psi_1 \in \Psi_{\varepsilon,1}$:
   (a) Fit $M_1(\psi_1)$ by OLS on each segment separately and compute
   $$RSS_1(\psi_1, \ldots, \psi_m) = \sum_{j=1}^{m+1} \sum_{i=\psi_{j-1}+1}^{\psi_j} \hat{u}_{j,i}^2$$
   (b) Compute $F(\psi_1)$ as in Section 4.2.1
   (c) If $F(\psi_1, \ldots, \psi_m) > \sup F$, update $\sup F \leftarrow F(\psi_1)$ and store $\hat{\psi}_1 \leftarrow \psi_1$
4. Reject H_0 at significance level α if $\sup F > c_\alpha(\varepsilon)$, where $c_\alpha(\varepsilon, q)$ is the critical value.

A.2.2. Classical Likelihood Ratio Test

A.2.2.1. Linear Regression

Given data $\{(y_i, x_i, z_i)\}_{i=1}^{n}$, trimming parameter ε, number of bootstrap replications B, and the significance level α.

1. Compute $h = \lfloor \varepsilon n \rfloor$ and define $\Psi_{\varepsilon,1} = \{\psi_1 :\ h \leq \psi_1 \geq n - h\}$.
2. Fit M_0 by OLS on the full sample. Compute
   $$\text{RSS}_0 = \sum_{i=1}^{n} (y_i - x_i'\hat{\beta} - z_i'\hat{\delta})^2, \qquad \hat{\sigma}^2 = \frac{\text{RSS}_0}{n}, \qquad \hat{u}_i = y_i - x_i'\hat{\beta} - z_i'\hat{\delta} \qquad i = 1, \ldots, n$$

3. For $\psi_1 \in \Psi_{\varepsilon,1}$:

   (a) Fit OLS in $R_1(\psi_1)$ and $R_2(\psi_1)$ separately. Compute

$$\mathsf{RSS}_1(\psi_1) = \sum_{i=1}^{n_1} (y_i - x_i'\hat{\beta}(\psi_1) - z_i'\hat{\delta}_1(\psi_1))^2, \qquad \hat{\sigma}_1^2(\psi_1) = \frac{\mathsf{RSS}_1(\psi_1)}{n_1}$$

$$\mathsf{RSS}_2(\psi_1) = \sum_{i=n_1+1}^{n} (y_i - x_i'\hat{\beta}(\psi_1) - z_i'\hat{\delta}_2(\psi_1))^2, \qquad \hat{\sigma}_2^2(\psi_1) = \frac{\mathsf{RSS}_2(\psi_1)}{n_2}$$

   (b) Compute the per-candidate statistic:

$$\lambda(\psi_1) = n\log(\hat{\sigma}^2) - n_1\log(\hat{\sigma}_1^2(\psi_1)) - n_2\log(\hat{\sigma}_2^2(\psi_1))$$

   (c) If $\lambda(\psi_1) > \Lambda$, update $\Lambda \leftarrow \lambda(\psi_1)$ and $\hat{\psi}_1 \leftarrow \psi_1$.

4. Wild bootstrap loop: for $b = 1, \ldots, B$:

   (a) Draw a multiplier vector, independent from the data, $v^{(b)} = (v_1^{(b)}, \ldots, v_n^{(b)})$ with

$$v_i^{(b)} \overset{\text{i.i.d.}}{\sim} \begin{cases} +1 & \text{with probability } 1/2, \\ -1 & \text{with probability } 1/2, \end{cases} \qquad i = 1, \ldots, n,$$

   (b) Generate bootstrap responses under H_0, keeping the covariates (x_i, z_i) fixed:

$$y_i^{(b)} = x_i'\hat{\beta} + z_i'\hat{\delta} + v_i^{(b)}\hat{u}_i, \qquad i = 1, \ldots, n$$

   (c) Apply steps 1–4 to $\{(y_i^{(b)}, x_i, z_i)\}$ to obtain $\Lambda^{(b)}$.

5. Bootstrap inference:

   (a) Bootstrap critical value: let c_α be the empirical $(1-\alpha)$ quantile of $\{\Lambda^{(1)}, \ldots, \Lambda^{(B)}\}$.

   (b) Bootstrap p-value:

$$\hat{p} = \frac{1 + \sum_{b=1}^{B} \mathbf{1}\{\Lambda^{(b)} \geq \Lambda\}}{B+1}$$

   (c) Decision: reject H_0 at level α if $\Lambda > c_\alpha$, equivalently if $\hat{p} \leq \alpha$.

A.2.2.2. GLM

Given data $\{(y_i, x_i, z_i)\}_{i=1}^n$, trimming parameter ε, number of bootstrap replications B, and the significance level α.

1. Compute $h = \lfloor \varepsilon n \rfloor$ and define $\Psi_{\varepsilon,1} = \{\psi_1 : \; h \leq \psi_1 \geq n - h\}$

2. Fit M_0 by IRLS on the full sample. Obtain fitted means $\hat{\mu}_i^{(0)} = g^{-1}(x_i'\hat{\beta} + z_i'\hat{\delta})$, dispersion $\hat{\phi}_0$, and maximized log-likelihood ℓ_0

3. For $\psi_1 \in \Psi_{\varepsilon,1}$:

   (a) Construct the block-diagonal matrix Z and form $X^* = (X \mid Z)$.

   (b) Run IRLS on X^* to obtain $(\hat{\beta}(\psi_1), \hat{\delta}_1(\psi_1), \hat{\delta}_2(\psi_1))$ and compute the maximized log-likelihood $\ell_1(\psi_1)$.

   (c) Compute the per-candidate statistic, $\lambda(\psi_1) = 2[\ell_1(\psi_1) - \ell_0]$

   (d) If $\lambda(\psi_1) > \Lambda$, update $\Lambda \leftarrow \lambda(\psi_1)$ and $\hat{\psi}_1 \leftarrow \psi_1$.

4. Parametric bootstrap loop: for $b = 1, \ldots, B$:

   (a) Generate bootstrap responses under H_0. Using the null fitted means $\hat{\mu}_i^{(0)} = g^{-1}(x_i'\beta + z_i'\hat{\delta})$ and estimated dispersion $\hat{\phi}_0$, generate bootstrap responses as

   $$y_i^{(b)} \sim F(\hat{\mu}_i^{(0)}, \hat{\phi}_0), \qquad i = 1, \ldots, n,$$

   where $F(\cdot)$ denotes the EDF member with mean $\hat{\mu}_i^{(0)}$ and dispersion $\hat{\phi}_0$. For example:

   - Logistic regression: $y_i^{(b)} \sim \text{Bernoulli}(\hat{\mu}_i^{(0)})$
   - Poisson regression: $y_i^{(b)} \sim \text{Poisson}(\hat{\mu}_i^{(0)})$
   - Gamma regression: $y_i^{(b)} \sim \text{Gamma}(\hat{\mu}_i^{(0)}, \hat{\phi}_0)$

   (b) Apply steps 3-4 to the bootstrap sample $\{(y_i^{(b)}, x_i, z_i)\}_{i=1}^n$ to obtain $\Lambda^{(b)}$.

5. Bootstrap inference:

   (a) Bootstrap critical value: let c_α be the empirical $(1-\alpha)$ quantile of $\{\Lambda^{(1)}, \ldots, \Lambda^{(B)}\}$.

   (b) Bootstrap p-value:

   $$\hat{p} = \frac{1 + \sum_{b=1}^{B} \mathbf{1}\{\Lambda^{(b)} \geq \Lambda\}}{B+1}$$

   (c) Decision: reject H_0 at level α if $\Lambda > c_\alpha$, equivalently if $\hat{p} \leq \alpha$.

A.2.3. Empirical Likelihood Ratio Test

A.2.3.1. Linear Regression

Given data $\{(y_i, x_i, z_i)\}_{i=1}^n$, trimming parameter ε, number of bootstrap replications B, and significance level α:

1. Compute $h = \lfloor \varepsilon n \rfloor$ and define $\Psi_{\varepsilon,1} = \{\psi_1 : h \leq \psi_1 \leq n - h\}$.

2. For each $\psi_1 \in \Psi_{\varepsilon,1}$:

   (a) Fit OLS separately on $R_1(\psi_1)$ and $R_2(\psi_1)$ to obtain $(\hat{\beta}(\psi_1), \hat{\delta}_1(\psi_1))$ and $(\hat{\beta}(\psi_1), \hat{\delta}_2(\psi_1))$.

   (b) Check the continuity condition

   - Continuous case: If the estimated intersection satisfies $z'_{\psi_1}\hat{\delta}_1(\psi_1) = z'_{\psi_1}\hat{\delta}_2(\psi_1)$, retain the OLS estimators as is.
   - Not continuous case: enforce continuity by refitting the full sample via the hinge parametrization

   $$y_i = x_i'\beta + z_i'\gamma + \xi\,(z_i - z_{\psi_1})_+ + u_i, \qquad (z_i - z_{\psi_1})_+ = (z_i - z_{\psi_1})\,\mathbf{1}(i \in R_2(\psi_1)),$$

   and set $\hat{\delta}_1(\psi_1) = \hat{\gamma}(\psi_1)$ and $\hat{\delta}_2(\psi_1) = \hat{\gamma}(\psi_1) + \hat{\xi}(\psi_1)$.

   (c) Compute the null consistent residuals

   $$\tilde{u}_i(\psi_1) = \begin{cases} y_i - x_i'\hat{\beta}(\psi_1) - z_i'\hat{\delta}_2(\psi_1), & i \in R_1(\psi_1), \\ y_i - x_i'\hat{\beta}(\psi_1) - z_i'\hat{\delta}_1(\psi_1), & i \in R_2(\psi_1). \end{cases}$$

   (d) Use $\{\tilde{u}_i(\psi_1)\}_{i=1}^n$ as the input to the *el.test* function in the *R* package *emplik* (Zhou, 2026) to compute $-2\log\hat{\mathcal{R}}(\psi_1)$.

3. Compute $M_n = \sup_{\psi_1 \in \Psi_{\varepsilon,1}} -2\log\hat{\mathcal{R}}(\psi_1)$, and $\hat{\psi}_1 = \arg\max_{\psi_1 \in \Psi_{\varepsilon,1}}\{-2\log\hat{\mathcal{R}}(\psi_1)\}$.

4. Wild bootstrap: fit M_0 on the full sample to obtain $\hat{\mu}_i^{(0)} = x_i'\hat{\beta} + z_i'\hat{\delta}$ and residuals $\hat{u}_i = y_i - \hat{\mu}_i^{(0)}$. For $b = 1, \ldots, B$:

   (a) Draw a multiplier vector $v^{(b)} = (v_1^{(b)}, \ldots, v_n^{(b)})$, independ from the data, with

   $$v_i^{(b)} \overset{\text{i.i.d.}}{\sim} \begin{cases} +1 & \text{with probability } 1/2, \\ -1 & \text{with probability } 1/2, \end{cases} \qquad i = 1, \ldots, n$$

   (b) Construct pseudo-responses $y_i^{(b)} = \hat{\mu}_i^{(0)} + v_i^{(b)}\hat{u}_i$, with $i = 1, \ldots, n$.

   (c) Apply steps 2–3 to $\{(y_i^{(b)}, x_i, z_i)\}$ to obtain $M_n^{(b)}$.

5. Bootstrap inference:

   (a) Bootstrap critical value c_α: empirical $(1-\alpha)$ quantile of $\{M_n^{(1)}, \ldots, M_n^{(B)}\}$.

(b) Bootstrap p-value:

$$\hat{p} \;=\; \frac{1+\sum_{b=1}^{B}\mathbf{1}\{M_n^{(b)} \geq M_n\}}{B+1}$$

(c) Reject H_0 at level α if $M_n > c_\alpha$, equivalently if $\hat{p} \leq \alpha$.

A.2.3.2. GLM

Given data $\{(y_i, x_i, z_i)\}_{i=1}^n$, link function $g(\cdot)$, trimming parameter ε, number of bootstrap replications B, and significance level α:

1. Compute $h = \lfloor \varepsilon n \rfloor$ and define $\Psi_{\varepsilon,1}$.
2. For each $\psi_1 \in \Psi_{\varepsilon,1}$:
   (a) Run IRLS separately on $R_1(\psi_1)$ and $R_2(\psi_1)$ to obtain $(\hat{\beta}(\psi_1), \hat{\delta}_j(\psi_1))$ and fitted means $\hat{\mu}_i^{(j)}(\psi_1)$ for $j = 1, 2$.
   (b) Check the continuity condition. If $z'_{\psi_1}\hat{\delta}_1(\psi_1) = z'_{\psi_1}\hat{\delta}_2(\psi_1)$ retain the estimators. Otherwise, enforce continuity via IRLS on the hinge parametrization

   $$\eta_i = x_i'\beta + z_i'\gamma + \xi\,(z_i - z_{\psi_1})_+,$$

   setting $\hat{\delta}_1(\psi_1) = \hat{\gamma}(\psi_1)$ and $\hat{\delta}_2(\psi_1) = \hat{\gamma}(\psi_1) + \hat{\xi}(\psi_1)$.
   (c) Compute null-consistent working residuals:

   $$\tilde{u}_i^*(\psi_1) = \begin{cases} \left(y_i - \hat{\mu}_i^{(2)}(\psi_1)\right) g'(\hat{\mu}_i^{(2)}(\psi_1)), & i \in R_1(\psi_1), \\ \left(y_i - \hat{\mu}_i^{(1)}(\psi_1)\right) g'(\hat{\mu}_i^{(1)}(\psi_1)), & i \in R_2(\psi_1). \end{cases}$$

   (d) Pass $\{\tilde{u}_i^*(\psi_1)\}_{i=1}^n$ to `el.test` in the R package `emplik` to compute $-2\log\hat{\mathcal{R}}(\psi_1)$.
3. Compute

   $$M_n = \sup_{\psi_1 \in \Psi_{\varepsilon,1}} \{-2\log\hat{\mathcal{R}}(\psi_1)\}, \qquad \hat{\psi}_1 = \arg\max_{\psi_1 \in \Psi_{\varepsilon,1}} \{-2\log\hat{\mathcal{R}}(\psi_1)\}$$

4. Multiplier wild bootstrap: for $b = 1, \ldots, B$:
   (a) Draw a single multiplier vector, independent of the data and reused across all ψ_1 in this replicate, $v^{(b)} = (v_1^{(b)}, \ldots, v_n^{(b)})$ with

   $$v_i^{(b)} \overset{\text{i.i.d.}}{\sim} \begin{cases} +1 & \text{with probability } 1/2, \\ -1 & \text{with probability } 1/2, \end{cases} \qquad i = 1, \ldots, n,$$

(b) For each $\psi_1 \in \Psi_{\varepsilon,1}$, form perturbed null-consistent residuals

$$\tilde{u}_i^{*(b)}(\psi_1) = v_i^{(b)} \tilde{u}_i^*(\psi_1),$$

and pass them to `el.test` to obtain $-2\log \hat{\mathcal{R}}^{(b)}(\psi_1)$.

(c) Compute

$$M_n^{(b)} = \sup_{\psi_1 \in \Psi_{\varepsilon,1}} \{-2\log \hat{\mathcal{R}}^{(b)}(\psi_1)\}$$

5. Bootstrap inference:

(a) Bootstrap critical value c_α^*: empirical $(1-\alpha)$ quantile of $\{M_n^{(1)}, \ldots, M_n^{(B)}\}$.

(b) Bootstrap p-value:

$$\hat{p}^* = \frac{1 + \sum_{b=1}^{B} \mathbf{1}\{M_n^{(b)} \geq M_n\}}{B+1}$$

(c) Reject H_0 at level α if $M_n > c_\alpha^*$, equivalently if $\hat{p}^* \leq \alpha$.

The same procedure applies to the raw-residual variant by substituting $\tilde{u}_i^*(\psi_1)$ with $\tilde{u}_i(\psi_1)$ throughout.

A.2.4. Recursive Residual CUSUM

Given the observed data $\{(y_i, x_i, z_i)\}_{i=1}^n$, the number of stable regressors p, the number of regime-varying regressors q, and the significance level α.

1. Set $k = p + q$. Form the null model design matrix $X^{*(0)} = (X \mid Z) \in \mathbb{R}^{n \times k}$, with rows $x_i^{*(0)} = (x_i', z_i')' \in \mathbb{R}^k$.

2. Using the first k observations, set $S_k = 0$, compute the initial OLS estimator

$$\hat{\theta}_k^{(0)} = \left(X_k^{*(0)\prime} X_k^{*(0)}\right)^{-1} X_k^{*(0)\prime} Y_k, \qquad Y_k = (y_1, \ldots, y_k)$$

and the initial inverse $(X_k^{*(0)\prime} X_k^{*(0)})^{-1}$.

3. For $i = k+1, \ldots, n$:

(a) Update the inverse, $(X_i^{*(0)\prime} X_i^{*(0)})^{-1}$, using the Sherman-Morrison formula.

(b) Compute the recursive residual:

$$w_i = \frac{y_i - x_i^{*(0)\prime} \hat{\theta}_{i-1}^{(0)}}{\left(1 + x_i^{*(0)\prime} \left(X_{i-1}^{*(0)\prime} X_{i-1}^{*(0)}\right)^{-1} x_i^{*(0)}\right)^{1/2}}$$

   (c) Update the OLS estimator: $\hat{\theta}_i^{(0)} = \hat{\theta}_{i-1}^{(0)} + \left(X_i^{*(0)\prime} X_i^{*(0)}\right)^{-1} x_i^{*(0)} \left(y_i - x_i^{*(0)\prime} \hat{\theta}_{i-1}^{(0)}\right)$

   (d) Update the cumulative sum: $S_i = S_{i-1} + w_i$

4. Estimate the residual standard deviation: $\hat{\sigma} = \left[\frac{1}{n-k} \sum_{i=k+1}^{n} w_i^2\right]^{1/2}$

5. Evaluate the standardized CUSUM process: $C(i) = \frac{S_i}{\hat{\sigma}\sqrt{n-k}}, \qquad i = k+1, \ldots, n$

6. Compute the test statistic: $\mathrm{CUSUM} = \sup_{i \in \{k+1,\ldots,n\}} |C(i)|$

7. Obtain the p-value: $p = 2\sum_{r=1}^{R} (-1)^{r-1} e^{-2r^2 \cdot \mathrm{CUSUM}^2}$, taking R large enough for the desired precision.

8. Decision rule: reject H_0 at level α if $p < \alpha$.

9. Breakpoint estimate if H_0 is rejected:

$$\hat{\psi}_1 = \arg \max_{i \in \{h,\ldots,n-h\}} |C(i)| \qquad \text{with} \qquad h = \max(k+1, \lfloor \varepsilon n \rfloor)$$

A.2.5. Score-CUSUM

Given $\{(y_i, x_i, z_i)\}_{i=1}^{n}$, significance level α, and trimming fraction ε.

1. Set $k = p + q$ and $h = \lfloor \varepsilon n \rfloor$. Form the null model design matrix $X^{*(0)} = (X \mid Z) \in \mathbb{R}^{n \times k}$ with rows $x_i^{*(0)} = (x_i', z_i')' \in \mathbb{R}^k$.

2. Fit the null GLM via IRLS to obtain $\hat{\theta}^{(0)} = (\hat{\beta}^{(0)\prime}, \hat{\delta}^{(0)\prime})'$, estimated dispersion $\hat{\phi}^{(0)}$, diagonal null weight matrix $W^{(0)} = \mathrm{diag}(w_1^{(0)}, \ldots, w_n^{(0)})$, and null fitted means $\hat{\mu}^{(0)} = (\hat{\mu}_1^{(0)}, \ldots, \hat{\mu}_n^{(0)})'$.

3. For $i = 1, \ldots, n$, compute:

   (a) Null working residuals $r_i^{*(0)} = (y_i - \hat{\mu}_i^{(0)}) g'(\hat{\mu}_i^{(0)})$

   (b) Individual score contributions $s_i(\hat{\theta}^{(0)}, \hat{\phi}^{(0)}) = w_i^{(0)} r_i^{*(0)} x_i^{*(0)} \in \mathbb{R}^k$

4. Compute the Fisher information matrix $\hat{I}^{(0)} = X^{*(0)\prime} W^{(0)} X^{*(0)} \in \mathbb{R}^{k \times k}$ and its inverse square root via eigen-decomposition: find orthonormal P and diagonal D such that $P \hat{I}^{(0)} P' = D$, then set $(\hat{I}^{(0)})^{-1/2} = P' D^{-1/2} P$.

5. For $i = 1, \ldots, n$, form the normalized cumulative score process,

$$C^{\mathrm{score}}(i) = \frac{1}{\sqrt{n}} (\hat{I}^{(0)})^{-1/2} \sum_{l=1}^{i} s_l(\hat{\theta}^{(0)}, \hat{\phi}^{(0)}) \in \mathbb{R}^k$$

6. Compute the test statistic, either:

$$\text{ScoreCUSUM} = \sup_{i=1,\dots,n} ||C^{\text{score}}(i)||_2, \qquad \text{ScoreCUSUM} = \frac{1}{n}\sum_{i=1}^{n} C^{\text{score}}(i)' C^{\text{score}}(i)$$

7. Obtain the critical value c_α from the simulated limit distributions for the chosen statistic and the given k, as tabulated in Zeileis (2006), or by parametric bootstrap resampling from the fitted null GLM.

8. Reject H_0 at level α if the test statistic exceeds c_α.

9. If H_0 is rejected, estimate the breakpoint as $\hat{\psi}_1 = \arg\max_{i\in\{h,\dots,n-h\}} ||C^{\text{score}}(i)||_2$.

A.2.6. Score MOSUM

Given $\{(y_i, x_i, z_i)\}_{i=1}^n$, significance level α, and bandwidth $b \in (0, 1)$.

1. Set $k = p + q$ and $h = \lfloor bn \rfloor$. Form the null model design matrix $X^{*(0)} = (X \mid Z) \in \mathbb{R}^{n\times k}$ with rows $x_i^{*(0)} = (x_i', z_i')' \in \mathbb{R}^k$.

2. Fit the null GLM via IRLS to obtain $\hat{\theta}^{(0)} = (\hat{\beta}^{(0)'}, \hat{\delta}^{(0)'})'$, estimated dispersion $\hat{\phi}^{(0)}$, diagonal null weight matrix $W^{(0)} = \text{diag}(w_1^{(0)}, \dots, w_n^{(0)})$, and null fitted means $\hat{\mu}^{(0)} = (\hat{\mu}_1^{(0)}, \dots, \hat{\mu}_n^{(0)})'$.

3. For $i = 1, \dots, n$, compute:

   (a) Null working residuals $r_i^{*(0)} = (y_i - \hat{\mu}_i^{(0)}) g'(\hat{\mu}_i^{(0)})$

   (b) Individual score contributions $s_i(\hat{\theta}^{(0)}, \hat{\phi}^{(0)}) = w_i^{(0)} r_i^{*(0)} x_i^{*(0)} \in \mathbb{R}^k$

4. Compute the Fisher information matrix $\hat{I}^{(0)} = X^{*(0)'} W^{(0)} X^{*(0)} \in \mathbb{R}^{k\times k}$ and its inverse square root via eigen-decomposition: find orthonormal P and diagonal D such that $P\hat{I}^{(0)}P' = D$, then set $(\hat{I}^{(0)})^{-1/2} = P'D^{-1/2}P$.

5. For $i = h+1, \dots, n$, form the moving-window score process,

$$M^{\text{score}}(i) = \frac{1}{\sqrt{h}} (\hat{I}^{(0)})^{-1/2} \sum_{l=i-h+1}^{i} s_l(\hat{\theta}^{(0)}, \hat{\phi}^{(0)}) \in \mathbb{R}^k$$

   at each index i, the window sums score contributions over the h most recent observations, dropping the oldest contribution and adding the newest as i advances.

6. Compute the test statistic:

$$\text{Score MOSUM} = \sup_{i\in\{h+1,\dots,n\}} ||M^{\text{score}}(i)||_2$$

7. Obtain the critical value c_α from the simulated limit distribution of $\sup_{\tau\in[b,1]}\left\|\frac{W(\tau)-W(\tau-b)}{\sqrt{b}}\right\|$, depending on b and k, as tabulated in Zeileis (2006), or by parametric bootstrap re-sampling from the fitted null GLM.

8. Reject H_0 at level α if the test statistic exceeds c_α.

9. If H_0 is rejected, estimate the breakpoint at the centre of the window achieving the supremum:

$$\hat{\psi}_1 = \arg\max_{i\in\{h+1,\dots,n\}} ||M^{\text{score}}(i)||_2 - \left\lfloor\frac{h}{2}\right\rfloor$$

A.2.7. Supremum Score Test

Given data $\{(y_i, x_i, z_i)\}_{i=1}^n$, link function $g(\cdot)$, variance function $V(\cdot)$, trimming parameter ε, and significance level α:

1. Compute $h = \lfloor \varepsilon n \rfloor$ and define $\Psi_{\varepsilon,1} = \{\psi_1 : \psi_1 \geq h,\ n - \psi_1 \geq h\}$ with admissible grid $\psi_1^{(1)} < \cdots < \psi_1^{(K)}$.

2. Fit the null model M_0 by IRLS on the full sample with design matrix $X^{*(0)} = (X \mid Z)$. Obtain $(\hat{\beta}^{(0)}, \hat{\delta}^{(0)}, \hat{\phi}^{(0)})$, fitted means $\hat{\mu}_i^{(0)}$, weight matrix $W^{(0)}$, null working residuals $R^{*(0)} = G^{(0)}(Y - \hat{\mu}^{(0)})$, and weighted hat matrix $H_w^{(0)}$. These quantities are computed once and reused for all candidates.

3. For each $\psi_1^{(i)} \in \Psi_{\varepsilon,1}$, $i = 1, \dots, K$:

   (a) Construct the covariate $U(\psi_1^{(i)}) = \left((z_1 - \psi_1^{(i)})_+, \dots, (z_n - \psi_1^{(i)})_+\right)' \in \mathbb{R}^{n\times q}$ and its weighted version $U_w(\psi_1^{(i)}) = W^{(0)1/2} U(\psi_1^{(i)})$.

   (b) Compute the score: $s_\xi(\psi_1^{(i)}) = U(\psi_1^{(i)})'\, W^{(0)}\, R^{*(0)}$

   (c) Compute the partial information: $\mathcal{I}_{\xi\xi\cdot\theta^{(0)}}(\psi_1^{(i)}) = U_w(\psi_1^{(i)})'(I_n - H_w^{(0)})U_w(\psi_1^{(i)})$

   (d) Compute the per-candidate score statistic: $T(\psi_1^{(i)}) = s_\xi(\psi_1^{(i)})'\, \mathcal{I}_{\xi\xi\cdot\theta^{(0)}}(\psi_1^{(i)})^{-1}\, s_\xi(\psi_1^{(i)})$

   (e) If $T(\psi_1^{(i)}) > \mathcal{T}$, update $\mathcal{T} \leftarrow T(\psi_1^{(i)})$ and $\hat{\psi}_1 \leftarrow \psi_1^{(i)}$.

4. Compute the Davies bound:

   (a) Compute the nominal p-value, plug-in variability estimate and correction factor

$$p_0 = P\Big(\chi^2(q) \geq \mathcal{T}\Big),\ \hat{V} = \sum_{i=1}^{K-1}\left|\sqrt{T(\psi_1^{(i+1)})} - \sqrt{T(\psi_1^{(i)})}\right|,\ g(\mathcal{T}, q) = \frac{\mathcal{T}^{(q-1)/2}\exp(-\mathcal{T}/2)}{2^{q/2}\,\Gamma(q/2)}$$

   (b) Compute the Davies bound: $\hat{p} = p_0 + \hat{V}\cdot g(\mathcal{T}, q)$.

5. Reject H_0 at level α if $\hat{p} \leq \alpha$.

A.2.8. Average Score Test

Given data $\{(y_i, x_i, z_i)\}_{i=1}^n$, link function $g(\cdot)$, variance function $V(\cdot)$, trimming parameter ε, and significance level α:

1. Compute $h = \lfloor \varepsilon n \rfloor$ and define $\Psi_{\varepsilon,1} = \{\psi_1 : \psi_1 \geq h,\ n - \psi_1 \geq h\}$.

2. Fit the null model M_0 by IRLS on the full sample with design matrix $X^{*(0)} = (X \mid Z)$. Obtain $(\hat{\beta}^{(0)}, \hat{\delta}^{(0)}, \hat{\phi}^{(0)})$, fitted means $\hat{\mu}_i^{(0)}$, weight matrix $W^{(0)}$, null working residuals $R^{*(0)} = G^{(0)}(Y - \hat{\mu}^{(0)})$, and weighted hat matrix $H_w^{(0)}$

   These quantities are computed once and reused for all candidates.

3. Construct the average hinge matrix. For each observation $i = 1, \ldots, n$, compute

$$A = (A_1', \ldots, A_n')' \in \mathbb{R}^{n \times q}, \quad A_w = W^{(0)1/2} A, \quad \text{with} \quad A_i = \frac{1}{K} \sum_{k=1}^{K} (z_i - \psi_k)_+ \in \mathbb{R}^q$$

4. Compute the score of the alternative model with respect to ξ, evaluated at the null estimates: $s_\xi = A'\, W^{(0)}\, R^{*(0)} = A_w'\, R_w^{*(0)}$.

5. Compute the partial Fisher information for ξ: $\mathcal{I}_{\xi\xi \cdot \theta^{(0)}} = A_w'(I_n - H_w^{(0)})A_w$

6. Compute the average score statistic:

$$s_0 = s_\xi'\, \mathcal{I}_{\xi\xi \cdot \theta^{(0)}}^{-1}\, s_\xi = (A_w' R_w^{*(0)})'\, (A_w'(I_n - H_w^{(0)})A_w)^{-1}\, (A_w' R_w^{*(0)})$$

7. Compute the p-value: $\hat{p} = P(\chi^2(q) \geq s_0)$. Reject H_0 at level α if $\hat{p} \leq \alpha$.

A.3. PBT GoF Test for Segmented Regression Models: Pseudo-algorithms

A.3.1 Linear Changepoint Regression Models

Given data $\{(y_i, x_i, z_i)\}_{i=1}^n$, trimming parameter ε, number of bootstrap replications B, and significance level α.

1. Fit the segmented model to $\{(y_i, x_i, z_i)\}_{i=1}^n$, obtain $\hat{\theta} = (\hat{\beta}', \hat{\delta}_1', \hat{\delta}_2', \hat{\psi}_1)'$, the fitted means $\hat{\mu}_i$, and the residuals $\hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)'$.

2. Form the full design matrix $X^* = (X \mid Z)$, where Z is block-diagonal at $\hat{\psi}_1$. Compute the symmetric $n \times n$ matrix $A = (A_{is})$ with $A_{is} = \sum_{l=1}^n A_{isl}$ and

$$A_{isl} = C_{p+2q} \left| \pi - \arccos\left(\frac{(X_i^* - X_l^*)'(X_s^* - X_l^*)}{\|X_i^* - X_l^*\|\, \|X_s^* - X_l^*\|} \right) \right|,$$

with $A_{isl} = 0$ when $X_i^* = X_l^*$ or $X_s^* = X_l^*$, and $C_{p+2q} = \frac{\pi^{(p+2q)/2-1}}{\Gamma((p+2q)/2+1)}$.

3. Compute the observed test statistic $T_{\text{seg}} = \frac{1}{n^2}\hat{\varepsilon}' A \hat{\varepsilon}$.

4. For $b = 1, \ldots, B$:

   (a) Draw a multiplier vector $v^{(b)} = (v_1^{(b)}, \ldots, v_n^{(b)})$ of i.i.d. variables with $E(v_i) = 0$, $\text{Var}(v_i) = 1$, bounded support, and independent of the data. A common choice is the Rademacher distribution, $v_i = \pm 1$ each with probability $1/2$.

   (b) Construct bootstrap responses $y_i^{(b)} = \hat{\mu}_i + \hat{\varepsilon}_i \, v_i^{(b)}$ for $i = 1, \ldots, n$, keeping x_i, z_i fixed.

   (c) Refit the segmented model on $\{(y_i^{(b)}, x_i, z_i)\}_{i=1}^n$ to obtain $\hat{\theta}^{(b)} = (\hat{\beta}^{(b)\prime}, \hat{\delta}_1^{(b)\prime}, \hat{\delta}_2^{(b)\prime}, \hat{\psi}_1^{(b)})'$ and bootstrap residuals $\hat{\varepsilon}^{(b)}$. Form $X^{*(b)} = (X \mid Z^{(b)})$, where $Z^{(b)}$ is block-diagonal at $\hat{\psi}_1^{(b)}$, and compute $A^{(b)}$ as in step 2.

   (d) Compute the bootstrap statistic $T_{\text{seg}}^{(b)} = \frac{1}{n^2} \, (\hat{\varepsilon}^{(b)})' A^{(b)} \, \hat{\varepsilon}^{(b)}$.

5. Bootstrap inference:

   (a) Bootstrap critical value c_α: empirical $(1-\alpha)$ quantile of $\{T_{\text{seg}}^{(1)}, \ldots, T_{\text{seg}}^{(B)}\}$.

   (b) Bootstrap p-value: $\hat{p} = \frac{1 + \sum_{b=1}^{B} \mathbf{1}(T_{\text{seg}}^{(b)} \geq T_{\text{seg}})}{B+1}$.

   (c) Reject H_0 at level α if $T_{\text{seg}} > c_\alpha$, equivalently if $\hat{p} \leq \alpha$.

A.3.2 GLM Changepoint Regression Models

Given data $\{(y_i, x_i, z_i)\}_{i=1}^n$, trimming parameter ε, number of bootstrap replications B, and significance level α.

1. Fit the segmented model to $\{(y_i, x_i, z_i)\}_{i=1}^n$, obtain $\hat{\theta}$, fitted means $\hat{\mu}_i$, estimated dispersion $\hat{\phi}$, and residuals $\hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)'$.

2. Form $X^* = (X \mid Z)$ and compute A and T_{seg}.

3. For $b = 1, \ldots, B$:

   (a) Draw bootstrap responses independently from the fitted segmented model, keeping the covariates X_i^* fixed:

   $$y_i^{(b)} \sim F(\hat{\mu}_i, \hat{\phi}), \qquad i = 1, \ldots, n,$$

   where $F(\cdot)$ denotes the EDF member with mean $\hat{\mu}_i$ and estimated dispersion $\hat{\phi}$. For example:

- Logistic regression: $y_i^{(b)} \sim \text{Bernoulli}(\hat{\mu}_i)$,
- Poisson regression: $y_i^{(b)} \sim \text{Poisson}(\hat{\mu}_i)$.

(b) Refit the segmented model on $\{(y_i^{(b)}, x_i, z_i)\}_{i=1}^n$, obtain $\hat{\theta}^{(b)}$ and bootstrap residuals $\hat{\varepsilon}^{(b)}$. Form $X^{*(b)}$ at $\hat{\psi}_1^{(b)}$ and compute $A^{(b)}$ and $T_{\text{seg}}^{(b)} = \frac{1}{n^2}\,(\hat{\varepsilon}^{(b)})' A^{(b)}\,\hat{\varepsilon}^{(b)}$.

4. Bootstrap inference:

(a) Bootstrap critical value c_α: empirical $(1-\alpha)$ quantile of $\{T_{\text{seg}}^{(1)}, \ldots, T_{\text{seg}}^{(B)}\}$.

(b) Bootstrap p-value: $\hat{p} = \frac{1+\sum_{b=1}^{B} \mathbf{1}(T_{\text{seg}}^{(b)} \geq T_{\text{seg}})}{B+1}$.

(c) Reject H_0 at level α if $T_{\text{seg}} > c_\alpha$, equivalently if $\hat{p} \leq \alpha$.