Changing Patterns of Fertility in Sweden

A Time Series Analysis of Structural Breaks and Sequential Detection, 1950–2025

Lisa von Ahnen, Nora Kraft

Master's thesis
2026:E57

Lund Institute of Technology
Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

CENTRUM SCIENTIARUM MATHEMATICARUM

Contents

Abstract

Sweden has experienced pronounced long-run fluctuations in fertility since the mid-twentieth century, alongside persistent within-year seasonality in births. While many fertility movements can be interpreted *ex post* in relation to economic conditions, cohort size, and institutional reforms, a central practical challenge is *ex ante*: *when* does an emerging decline become statistically detectable as the change unfolds, rather than being mistaken for an ordinary cycle?

Using monthly live births in Sweden from 1900–2024 and annual fertility indicators from 1950 onward, sourced from Statistics Sweden (SCB) (2026) and the Human Fertility Database, this thesis evaluates the prospective detectability of gradual fertility declines using the CUSUM algorithm. Because standard CUSUM procedures target mean shifts, slope changes in the original series are transformed into approximate mean shifts by detrending and deseasonalising a reference regime, computing residuals, and differencing. Decision thresholds are calibrated via Monte Carlo simulation from the estimated reference-period AR(1) model to control the false alarm rate under serial dependence.

The framework is applied to three historical episodes: the diffusion of oral contraceptives (1965), the Swedish economic crisis (1991), and the post-2010 fertility decline. Detectability is compared across fertility indicators, including monthly births, Total Fertility Rate (TFR), tempo-adjusted TFR, and age-specific fertility rates. Three patterns emerge consistently. First, observed TFR is the most reliable early-warning indicator, detecting in all three episodes with delays of one to three years, while monthly births detect promptly only once aggregated to the annual level. Second, tempo-adjusted indicators detect later than their observed counterparts or not at all, providing a structured statistical lens on the tempo–quantum distinction central to demographic theory. Third, the age-group pattern of detection delays distinguishes broad economic shocks (uniform delays across age groups) from postponement-driven declines (rapid detection at peak childbearing ages but long delays at younger ages). Cross-validation with offline structural break methods (Bai–Perron, slope change test, Chow test) indicates that breaks confirmed in hindsight may have been statistically indistinguishable from ordinary fluctuation for several years after the underlying change occurred, quantifying a surveillance lag with direct implications for the design of demographic early warning systems.

Acknowledgements

A special thank you goes to our supervisor, Andreas Jakobsson, for the many late evenings he has dedicated to our supervision meetings throughout the semester. His guidance, support, and flexibility have been invaluable and we are deeply grateful for the time and engagement he has devoted to this thesis.

List of abbreviations

AR(1)	Autoregressive process of order 1
ARL	Average Run Length
ARMA	Autoregressive Moving Average
ASFR	Age-Specific Fertility Rate
CUSUM	Cumulative Sum
HFD	Human Fertility Database
MAB	Mean Age at Childbearing
MDC	Minimum Detectable Change
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
PELT	Pruned Exact Linear Time
RFSU	Riksförbundet för sexuell upplysning
SCB	Statistics Sweden
SDT	Second Demographic Transition
TFR	Total Fertility Rate

1 Introduction

Swedish fertility patterns have changed substantially over the past seventy years. Periods of rising fertility have alternated with sustained declines, driven by changes in social norms, family policies, economic conditions, and demographic composition. The Total Fertility Rate (TFR) peaked at approximately 2.5 in the early 1950s, fell to around 1.6 in the mid-1970s, recovered to near replacement level by 1990, collapsed during the 1990s economic crisis, and has since declined to a postwar record low of 1.43 in 2024 (Statistics Sweden (SCB) 2026). Underlying this cyclical pattern is a secular postponement of childbearing that distorts period fertility measures and complicates the interpretation of observed TFR movements (Bongaarts and Feeney 1998).

A central feature of fertility data is strong seasonality. Monthly birth counts exhibit systematic within-year variation that persists throughout the sample period. At the same time, longer-run fertility movements often unfold gradually, making it difficult to distinguish emerging structural changes from ordinary cyclical fluctuations.

This thesis focuses on the statistical problem of identifying structural changes within a short time window, rather than explaining fertility change retrospectively from the complete dataset. The main focus will be on three historical episodes: the diffusion of oral contraceptives (1965), the Swedish economic crisis (1991), and the post-2010 fertility decline.

1.1 Background

Fertility is a core demographic driver with long-run implications for population age structure, labour supply, and the financing of welfare systems. In Sweden, fertility developments have been interpreted in relation to institutional reforms (J. M. Hoem 1993), macroeconomic shocks (B. Hoem 2000; Andersson 2000), and broader shifts in family-formation behaviour (Hellstrand et al. 2021; Ohlsson-Wijk and Andersson 2022). Yet the timing of these changes is rarely obvious as the data arrive: a new decline may be indistinguishable from ordinary variation for several years after the true structural break.

1.2 Motivation

While many studies explain fertility decline retrospectively, much less attention is given to when a decline first becomes statistically visible in incoming data. This is particularly relevant when changes are gradual rather than abrupt and when the choice of fertility indicator, such as monthly births, TFR, tempo-adjusted TFR, or age-specific rates, directly affects how quickly a change accumulates sufficient evidence. Sequential monitoring methods such as the Cumulative Sum (CUSUM) procedure, introduced by Page (1954), are therefore of direct interest for fertility analysis and their application to demographic indicators is, to our knowledge, not yet researched.

1.3 Purpose of the thesis

The purpose of this thesis is to analyse structural changes in Swedish fertility and evaluate how quickly such changes can be detected using time-series methods. Particular emphasis is put on sequential monitoring through CUSUM and evaluating the registered alarms and their delays in months or years from the chosen events.

1.4 Research questions

The thesis addresses the following research questions:

- How have long-run trends and seasonal patterns in Swedish fertility evolved between 1950 and 2025?
- Are there statistically detectable structural breaks in Swedish fertility series and how do they align with the three historical episodes?
- How quickly can fertility declines be detected using sequential monitoring methods such as CUSUM and how does detection delay vary across indicators?
- What does the comparison between CUSUM alarms and offline structural break methods reveal about the detectability of gradual fertility change?

1.5 Structure of the thesis

Section 2 presents background and theory. Section 3 describes the data, methods, and results. Section 4 discusses the findings and their limitations. Section 5 concludes.

2 Background and Theory

2.1 Fertility indicators

2.1.1 Number of births

The number of live births is the most direct measure of reproductive output in a given population and period. In Sweden, Statistics Sweden (SCB) publishes monthly birth counts with a consistent methodology stretching back to the early twentieth century (Statistics Sweden (SCB) 2026), making the series one of the longest and most reliable high-frequency demographic records available for any country.

Monthly birth counts are particularly useful for studying short-run fluctuations, seasonal patterns, and the precise timing of turning points in fertility trends. Strong within-year seasonality in births, with peaks typically in March and troughs in late autumn, has been documented across a wide range of populations and reflects a combination of behavioural, biological and institutional factors that govern the timing of conception (Lam and Miron 1991; Lam, Miron, and Riley 1994). Because seasonal regularities are stable, systematic departures from the expected within-year pattern can serve as an early indicator that underlying fertility behaviour has changed.

A fundamental limitation of raw birth counts, however, is that they are driven not only by the proneness to have children but also by the size and age structure of the female population. A cohort of unusually large or small size will generate a correspondingly large or small number of births even if age-specific fertility rates remain unchanged. This makes longitudinal comparisons of birth counts across periods with different cohort compositions potentially misleading and motivates the use of standardised fertility rates.

2.1.2 Total Fertility Rate (TFR)

The Total Fertility Rate (TFR) is the most widely used summary measure of period fertility. It is defined as the sum of age-specific fertility rates across the full reproductive age span, and can be interpreted as the number of children a woman would bear over her lifetime if she were exposed to the currently observed age-specific rates throughout her reproductive years (Preston et al. 2001):

$$\mathrm{TFR} = \sum_{a=15}^{49} \mathrm{ASFR}_a \tag{1}$$

where ASFR_a is the fertility rate at age a, expressed as births per woman per year. Because TFR standardises for the age composition of the female population, it provides a more behaviourally interpretable measure than raw birth counts.

An important warning is that TFR is a *period* construct: it combines rates from women of different birth cohorts who are observed simultaneously at different stages of their reproductive careers. It therefore does not reflect the completed fertility of any actual cohort. Moreover, as demonstrated formally by Ryder (1964) and later by Bongaarts and Feeney (1998), period TFR is sensitive to changes in the *timing* of births. When women systematically shift the age at which they bear children, even without any change in their intended family size, period TFR rises or falls in ways that can be substantially misleading. Sobotka and Lutz (2011) argue that these distortions are large enough to produce policy-relevant errors when TFR is used as the sole indicator of fertility trends.

2.1.3 Tempo-adjusted TFR

The tempo effect arises because postponement compresses the birth events of multiple cohorts into the future: fewer births are observed in the present than would be expected from the underlying quantum of fertility. Bongaarts and Feeney (1998) proposed correcting for this distortion by dividing period TFR by one minus the annual change in mean age at childbearing:

$$\text{TFR}^* = \frac{\text{TFR}}{1 - r} \tag{2}$$

where $r = \Delta\bar{a}$ is the annual change in the mean age at childbearing. When women are postponing births ($r > 0$), the observed TFR understates the quantum of fertility, and $\text{TFR}^* > \text{TFR}$. Conversely, when timing shifts reverse and women bring births forward, TFR can be temporarily inflated above the underlying quantum.

The Bongaarts–Feeney adjustment has been extended in various ways. Kohler and Philipov (2001) incorporate variance effects, showing that changes in the dispersion of the age schedule, not only changes in its mean, can also distort period TFR. Bongaarts and Sobotka (2012) propose a combined tempo- and parity-adjusted measure. Critiques by Ní Bhrolcháin (2011) point out that whether tempo change constitutes genuine "distortion" depends on the purpose of the measure and on whether period or cohort fertility is the more appropriate policy target. For the present thesis, the Bongaarts–Feeney formula is adopted as a standard and tractable adjustment that allows the assessment of how much of any detected CUSUM alarm represents genuine quantum change versus timing shifts.

2.1.4 Age-specific fertility rates

Age-specific fertility rates (ASFRs) decompose the overall fertility level into the contribution of each single-year or five-year age group:

$$\text{ASFR}_a = \frac{B_a}{N_a} \tag{3}$$

where B_a is the number of births to women aged a and N_a is the female mid-year population at age a. Examining ASFRs separately allows one to identify which parts of the age schedule are driving changes in total fertility and to distinguish tempo from quantum effects more precisely: a downward shift of fertility rates at younger ages combined with an upward shift at older ages is the fingerprint of postponement, whereas a uniform decline across all ages is more consistent with a genuine reduction in desired family size (Billari and Kohler 2004; Hellstrand et al. 2021).

In the present analysis, ASFRs are available by five-year age group (15–19, 20–24, 25–29, 30–34, 35–39, 40–44). Applying CUSUM separately to each age group provides information on which segment of the age schedule changes first and most sharply following a given shock, a question that is directly relevant for early detection.

2.1.5 Mean age at childbearing

The mean age at childbearing (MAB) is defined as the average age of mothers at the time of birth, computed from the age-specific birth distribution. It serves as the primary summary statistic for the timing dimension of fertility:

$$\overline{a} = \frac{\sum_a a \cdot B_a}{\sum_a B_a} \tag{4}$$

Rising MAB indicates postponement, and the annual increment $r = \Delta\overline{a}$ enters directly into the Bongaarts–Feeney tempo adjustment. In Sweden, MAB has risen from approximately 27 years in the early 1970s to over 31 years by the early 2020s (Statistics Sweden (SCB) 2026), which indicates a multi-decade postponement transition that is among the most pronounced in Europe. Because MAB is a smooth, slowly evolving series, it also provides a stable input to the tempo correction even in years with noisy period fertility data.

2.2 Fertility trends in Sweden and the Nordic countries

2.2.1 Long-term fertility patterns, 1950–2025

Swedish fertility since the mid-twentieth century has followed a distinctive roller-coaster trajectory that sets it apart from the monotonically declining path observed in many Southern and Eastern European countries. Understanding this trajectory is essential context for interpreting the structural breaks that the present analysis aims to detect.

The postwar period opened with a sustained baby boom. TFR reached approximately 2.5 in the early 1950s before declining gradually through the late 1950s and stabilising around 2.2 during the early 1960s. A sharper decline began in the mid-1960s, coinciding with the diffusion of modern contraception (Section 2.3.1), and continued into the mid-1970s, when TFR bottomed out near 1.6, well below replacement level (Statistics Sweden (SCB) 2026). A partial recovery during the late 1970s was followed by a remarkable rebound in the 1980s. TFR climbed from approximately 1.6 in 1978 to 2.1 in 1990, approaching the replacement threshold for the first time since the early 1960s. J. M. Hoem (1993) attributed much of this boom to the "speed premium" embedded in Swedish parental leave legislation: the income replacement rule in the Parental Insurance Act created strong incentives to space births within 24–30 months, front-loading fertility among women who were already in employment.

The 1990s brought Sweden's deepest postwar recession and a correspondingly sharp fertility decline (Section 2.3.2). TFR fell from 2.1 in 1990 to an all-time postwar low of approximately 1.5 by the late 1990s (Statistics Sweden (SCB) 2026). Recovery followed through the 2000s, with TFR rising back toward 1.9–2.0 by the late 2000s. From around 2010, however, a new and sustained decline set in, bringing TFR to approximately 1.7 in 2019 and to an all-time record low of 1.43 in 2024 (Statistics Sweden (SCB) 2026; Ohlsson-Wijk and Andersson 2025). This most recent decline is the focus of Case Study III.

Across all these cycles, the underlying demographic driver has shifted. The 1960s–70s decline was primarily a contraceptive and timing story. The 1990s decline was an economic shock superimposed on ongoing postponement. The post-2010 decline appears to reflect a more fundamental change in first-birth intentions that cuts across socio-economic groups, as discussed in Section 2.3.3.

2.2.2 Nordic comparative context

Compared with most of continental, Southern, and Eastern Europe, the Nordic countries have historically maintained relatively high period fertility, a pattern commonly attributed to the institutional features of the Nordic social-democratic welfare regime (Esping-Andersen 1990). Universal, subsidised childcare, generous parental leave with high earnings replacement, active labour market policies and strong norms of female labour-force participation all reduce the opportunity costs of childbearing and are associated with higher fertility (Rindfuss et al. 2010; Gauthier 2007). Andersson (2008) provides a comparative overview of childbearing trends across Europe and documents the distinctively high and stable fertility of the Nordic region relative to the EU average.

All five Nordic countries recently hit historic fertility lows simultaneously: Sweden 1.43, Norway 1.44, Denmark 1.47, Iceland 1.56, and Finland 1.25 in 2024, with declines since 2010 ranging from approximately 22% (Denmark) to 33% (Finland) (Nordic Statistics 2024). The synchronicity of these declines across diverse institutional environments within the Nordic region is itself a puzzle that has attracted considerable research attention, as discussed in Section 2.3.3.

2.2.3 Postponement of childbearing

Running through all phases of the Swedish fertility cycle is a secular trend toward later childbearing. Mean age at first birth rose from roughly 24 years in the mid-1970s to approximately 30 years by the late 2010s (Statistics Sweden (SCB) 2026). This postponement transition is a common feature of low-fertility societies and has been theorised within the framework of the *Second Demographic Transition* (SDT), a concept introduced by Lesthaeghe and van de Kaa (1986) and further developed by van de Kaa (1987). The SDT posits that declining fertility and postponement in post-industrial societies reflect a fundamental shift in values toward individual autonomy, self-realisation, and unstable partnerships, changes facilitated by the spread of effective contraception. Lesthaeghe (2010) provides an updated account of the SDT's spread.

Postponement has direct implications for the measurement of fertility. As formalised by Ryder (1964), a rising mean age at childbearing depresses period TFR below cohort TFR. Empirically, Sobotka (2004) showed that the "lowest-low fertility" recorded in Southern and Eastern Europe during the 1990s (TFR ≤ 1.3) was in large part a tempo artefact: tempo-adjusted TFRs for these countries were typically 0.3–0.4 points higher than observed TFRs. Whether recuperation, the compensating rush of previously delayed births, fully offsets postponement is debated; Billari and Kohler (2004) show that recuperation is typically partial, so that sustained postponement can translate into a permanent quantum decline.

2.3 Events potentially affecting fertility

The three case studies in this thesis are organised around historical episodes that each represent a distinct type of shock to the fertility process: a technological and institutional shift (oral contraceptives), a severe macroeconomic crisis (Sweden in the 1990s), and a structural demographic shift of uncertain origin (the post-2010 decline). This section reviews the historical and empirical background for each.

2.3.1 Diffusion of oral contraceptives (1960s)

The emergence of effective hormonal contraception was among the most consequential innovations in twentieth-century reproductive behaviour. The first oral contraceptive, Enovid (a combination of mestranol and norethynodrel manufactured by G.D. Searle), received FDA approval in the United States on 23 June 1960 for contraceptive use, having already been approved in 1957 for gynaecological indications (Junod and Marks 2002). By 1965, approximately 26% of married American women under age 30 were using the pill, and the social consequences were profound: reduced unintended pregnancies, greater separation of sexual activity from reproduction, and a shift in contraceptive decision-making toward women (Westoff and Ryder 1977; Goldin and Katz 2002). Diffusion across Western Europe followed rapidly, with broadly similar demographic outcomes (van de Kaa 1987).

In Sweden, the groundwork for rapid diffusion was already in place. Contraception had been legal since 1938, compulsory sex education in schools had been established since 1955, and the Riksförbundet för sexuell upplysning (RFSU), founded in 1933 by Elise Ottesen-Jensen and colleagues (RFSU 2026), had built a nationwide infrastructure for reproductive health information. Oral contraceptives were approved by the Swedish Board of Health and Welfare (Socialstyrelsen) for contraceptive use from August 1964 and became available by prescription through the public health system from 1965 (Oláh and Bernhardt 2008). Diffusion accelerated markedly through the late 1960s as prescribing physicians became more comfortable with the new method, as the price fell, and as cultural acceptance broadened.

The contraceptive transition was embedded in a broader set of institutional reforms that together constituted a structural shift in Swedish gender relations and family law. Abortion was substantially liberalised step by step, culminating in the Abortion Act (SFS 1974:595), which from 1 January 1975 gave women an unconditional right to terminate a pregnancy up to the end of the 18th week. The combination of near-universal access to reliable contraception and broad access to legal abortion fundamentally altered the relationship between sexual activity and reproductive outcomes, contributing to what the SDT framework characterises as the "contraceptive revolution" (Lesthaeghe 2010).

The demographic consequence for Sweden was a pronounced and sustained TFR decline. From a peak of approximately 2.47 in 1964, TFR fell to approximately 2.07 by 1968 and to approximately 1.68 by 1976 (Human Fertility Database 2026). This decline is the basis for Case Study I, with the reference date set at 1965 to capture the first year of broad availability. The central empirical question is when this decline became *statistically detectable* across different fertility indicators, a question that is difficult to answer from historical narrative alone and that motivates the sequential monitoring approach of this thesis.

Research has also connected the spread of the pill to broader economic outcomes. Goldin and Katz (2002) show that earlier legal access to the pill in US states increased women's professional educational investments and career formation, consistent with the view that effective contraception enabled life-course planning on a new timescale. Bailey (2006) further documents that pill access before age 21 significantly reduced the likelihood of a first birth before age 22, confirming the timing-control mechanism that is directly relevant for understanding the Swedish TFR decline during the late 1960s.

2.3.2 Swedish economic crisis of the 1990s

The Swedish financial and macroeconomic crisis of the early 1990s is one of the most severe economic disruptions experienced by any OECD country in the postwar period and produced what was, at the time, the most rapid fertility decline ever recorded in Sweden.

The macroeconomic shock. Following three decades of tight financial regulation, the Swedish credit market was substantially deregulated between 1983 and 1986. Most consequentially, quantitative lending caps were lifted in November 1985 in what was popularly called the "November Revolution". The removal of credit constraints, combined with high nominal interest rates that were partially tax-deductible, created powerful incentives for borrowing. Total bank loans outstanding rose by approximately 136% between 1986 and 1990, roughly 20% per year in nominal terms, and commercial property prices almost doubled (Englund 1999; Jonung et al. 2009). Open unemployment hit a postwar low of approximately 1.4% in 1989.

The correction was abrupt and severe. As asset prices fell, banks' collateral values collapsed. Nordbanken required a government capital injection in December 1991, and Gota Bank was declared insolvent in September 1992. The Riksbank's defence of the fixed ECU peg reached its climax on 16 September 1992, when the marginal lending rate was raised to 500% overnight in a failed attempt to stem speculative outflows (Sveriges Riksbank 1992). On 19 November 1992, Sweden abandoned the fixed exchange rate and the krona depreciated by approximately 23% against the ECU (Sveriges Riksbank 1992). The government announced a blanket guarantee of all bank obligations on 24 September 1992. GDP fell by approximately 5% in cumulative terms over 1991–1993, and open unemployment rose from 1.4% to approximately 8% (Englund 1999; Jonung et al. 2009). Approximately 500,000 jobs were lost in five years.

Fertility response. Swedish TFR fell from 2.13 in 1990 to 1.50 by 1997–1999, a decline of nearly 30% in under a decade (Statistics Sweden (SCB) 2026). B. Hoem (2000) documents that the decline was driven primarily by falling first-birth rates, that women in unemployment or receiving unemployment benefits had substantially lower first-birth propensities, and that the share of Swedish women with labour-market interruptions rose sharply during the crisis. The effect was concentrated in the first years of the crisis: B. Hoem (2000) estimates that declining conceptions (as opposed to rising abortions) accounted for approximately 23% of the birth reduction in the period September 1991 to March 1997.

Andersson (2000) showed that Swedish fertility is strongly pro-cyclical and income-contingent: women with stable employment and higher incomes had higher first- and second-birth rates. The crisis reduced both the level and the stability of income for a large share of the workforce, doubly reducing fertility propensity. An additional mechanism identified by J. M. Hoem (1993) is the "speed premium" from parental leave policy:

the 1980s boom had front-loaded births among employed women exploiting the eligibility window, meaning that this cohort had fewer remaining births when the crisis hit, amplifying the observed TFR decline.

Viewed from a comparative perspective, the Swedish case confirms the general procyclicality of fertility documented across OECD countries by Sobotka, Skirbekk, et al. (2011). That said, fertility responses to economic crises are highly heterogeneous across contexts. Comolli et al. (2021) show that across the Nordic countries, the fertility consequences of the 1990s recessions and the 2008–09 global financial crisis differed substantially, with some countries recovering more quickly than others. This heterogeneity suggests that institutional buffers, the breadth of welfare state coverage, the speed of labour-market adjustment, and the generosity of family policies, shape the magnitude and duration of the fertility response. In the Swedish case, the depth of the 1990s crisis apparently exceeded the buffering capacity of the Nordic welfare state, producing the steepest fertility decline in the country's modern demographic history.

2.3.3 Fertility decline since 2010

The sustained decline in Swedish fertility that began around 2010 differs in important ways from both the 1960s contraceptive transition and the 1990s economic crisis. Unlike the former, it does not map onto a single identifiable institutional change. Unlike the latter, it began in a period of relative macroeconomic prosperity: Swedish GDP per capita was at a postwar high in 2010 and unemployment was well below crisis levels. Understanding its causes has become one of the central questions in Nordic demography.

Swedish TFR stood at approximately 1.98 in 2010 and declined to 1.70 by 2019 and to 1.43 by 2024, the latter an all-time postwar record low and a decline of approximately 28% in 14 years (Statistics Sweden (SCB) 2026; Ohlsson-Wijk and Andersson 2025). The decline was not unique to Sweden: all five Nordic countries reached historic lows simultaneously by 2023–24 (Nordic Statistics 2024), and similar patterns were observed across much of Western and Northern Europe, suggesting broad structural forces rather than country-specific policy changes.

First births and cohort fertility. The most important empirical finding is that the post-2010 decline is overwhelmingly driven by falling *first-birth* rates. Ohlsson-Wijk and Andersson (2022) show that among Swedish women the decline in first births was remarkably uniform across socio-demographic groups: women differing in education, income, immigration background, and region of residence all showed similar declines, pointing toward a broad shift in childbearing intentions rather than deteriorating conditions for a particular subgroup. Strikingly, the decline was not concentrated among women in their late teens or early twenties (where it might be explained by postponement) but was visible across all ages up to the mid-thirties.

Hellstrand et al. (2021) provide compelling evidence that the post-2010 decline is not simply a tempo story. Analysing parity-progression ratios and cohort data across all five Nordic countries, they show that completed cohort fertility has also begun to decline for cohorts born in the late 1980s and 1990s. Declining first-birth intensities explain between 57% (Iceland) and 91% (Denmark) of the total period fertility decline. This implies that the recent fertility loss cannot be recovered through recuperation of delayed births alone: some of it represents a permanent downward shift in lifetime family size.

Possible explanations. Several non-mutually-exclusive explanations have been proposed.

Economic uncertainty. Although aggregate unemployment was low, Matysiak et al. (2021) show that perceived insecurity and uncertainty about future economic conditions, as opposed to objective unemployment, is more predictive of fertility in Nordic countries than in Southern Europe. Neyer et al. (2022) point to declining institutional trust and a "dire outlook on the future" among the childless as a key factor in Sweden specifically. Comolli et al. (2021) found that Nordic fertility declines after 2010 were not closely associated with standard recession indicators, supporting the view that subjective uncertainty matters more than objective economic conditions.

Partnership dynamics. Cantalini et al. (2024) find that while rates of cohabiting union formation in Sweden did not fall after 2010, cohabiting women became increasingly hesitant to make further long-term family commitments, possibly reflecting greater ambivalence about partnership stability.

Housing costs. Rising housing prices in Swedish cities have been proposed as a barrier to family formation, particularly for younger adults without parental capital transfers. Ohlsson-Wijk and Andersson (2022) find no significant urban–rural gradient in the fertility decline, which argues against housing costs as the primary driver, though they may act as one of several amplifying factors.

COVID-19 and aftermath. The pandemic created a brief and modest fertility uptick in 2021 that Ohlsson-Wijk and Andersson (2025) characterise as preponement of already-intended births, followed by an unexpectedly sharp drop from 2022 onward that drove TFR to the 2024 record low. Bujard and Andersson (2024) find no association between the 2022 Swedish fertility drop and local unemployment or infection rates, suggesting the post-pandemic decline reflects a continuation and intensification of the pre-existing structural decline rather than a COVID-specific mechanism.

The post-2010 decline is the focus of Case Study III. A central substantive question is whether CUSUM monitoring of tempo-adjusted indicators detects the decline, which would point toward genuine quantum change, or whether only unadjusted indicators signal, which would be in line with a predominantly postponement-driven story. The results bear on the interpretation of whether Sweden faces a temporary timing adjustment or a more fundamental shift in completed family size.

2.4 Structural breaks in time series

2.4.1 Trend breaks and regime changes

Structural break analysis provides a statistical framework for identifying points in time at which the parameters governing a time series change discontinuously (Bai and Perron 1998; Bai and Perron 2003). These changes may reflect policy reforms, economic shocks, or gradual behavioural shifts that eventually cross a threshold visible in the data. In the fertility context, a structural break in the TFR trend may correspond to the onset of a contraceptive revolution, a recession, or a change in family-formation norms.

Formally, consider the linear regression model

$$y_t = X_t'\beta + u_t, \qquad t = 1, \ldots, T \tag{5}$$

where y_t is the fertility indicator of interest, X_t contains deterministic regressors (a constant, a time trend, and possibly seasonal dummies), β is a vector of coefficients, and u_t is a disturbance term. Following Bai and Perron (1998), if the data-generating process shifts at unknown dates $T_1 < T_2 < \cdots < T_m$, the model generalises to

$$y_t = X_t'\beta_j + u_t, \qquad t = T_{j-1} + 1, \ldots, T_j, \quad j = 1, \ldots, m + 1 \tag{6}$$

where the coefficient vector β_j may differ across the $m+1$ regimes and $T_0 = 0$, $T_{m+1} = T$.

In the present analysis, the parameters of primary interest are the *slope* coefficients (trend parameters): a break in slope corresponds to a change in the rate of change of the fertility indicator, which is exactly the type of change expected when a new demographic regime sets in gradually rather than instantaneously.

2.4.2 The Bai–Perron framework for multiple breaks

The standard approach to multiple structural break estimation is the framework developed by Bai and Perron (1998) and refined in Bai and Perron (2003). The main innovation relative to the classical Chow test (Chow 1960), which requires the break date to be known in advance, is that the Bai–Perron procedure treats break dates as unknown parameters to be estimated jointly with the regression coefficients.

Break dates are estimated by minimising the global residual sum of squares:

$$\min_{T_1,\ldots,T_m} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}+1}^{T_j} \left(y_t - X_t'\hat{\beta}_j\right)^2 \tag{7}$$

subject to a minimum segment length $\lfloor \varepsilon T \rfloor$ for some $\varepsilon > 0$ that prevents break dates from being placed too close together. Bai and Perron (1998) derive the asymptotic distribution of the estimated break dates and propose a sequence of tests, the double-maximum (UDmax) test for the presence of any break and the sequential ℓ versus $\ell + 1$ test for the number of breaks, all with known limiting distributions under the null of no structural change.

Bai and Perron (2003) showed that this global minimisation problem can be solved efficiently using dynamic programming in $O(T^2)$ operations, a result later strengthened by Killick et al. (2012), whose Pruned Exact Linear Time (PELT) algorithm achieves $O(T)$ expected cost under mild conditions on the penalty structure. The PELT algorithm, implemented in the Python package `ruptures`, is used in the present analysis.

Because the Bai–Perron procedure does not use information about reference dates, breaks detected close to the case-study reference dates constitute independent, data-driven confirmation of a structural change. This complementary role, the offline retrospective method validating what the online sequential method detects or misses, motivates the cross-validation in Section 3.9.

2.4.3 Slope change test and Chow test

Two additional offline methods are used for cross-validation. The slope change test estimates a piecewise linear regression with a known break at the reference date τ:

$$y_t = \beta_0 + \beta_1 t + \beta_2 \max(0, t - \tau) + \varepsilon_t \tag{8}$$

where β_2 measures the change in slope at τ. The null hypothesis H_0: $\beta_2 = 0$ is tested using OLS with HC3 heteroskedasticity-robust standard errors (Davidson and MacKinnon 2004). This test is informative about the magnitude of the slope change and allows visual comparison of pre- and post-break trends with confidence bands.

The Chow test (Chow 1960) is an F-test for equality of the full parameter vector (intercept *and* slope) across the two subsamples defined by τ:

$$F = \frac{(\mathrm{RSS}_R - \mathrm{RSS}_U)/k}{\mathrm{RSS}_U/(T - 2k)}, \qquad k = 2 \tag{9}$$

compared against the $F(2, T - 4)$ critical value. The Chow test is appropriate when the break date is known in advance; because the reference dates used here are defined by historical events rather than estimated from the data, the fixed-break-date assumption is justified.

2.4.4 Limitations of offline methods

A fundamental limitation shared by all offline structural break methods is that they use the *full sample* and therefore cannot, by definition, be used for prospective monitoring. They represent what a well-equipped analyst can conclude in retrospect, not what was statistically detectable sequentially at the time the change occurred. The gap between offline detectability and sequential detectability is precisely what the CUSUM analysis quantifies. It should be noted that the present analysis applies CUSUM to fixed windows centred on known reference dates, and therefore simulates rather than implements a fully prospective monitoring system. A fertility decline that is statistically confirmed in hindsight may have been indistinguishable from normal variation for several years after the true break date, a finding with direct implications for demographic surveillance and timely policy response.

2.5 Sequential change detection

2.5.1 The change detection problem

Sequential change detection addresses the problem of identifying a change in the distribution of an observed process *as early as possible* while controlling the probability of raising a false alarm. This is fundamentally different from the offline structural break problem: data arrive one observation at a time, and at each step the analyst must decide whether to issue an alarm or to continue collecting data. There is an irreducible trade-off between sensitivity (raising alarms quickly when a change occurs) and specificity (not raising alarms when no change has occurred).

Formally, let $\{y_t\}_{t=1}^{\infty}$ be a sequence of observations. Before a change point τ^*, observations follow distribution F_0 with parameter θ_0; after τ^* they follow F_1 with parameter $\theta_1 \neq \theta_0$. The sequential test can be framed as:

$$
\begin{aligned}
&H_0\colon \theta = \theta_0 &&\text{(no change has occurred)}\\
&H_1\colon \theta = \theta_1 &&\text{(change occurred at some unknown } \tau^* \leq t\text{)}
\end{aligned}
$$

A stopping rule T^* signals an alarm at the first time the accumulated evidence exceeds a threshold. Letting $E(\cdot)$ denote the expectation operator, two performance criteria are standard: (i) the *Average Run Length under the null* $\mathrm{ARL}_0 = E(T^* \mid H_0)$, measuring how long the procedure runs before a false alarm; and (ii) the *expected detection delay* $E(T^* - \tau^* \mid \tau^*, H_1)$, measuring how quickly a true change is detected (Basseville and Nikiforov 1993; Hawkins and Olwell 1998).

2.5.2 The CUSUM statistic and its optimality

The CUSUM procedure was introduced by Page (1954) and has since become the dominant method for sequential change detection in industrial quality control, epidemiological surveillance, and econometrics. It is designed to detect a shift in the mean of a sequence of independent observations, but can be extended to autocorrelated and non-Gaussian settings.

The one-sided upper CUSUM statistic accumulates positive deviations from a reference value k:

$$C_t^+ = \max\bigl(0,\ C_{t-1}^+ + z_t - k\bigr), \qquad C_0^+ = 0 \tag{10}$$

where $z_t = (y_t - \mu_0)/\sigma_0$ is the standardised observation under the in-control distribution, with μ_0 and σ_0 denoting the mean and standard deviation of the process before any change, and $k > 0$ is the *allowance* or *reference value*. An alarm is declared at the first time $C_t^+ > h$, where $h > 0$ is the *decision interval* (threshold). The symmetric lower-arm statistic C_t^- detects downward shifts:

$$C_t^- = \max\bigl(0,\ C_{t-1}^- - z_t - k\bigr) \tag{11}$$

A two-sided alarm is declared when either $C_t^+ > h$ or $C_t^- > h$.

The reference value k is conventionally set to half the shift size one wishes to detect efficiently, so that $k = \delta/2$ for a target shift of δ standard deviations. For the common choice of detecting a one-standard-deviation shift, $k = 0.5$, which is the value used throughout this thesis (Hawkins and Olwell 1998; Montgomery 2009).

The optimality of CUSUM is established by two landmark results. Lorden (1971) showed that CUSUM minimises the worst-case expected delay to detection subject to a constraint on the ARL under the null. Moustakides (1986) subsequently proved that this minimax optimality holds *exactly* (not just asymptotically) for i.i.d. observations in exponential families: no sequential procedure can achieve a lower expected delay for the same false alarm rate. These results justify CUSUM as the natural benchmark for sequential fertility monitoring.

In the econometric literature, CUSUM was introduced to regression residuals by Brown et al. (1975), who proposed applying the statistic to the sequence of recursive residuals from a rolling OLS estimation. Chu et al. (1996) extended this approach to a formal monitoring framework with controlled asymptotic size and demonstrated that one-shot end-of-sample tests cannot be repeatedly applied without eventually producing spurious alarms with probability one. The monitoring framework of Chu et al. (1996) is the direct precedent for the approach adopted in this thesis.

2.5.3 Detecting changes in trends

The standard CUSUM procedure is designed to detect shifts in the *mean* of a stationary series. Fertility indicators, however, typically exhibit deterministic trends: TFR or ASFRs may rise or fall gradually over time, and the change of interest is a change in the *slope* of this trend rather than a level shift.

A slope change of magnitude Δb in the original series $y_t = a + bt + \varepsilon_t$ produces a linearly growing divergence between the post-break series and the projected pre-break trend. In the residual series $e_t = y_t - \hat{y}_t$ (where $\hat{y}_t$ is extrapolated from the reference period), a slope break generates a gradually accumulating drift rather than an immediate level shift. To convert this gradual drift into the mean-shift format that CUSUM detects efficiently, a first-difference transformation is applied:

$$z_t = \Delta e_t = e_t - e_{t-1} \tag{12}$$

After differencing, a constant slope change Δb in the original series translates into an approximate constant mean shift Δb in z_t, which is exactly the signal CUSUM is designed to accumulate (Basseville and Nikiforov 1993). This differencing approach mirrors the "score-based" monitoring statistics proposed by Zeileis (2006) and is related to the fluctuation tests of Chu et al. (1996).

The baseline model fitted to the reference period includes, for monthly data, eleven monthly dummy variables to capture within-year seasonality, in addition to the linear trend. For annual data, a pure linear trend is used. The baseline is extrapolated to the monitoring window, and residuals are computed as the difference between observed and projected values. An AR(1) model is then fitted to the reference-period residuals to whiten the series before differencing and CUSUM application (see below).

2.5.4 Handling autocorrelation and threshold calibration

A critical practical issue is that fertility time series exhibit substantial autocorrelation. Standard CUSUM theory assumes i.i.d. observations; under positive autocorrelation, the statistic accumulates more rapidly than expected under the null, inflating the false alarm rate. Johnson and Bagshaw (1974) showed that serial correlation in the AR(1) range can reduce the in-control ARL by an order of magnitude relative to the nominal value. Schmid (1997) derived modified CUSUM procedures for stationary Gaussian processes, but these require the full autocorrelation structure to be known.

This thesis addresses autocorrelation in two steps. First, an AR(1) model is fitted to the reference-period residuals and the estimated AR(1) innovations $\hat{\eta}_t$ are used as the input to the differencing step, effectively pre-whitening the monitoring series. Second, because the pre-whitened series may still retain some dependence and because the AR(1) approximation may be imperfect, the decision threshold h is calibrated by Monte Carlo simulation under H_0. Specifically, 1,000 synthetic series are generated from the estimated reference-period AR(1) model, with the same length as the monitoring window, the full CUSUM pipeline is applied to each, and h is set to the 95th percentile of the resulting distribution of $\max_t C_t^+ \vee C_t^-$. This yields a false alarm rate of $\alpha = 0.05$ for each case study, with the threshold h calibrated specifically to that case study's estimated parameters.

The simulation approach is related to the block bootstrap of Künsch (1989) and the stationary bootstrap of Politis and Romano (1994), both of which are designed to preserve the dependence structure of stationary time series when constructing confidence intervals or critical values. The present implementation uses parametric simulation from the fitted AR(1) model rather than resampling, which is computationally simpler and avoids the block-length selection problem.

2.5.5 Detection delay and the minimum detectable change

Detection delay $D = t_a - \tau^*$ is the primary performance measure reported for each case study. A delay of zero means CUSUM raised an alarm in the same period as the reference date; positive delays measure the number of years between the reference date and the alarm.

An equally informative quantity is the *Minimum Detectable Change* (MDC): the slope change magnitude that the CUSUM is configured to detect efficiently given the estimated reference-period noise level. Following the design convention $k = \delta/2$ introduced in Section 2.5.2, a chart with reference value k is tuned for a standardised mean shift of $\delta = 2k$. Because a slope change Δb in the original series produces an approximately equal mean shift in the differenced monitoring series (Section 2.5.3), translating this design shift back into the units of that series gives MDC $= 2k\,\sigma_d$, where σ_d is the standard deviation of the differenced reference-period series (Hawkins and Olwell 1998; Montgomery 2009). The MDC therefore measures the smallest trend change the procedure is set up to register efficiently rather than a guaranteed power level; the realised power at a given shift also depends on the calibrated threshold h and the window length, as illustrated by the power analysis in Section 3.7. A high MDC indicates that only large slope changes can be reliably detected with the available data, while a low MDC signals that the procedure is sensitive to subtle trend changes. MDC values differ across indicators primarily because of differences in the signal variance of the residual series: noisy series require larger true shifts before these shifts emerge above the background variation.

2.5.6 CUSUM in demographic surveillance

The application of CUSUM to fertility indicators appears to be novel. The closest precedents are in public health and epidemiological surveillance, where CUSUM-based methods have been applied to the detection of disease outbreaks and unusual mortality patterns. Sonesson and Bock (2003) provide a review of prospective statistical surveillance methods in public health, including Poisson-adapted CUSUM for count data. Woodall (2006) surveys applications of control charts to health-care outcomes, and Steiner et al. (2000) demonstrates CUSUM monitoring of surgical complication rates.

The broader rationale for applying CUSUM to demographic indicators is grounded in the distinction between *ex post* explanation and sequential monitoring. Much existing demographic research identifies fertility change retrospectively, using data that were not available at the time the change began. A sequential monitoring framework asks a different question: at what point would a demographer observing incoming data have had sufficient statistical evidence to conclude that a new trend had begun? This question has direct relevance for early warning systems in national statistical offices and for the evaluation of family policies.

3 Analysis: Data, Methods, and Results

3.1 Data and Descriptive Overview

Five data series covering Sweden were used. Monthly live births (1900–2024) were obtained from Statistics Sweden (SCB). The Total Fertility Rate (TFR, 1950–2024) was likewise sourced from SCB, where values are published in per-mille format and were converted to rates per woman prior to analysis. The tempo-adjusted TFR (1971–2023) was downloaded from the Human Fertility Database (HFD). Age-specific fertility rates (ASFR) were available by five-year age group (15–19 through 45–49), published as rates per 1,000 women; the 45–49 group was excluded from the analysis as negligible. The mean age at childbearing (MAB) was obtained from SCB and used to compute the Bongaarts–Feeney tempo adjustment described in Section 2.1. All series were verified against expected value ranges before use.

Figures 1–4 provide a descriptive overview of the series. Monthly births (Figure 1) display the strong within-year seasonal pattern documented in Section 2.1, with a visible long-run decline beginning in the 1990s. The TFR series (Figure 2) traces the roller-coaster trajectory described in Section 2.2, most recently falling to a postwar record low of 1.43 in 2024. Age-specific rates (Figure 3) confirm the well-documented shift toward later childbearing: rates at ages 15–24 have declined while rates at ages 30–44 have risen. Mean age at childbearing (Figure 4) has risen steadily from approximately 27 years in 1970 to over 31 years by 2020, consistent with the postponement transition outlined in Section 2.2.

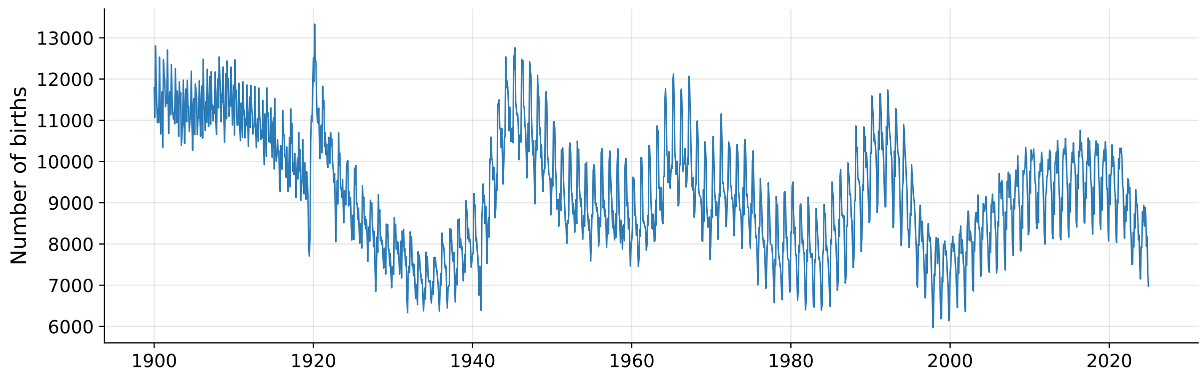

Figure 1: Monthly number of live births in Sweden, 1900–2024.

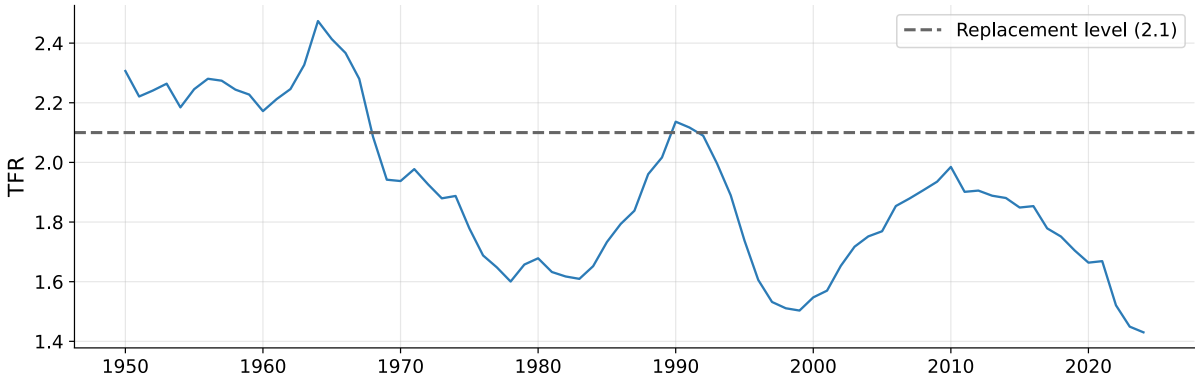

Figure 2: Total Fertility Rate in Sweden, 1950–2024, with replacement level (2.1) indicated.

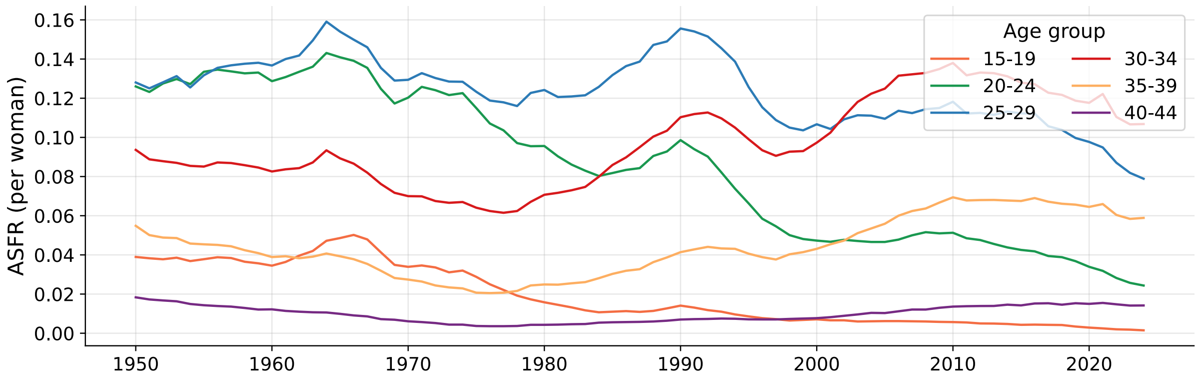

Figure 3: Age-specific fertility rates by age group, 1950–2024.

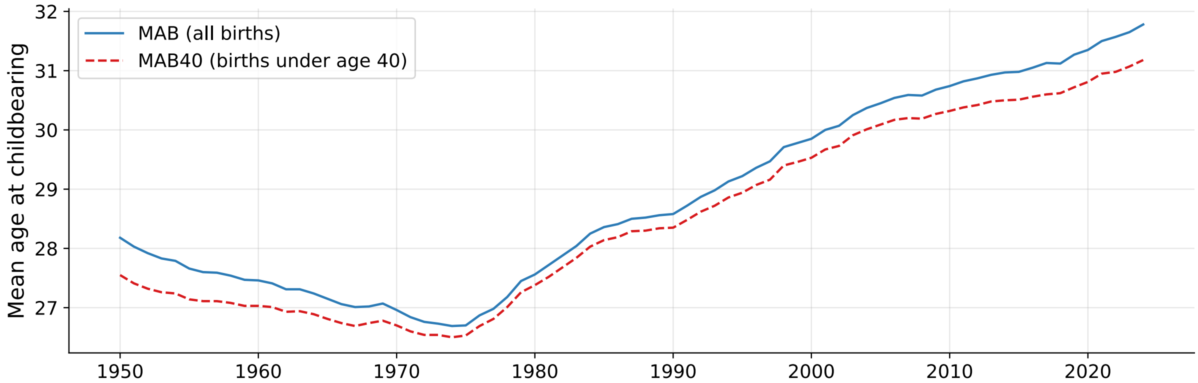

Figure 4: Mean age at childbearing (MAB) and mean age at childbearing under the age of 40 (MAB40) in Sweden, 1950–2024.

3.2 Tempo Adjustment

The HFD tempo-adjusted TFR is only available from 1971 onward, but Case Study I (oral contraceptives, reference date 1965) requires data from around 1950. A self-computed tempo-adjusted TFR was therefore constructed using the Bongaarts–Feeney formula introduced in Section 2.1. The observed TFR was first reconstructed by summing all age-group ASFR values within each year and multiplying by five. The annual change in MAB, $r(t) = \Delta\text{MAB}(t)$, was then capped at $[-0.5, 0.5]$ to prevent extreme corrections, and the adjusted rate computed as $\text{TFR}^*(t) = \text{TFR}(t)/(1 - r(t))$. The same formula was applied age-group by age-group to produce tempo-adjusted ASFR, allowing genuine fertility decline to be distinguished from postponement within each age group. An extended tempo-TFR series was then constructed by splicing the self-computed estimate for years before 1971 with the HFD series from 1971 onward.

Figure 5 confirms that the self-computed estimate and the HFD series track each other closely over their overlapping period, validating the spliced series. As expected from the theory in Section 2.1, the tempo-adjusted series lies consistently above the observed TFR, confirming that postponement has depressed the observed rate. Figure 6 shows the corresponding effect on age-group rates: because the adjustment applies a common factor $1/(1 - r)$, all rates scale upward by the same proportion each year.

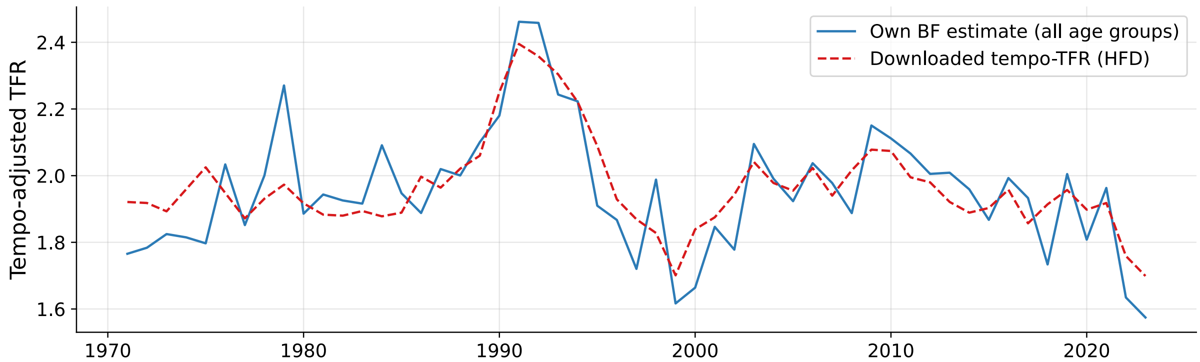

Figure 5: Own Bongaarts–Feeney tempo-adjusted TFR compared with the HFD downloaded series, 1971–2023. The two series track each other closely, supporting the use of the self-computed estimate to extend the series before 1971.

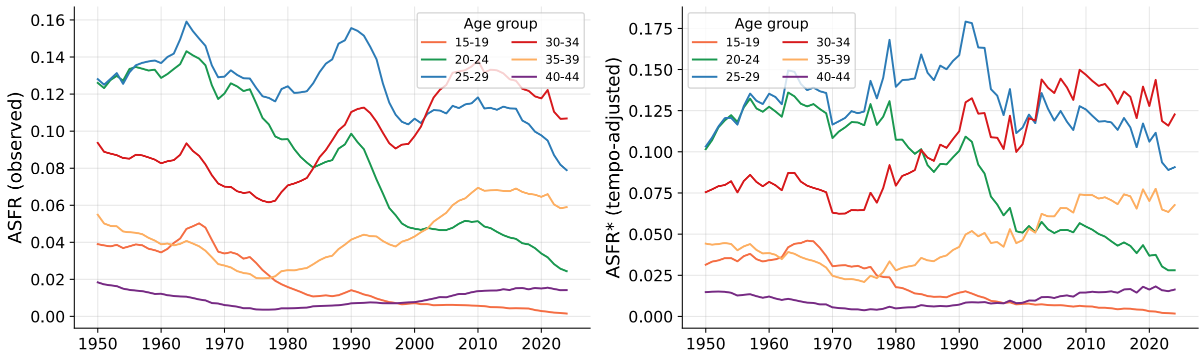

Figure 6: Observed (left) and tempo-adjusted (right) ASFR by age group, 1950–2024. The adjustment multiplies every age group by the same factor $1/(1 - r)$, so the adjusted panel (right) lies above the observed one (left).

3.3 CUSUM Sequential Change Detection

The CUSUM procedure and its theoretical properties are described in Section 2.5. The following pipeline was applied identically to every indicator and case study. For each case study, a symmetric window of ± 15 years around the reference date was extracted, with all observations before the reference date forming the reference period. The window of ± 15 years was chosen to give the reference period enough observations for reliable AR(1) estimation and threshold calibration, while keeping it short enough that the fertility trend within the window can be treated as roughly stable. Sensitivity to this choice is assessed in Appendix B.

A linear baseline was fitted by OLS on the reference period only. For monthly data the baseline included an intercept, a linear trend, and eleven monthly dummy variables (January as reference category); for annual data a pure linear trend was used. The baseline was extrapolated to the full window and residuals $e_t = y_t - \hat{y}_t$ computed. An AR(1) model was then fitted to the reference-period residuals by maximum likelihood, and a Ljung–Box test with $K = \min(10, \lfloor N/3 \rfloor)$ lags applied to the innovations to verify whiteness. The estimated coefficient $\hat{\phi}$ was used to filter the full residual series via $u_t = e_t - \hat{\phi}\, e_{t-1}$, converting autocorrelated residuals to approximate white noise. The filtered series was then first-differenced to convert the slope-break detection problem into a mean-shift problem, and the differences standardised to zero mean and unit variance using reference-period statistics, yielding the monitoring series z_t. The threshold h was calibrated via 1,000 Monte Carlo replications under H_0, set at the 95th percentile of the simulated $\max(C^+, C^-)$ distribution, giving a false alarm rate of $\alpha = 0.05$. The two-sided CUSUM was run with reference value $k = 0.5$ (Basseville and Nikiforov 1993, eq. 2.1.16), and detection delay was recorded as the number of years between the reference date and the first alarm. The complete pseudocode for the CUSUM detection pipeline is given in Appendix A.

3.4 Case Study Results

Three case studies were analysed. The pipeline above was applied to 15 indicators per case study: monthly births, TFR, tempo-adjusted TFR, ASFR for six age groups (15–19, 20–24, 25–29, 30–34, 35–39, 40–44), and tempo-adjusted ASFR for the same six groups: 45 CUSUM runs in total. Complete results are reported in Table 1.

Table 1: Complete CUSUM results for all case studies and indicators. A dash in the Alarm and Delay columns indicates no alarm within the ±15 year window.

Case study	Indicator	$\hat{\phi}$	h	MDC	Alarm	Delay
Oral contraceptives	Monthly births	0.883	3.561	352.0891	–	–
Oral contraceptives	TFR	0.558	3.277	0.0598	1968-01-01	3
Oral contraceptives	Tempo-TFR	0.161	3.277	0.1208	–	–
Oral contraceptives	ASFR 15-19	0.911	3.277	1.7519	1968-01-01	3
Oral contraceptives	ASFR 20-24	0.419	3.277	3.6622	1968-01-01	3
Oral contraceptives	ASFR 25-29	0.411	3.277	4.1527	1968-01-01	3
Oral contraceptives	ASFR 30-34	0.595	3.277	1.8691	1965-01-01	0
Oral contraceptives	ASFR 35-39	0.363	3.277	1.2482	1969-01-01	4
Oral contraceptives	ASFR 40-44	0.507	3.277	0.4345	–	–
Oral contraceptives	Tempo-ASFR 15-19	0.638	3.277	2.5047	–	–
Oral contraceptives	Tempo-ASFR 20-24	0.346	3.277	6.8609	1980-01-01	15
Oral contraceptives	Tempo-ASFR 25-29	0.07	3.277	7.5764	1970-01-01	5
Oral contraceptives	Tempo-ASFR 30-34	0.03	3.277	4.7886	1979-01-01	14
Oral contraceptives	Tempo-ASFR 35-39	0.084	3.277	2.403	1979-01-01	14
Oral contraceptives	Tempo-ASFR 40-44	-0.101	3.277	0.7928	–	–
Economic crisis 1990s	Monthly births	0.92	3.558	331.2667	–	–
Economic crisis 1990s	TFR	0.831	3.277	0.0505	1993-01-01	2
Economic crisis 1990s	Tempo-TFR	0.633	3.277	0.0747	1996-01-01	5
Economic crisis 1990s	ASFR 15-19	0.838	3.277	0.5532	1991-01-01	0
Economic crisis 1990s	ASFR 20-24	0.908	3.277	3.2514	1993-01-01	2
Economic crisis 1990s	ASFR 25-29	0.726	3.277	4.1611	1995-01-01	4
Economic crisis 1990s	ASFR 30-34	0.682	3.277	1.9408	1993-01-01	2
Economic crisis 1990s	ASFR 35-39	0.788	3.277	1.2315	–	–
Economic crisis 1990s	ASFR 40-44	0.109	3.277	0.2361	1996-01-01	5
Economic crisis 1990s	Tempo-ASFR 15-19	0.668	3.277	2.2318	–	–
Economic crisis 1990s	Tempo-ASFR 20-24	0.322	3.277	9.6862	–	–
Economic crisis 1990s	Tempo-ASFR 25-29	-0.158	3.277	12.0637	1997-01-01	6
Economic crisis 1990s	Tempo-ASFR 30-34	-0.304	3.277	6.6091	1995-01-01	4
Economic crisis 1990s	Tempo-ASFR 35-39	-0.026	3.277	2.9289	1997-01-01	6
Economic crisis 1990s	Tempo-ASFR 40-44	-0.046	3.277	0.6468	–	–
Post-2010	Monthly births	0.883	3.561	319.0526	–	–
Post-2010	TFR	0.535	3.252	0.0292	2011-01-01	1
Post-2010	Tempo-TFR	0.381	3.223	0.0885	–	–
Post-2010	ASFR 15-19	0.323	3.252	0.3468	–	–
Post-2010	ASFR 20-24	0.642	3.252	1.1538	2020-01-01	10
Post-2010	ASFR 25-29	0.48	3.252	3.0481	2022-01-01	12
Post-2010	ASFR 30-34	0.553	3.252	2.5039	2011-01-01	1
Post-2010	ASFR 35-39	0.519	3.252	1.2326	2014-01-01	4
Post-2010	ASFR 40-44	0.544	3.252	0.4382	2023-01-01	13
Post-2010	Tempo-ASFR 15-19	-0.234	3.252	0.4774	–	–
Post-2010	Tempo-ASFR 20-24	0.319	3.252	6.5405	–	–
Post-2010	Tempo-ASFR 25-29	-0.13	3.252	12.1283	–	–
Post-2010	Tempo-ASFR 30-34	0.028	3.252	13.8395	–	–
Post-2010	Tempo-ASFR 35-39	-0.324	3.252	5.4175	–	–
Post-2010	Tempo-ASFR 40-44	-0.16	3.252	1.1219	–	–

3.4.1 Case Study I – Oral Contraceptives (1965)

Monthly births produced no alarm within the ±15 year window (Figure 7), consistent with the high noise level of this series at monthly frequency. TFR triggered an alarm in 1968, three years after the reference date (Figure 8), as the CUSUM statistic C^- accumulated steadily once the series departed from its pre-1965 linear trend. The tempo-adjusted TFR produced no alarm (Figure 9). This indicates that once timing distortions are removed, the evidence for a genuine quantum decline does not accumulate sufficiently to cross the threshold. The result aligns with the interpretation that the 1960s TFR decline reflected postponement of childbearing rather than a reduction in completed family size.

Among the age-specific rates, ASFR 30–34 triggered an alarm immediately at the reference date (delay = 0), while ASFR 15–19, 20–24, and 25–29 all signalled with a three-year delay. ASFR 35–39 detected with a four-year delay and ASFR 40–44 produced no alarm, likely stemming from its low level and high relative noise. Tempo-adjusted ASFRs detected considerably later than their observed counterparts, with delays of 5 to 15 years where detection occurred, or not at all.

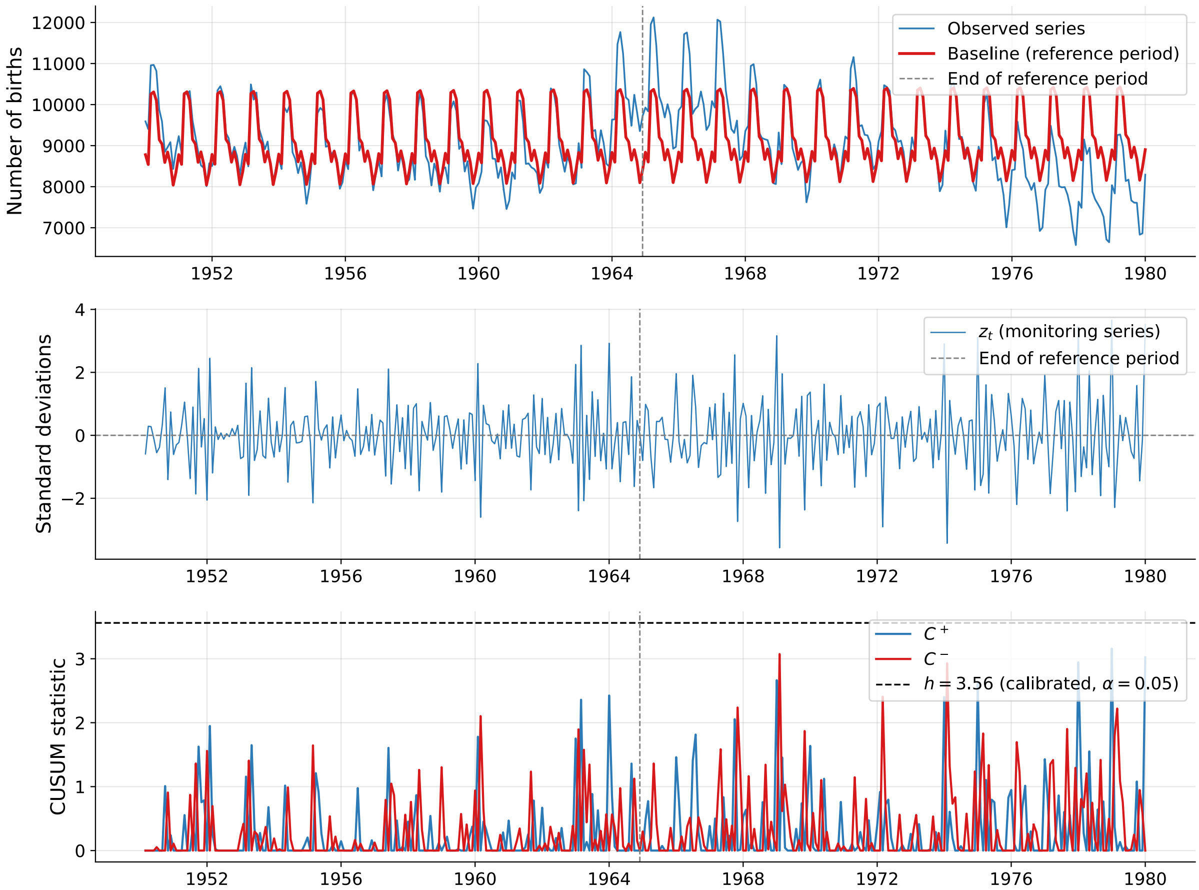

Figure 7: CUSUM on monthly births, oral contraceptives case study. The three panels show the observed series against the reference-period baseline (top), the monitoring series z_t (middle), and the CUSUM statistic against the calibrated threshold h (bottom). Neither arm of the statistic crosses h, so no alarm is raised within the ±15 year window.

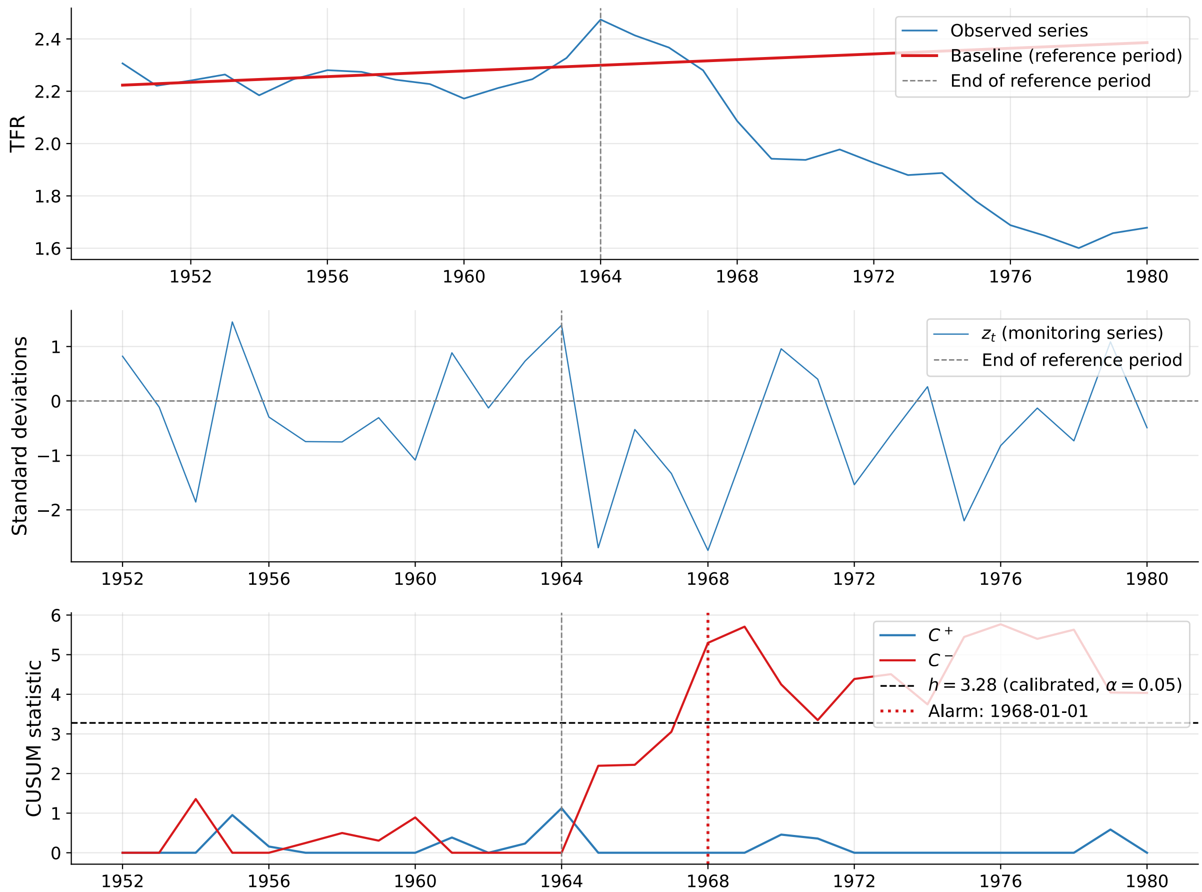

Figure 8: CUSUM on TFR, oral contraceptives case study. The lower-arm statistic C^- crosses the threshold h in 1968, giving an alarm three years after the reference date (delay = 3 years).

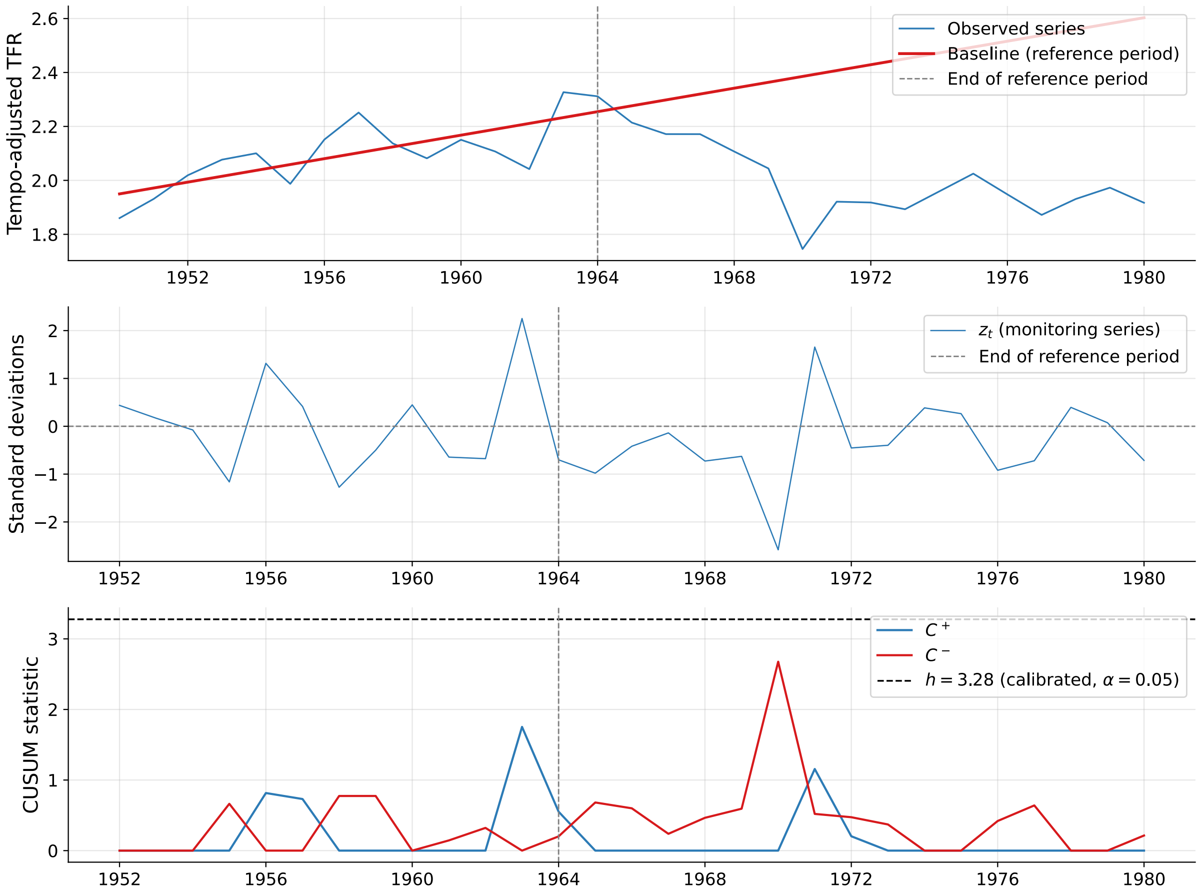

Figure 9: CUSUM on tempo-adjusted TFR, oral contraceptives case study. Neither arm crosses h, so no alarm is raised; once timing effects are removed the decline no longer accumulates enough evidence to detect.

3.4.2 Case Study II – Swedish Economic Crisis (1991)

Monthly births again produced no alarm (Figure 10). TFR detected in 1993, two years after the reference date (Figure 11), with C^- crossing the calibrated threshold as the series registered a sharp and sustained downward departure from its pre-1991 trend. Unlike in Case Study I, the tempo-adjusted TFR also produced an alarm, but with a considerably longer delay of five years (1996), indicating that the 1990s decline reflected a genuine quantum reduction alongside any timing effects.

ASFR 15–19 triggered an alarm immediately in 1991 (delay = 0), while ASFR 20–24 and 30–34 both detected with a two-year delay. ASFR 25–29 and 40–44 detected with four- and five-year delays respectively, and ASFR 35–39 produced no alarm. Tempo-adjusted ASFRs showed mixed results; Figure 12 illustrates the case of ASFR 30–34, where the tempo-adjusted series detected with a four-year delay, pointing to a slower-emerging reduction at prime childbearing ages.

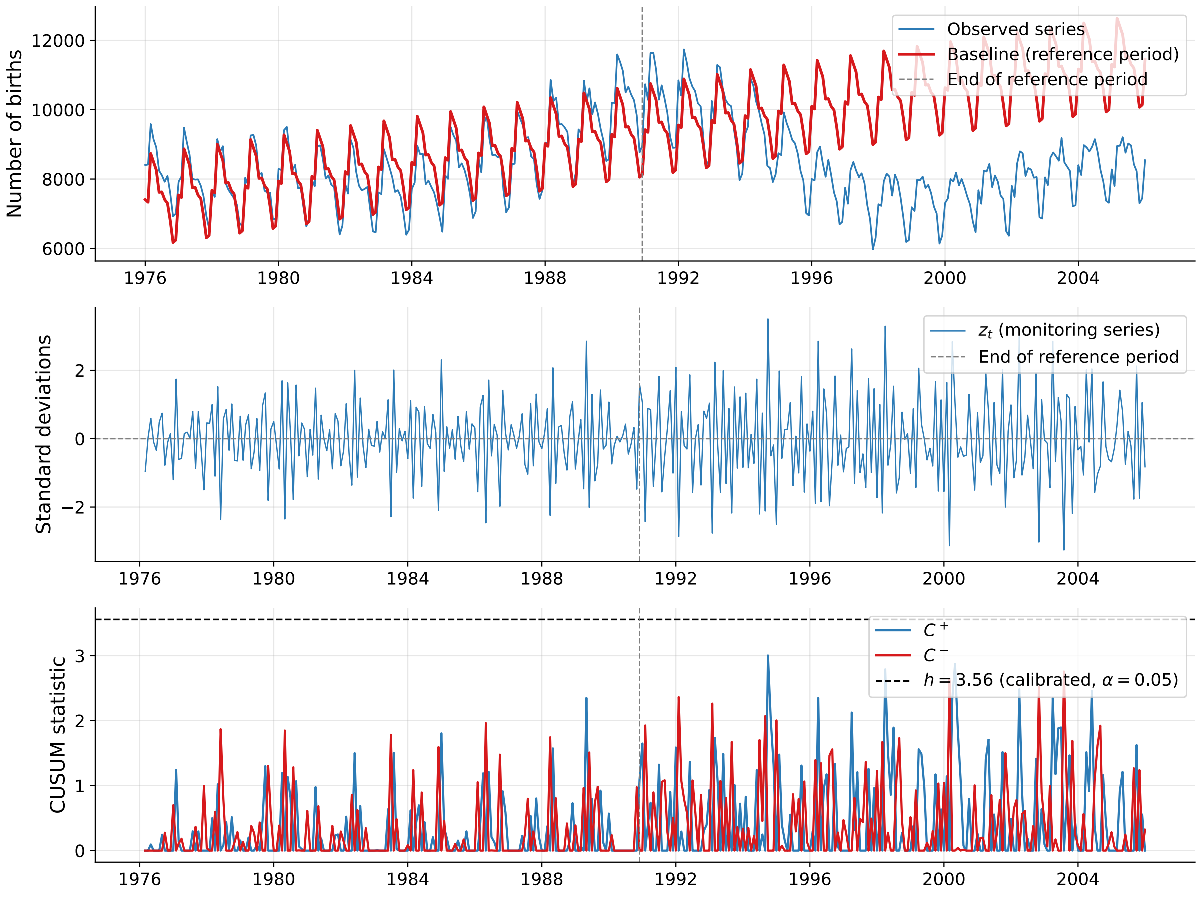

Figure 10: CUSUM on monthly births, economic crisis case study. As in Case Study I, neither arm crosses h and no alarm is raised within the ±15 year window.

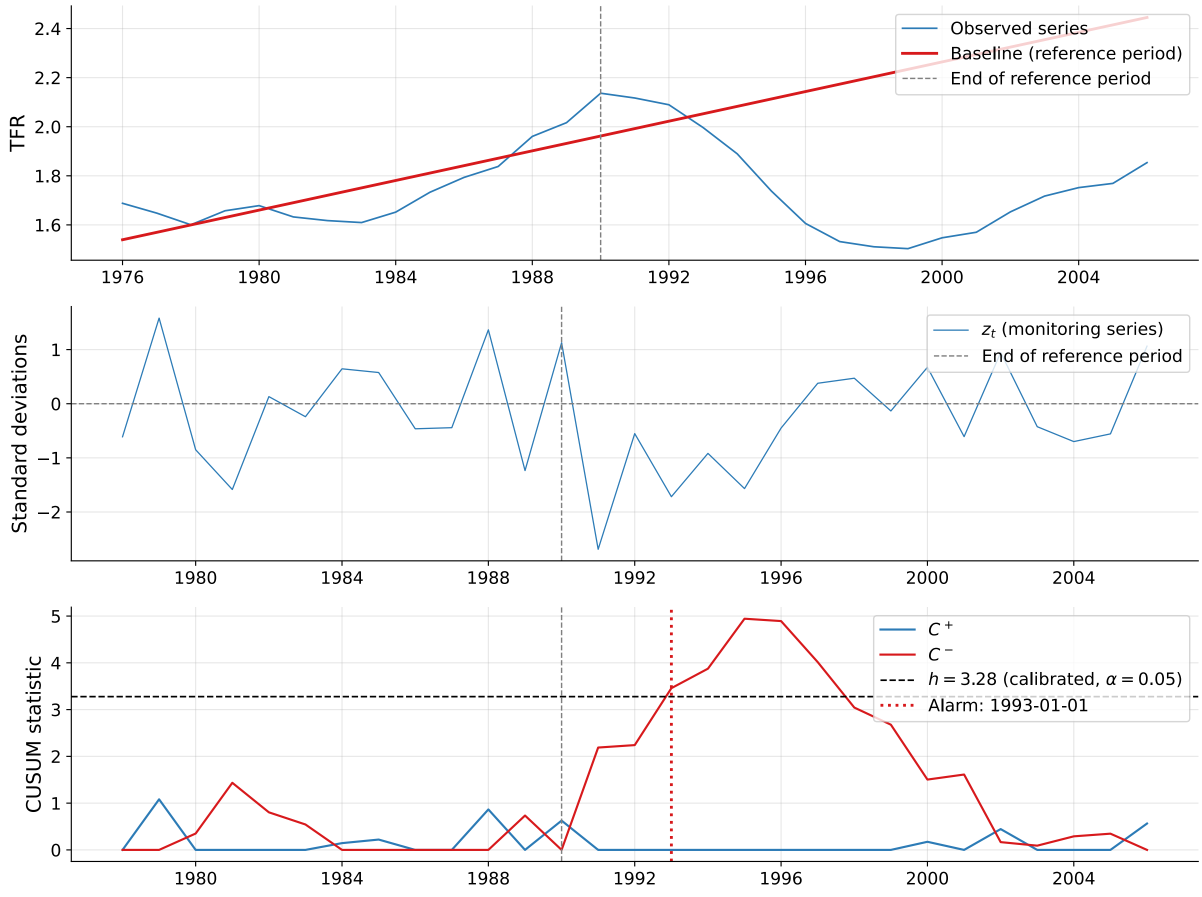

Figure 11: CUSUM on TFR, economic crisis case study. The lower-arm statistic C^{-} crosses h in 1993, two years after the reference date (delay = 2 years).

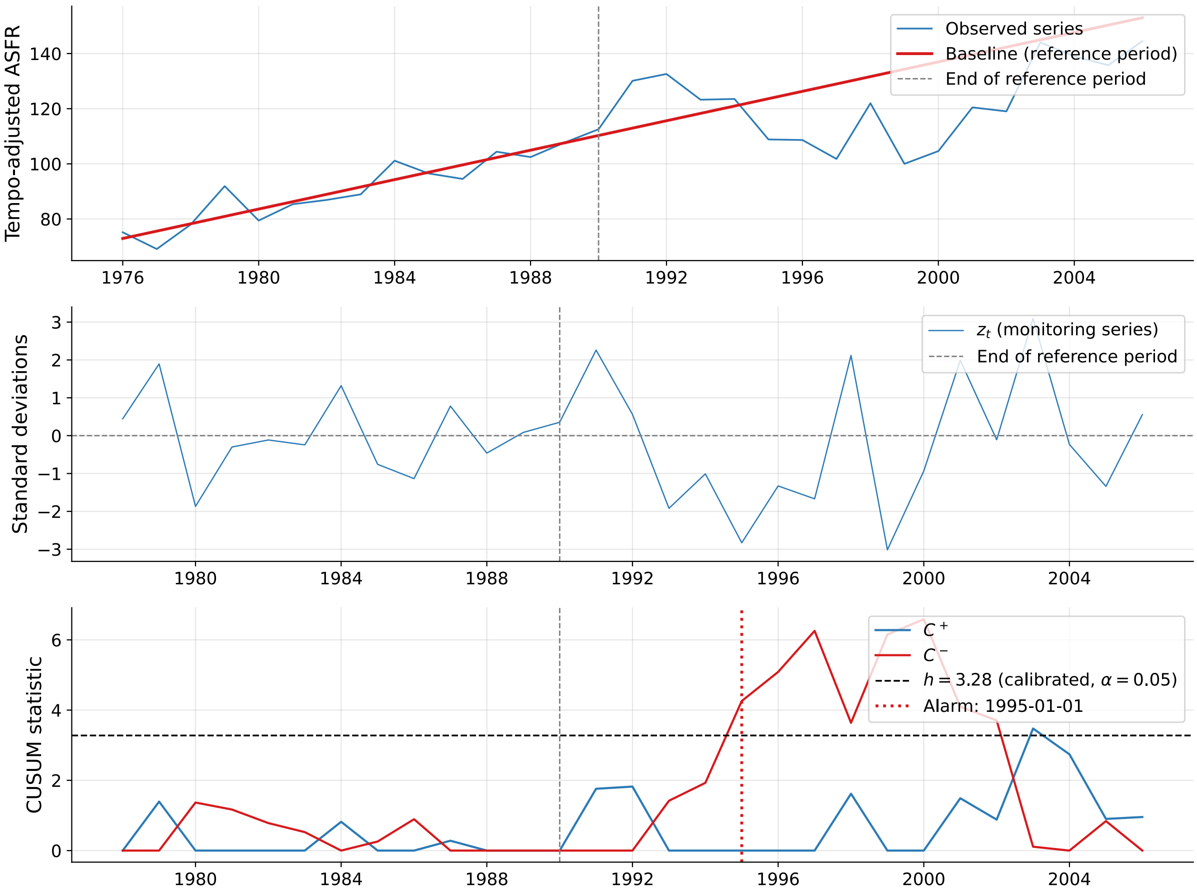

Figure 12: CUSUM on tempo-adjusted ASFR 30–34, economic crisis case study. The statistic crosses h in 1995 (delay = 4 years); unlike in Case Study I, a tempo-adjusted series still detects, pointing to a genuine quantum reduction.

3.4.3 Case Study III – Post-2010 Fertility Decline (2010)

Monthly births produced no alarm (Figure 13). TFR detected rapidly, in 2011, one year after the reference date (Figure 14), making this the shortest TFR detection delay across the three case studies. In sharp contrast, the tempo-adjusted TFR produced no alarm (Figure 15), mirroring the result for Case Study I and pointing toward postponement as a major driver of the observed TFR decline, a finding in line with the evidence from Hellstrand et al. (2021) and Ohlsson-Wijk and Andersson (2022) reviewed in Section 2.3.3.

Among the observed ASFRs, the earliest detection came from ASFR 30–34 (2011, delay = 1; Figure 16), matching TFR. ASFR 35–39 detected in 2014 (delay = 4), while younger age groups showed substantially longer delays: ASFR 20–24 detected in 2020 (delay = 10) and ASFR 25–29 in 2022 (delay = 12). ASFR 40–44 detected in 2023 (delay = 13), in line with its low level and high relative noise, and ASFR 15–19 produced no alarm. All tempo-adjusted ASFRs failed to detect within the window. This pattern, rapid detection at ages 30–34 but very long delays or non-detection at younger ages, points to postponement as the dominant mechanism: fertility at younger ages fell primarily because births were shifted to older ages, not because of a genuine reduction in intended family size. The decline at age 30–34, by contrast, is consistent with a real reduction in birth rates at the prime childbearing ages.

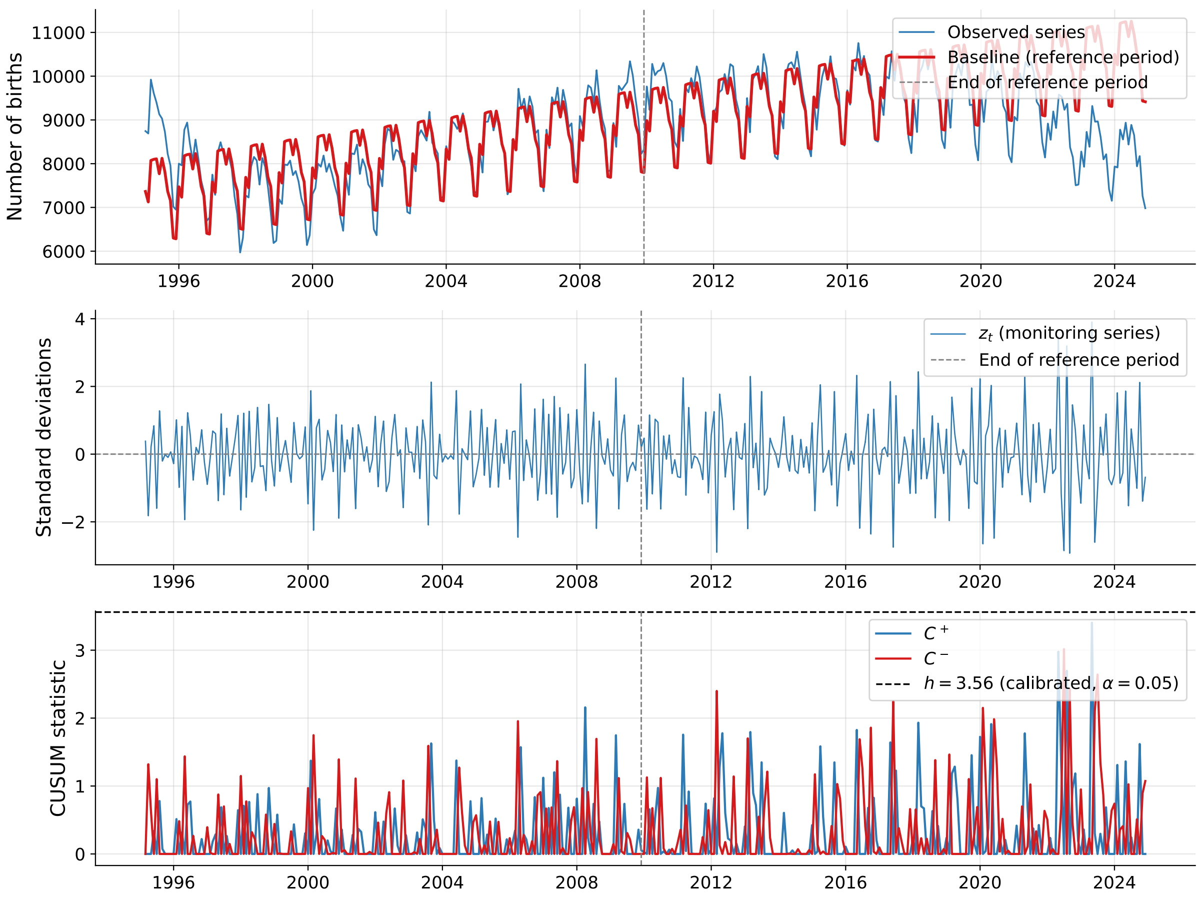

Figure 13: CUSUM on monthly births, post-2010 case study. Neither arm crosses h; monthly births again fail to detect within the ±15 year window.

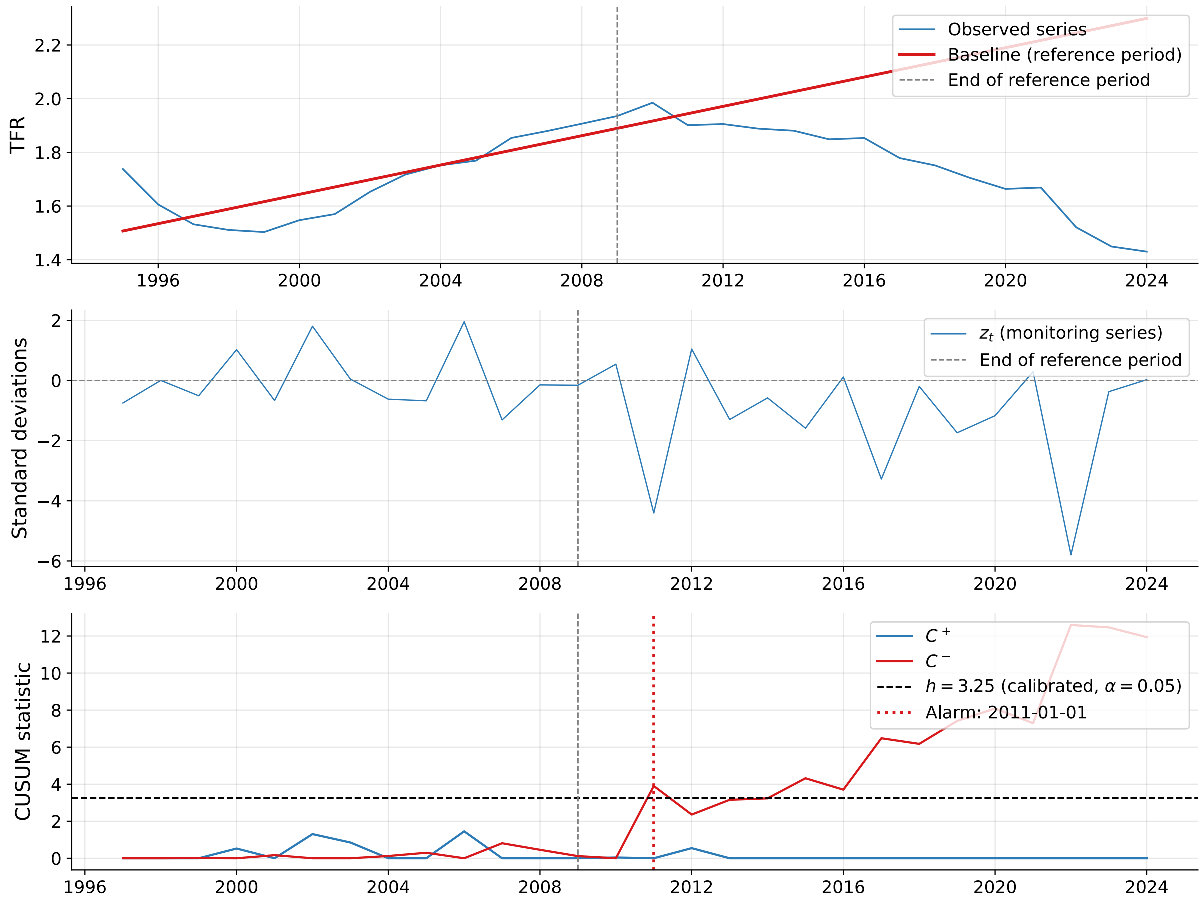

Figure 14: CUSUM on TFR, post-2010 case study. The lower-arm statistic C^- crosses h in 2011, the shortest TFR delay of the three case studies (delay = 1 year).

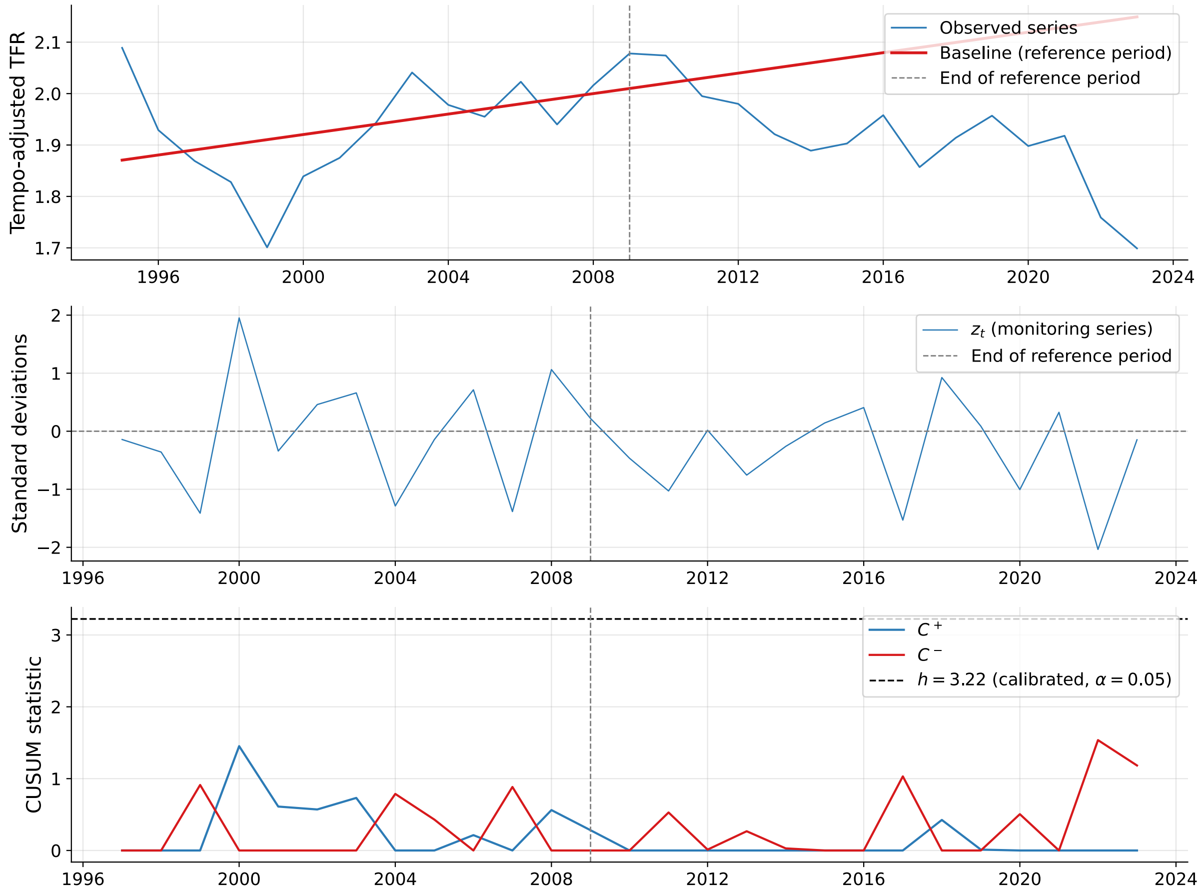

Figure 15: CUSUM on tempo-adjusted TFR, post-2010 case study. Neither arm crosses h, as in Case Study I, pointing to postponement as a major driver of the observed decline.

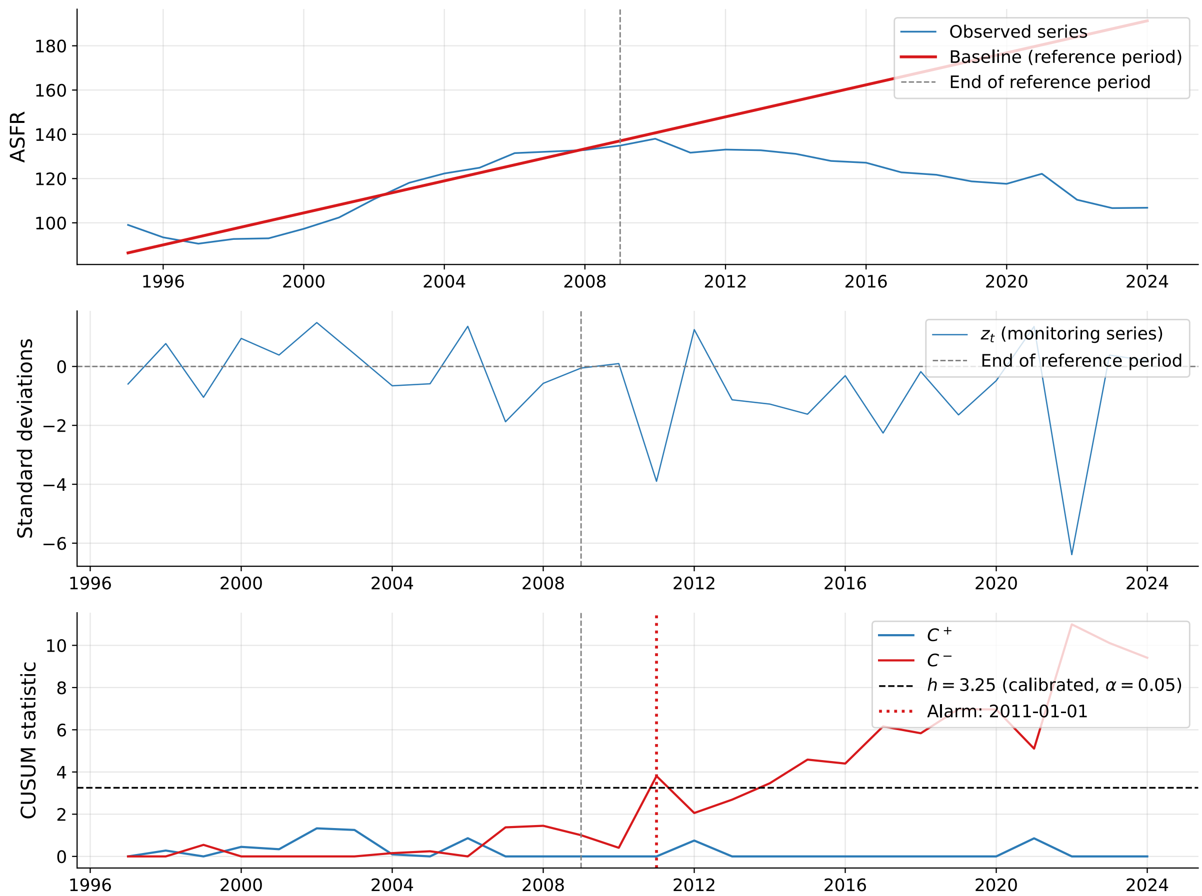

Figure 16: CUSUM on ASFR 30–34, post-2010 case study. The statistic crosses h in 2011 (delay = 1 year); the rapid alarm at this age, against long delays at younger ages, is the signature of postponement.

3.5 Summary of Detection Delays

Figure 17 summarises detection delays across the indicators and case studies as a grouped bar chart. Three patterns stand out. First, observed indicators consistently detect changes earlier than their tempo-adjusted counterparts, which either detect considerably later or not at all. Second, monthly births never produce an alarm in any case study, regardless of the underlying event. Third, the post-2010 decline yields the most heterogeneous pattern across age groups: while TFR and ASFR 30–34 detect within one year, younger age groups (ASFR 20–24, 25–29) require ten to twelve years or fail to detect entirely. This age-gradient in detection delay is a substantive finding in its own right, pointing toward postponement as an important mechanism at younger ages.

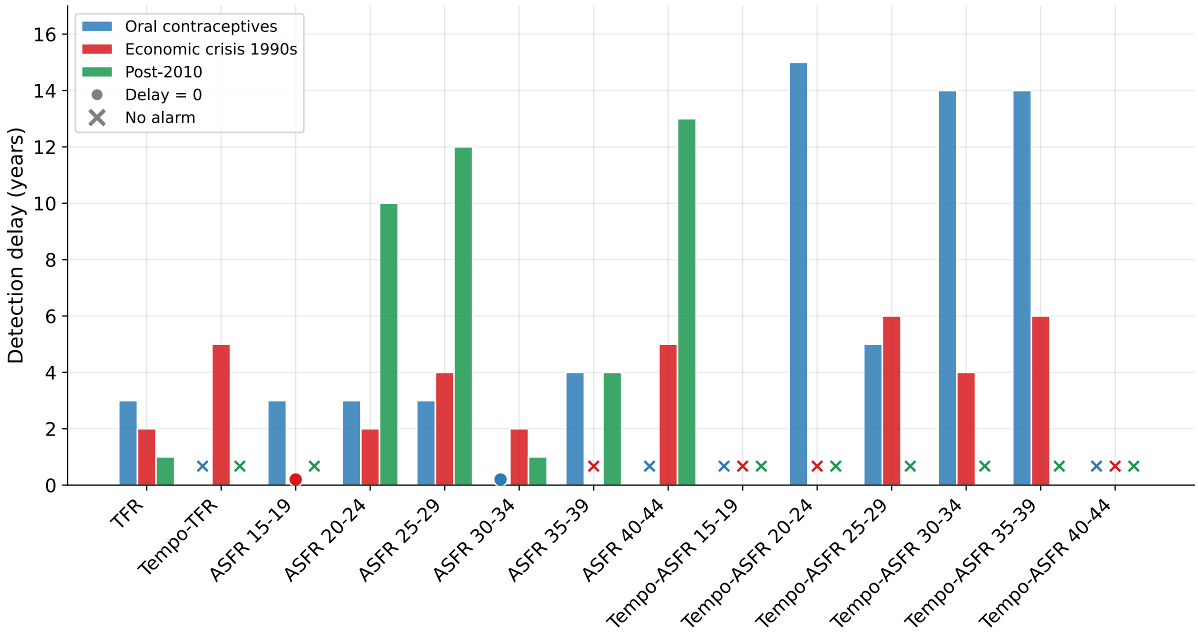

Figure 17: Detection delay in years by indicator and case study (monthly births omitted, as they never produce an alarm). Bars show years between the reference date and the CUSUM alarm; × indicates no alarm within the ±15 year window; ○ indicates zero delay (alarm at reference date). Observed indicators detect earlier than their tempo-adjusted counterparts, with the post-2010 decline showing the widest spread across age groups.

3.6 Sampling Frequency Analysis

Monthly births produced no alarms in any case study despite annual TFR consistently detecting changes. To investigate whether this stems from a genuine absence of signal or from an insufficient signal-to-noise ratio at monthly frequency, monthly births were aggregated to progressively coarser resolutions, 1-month (original), 2-month, 3-month (quarterly), 4-month, 6-month, and 12-month (annual), by summing births within consecutive non-overlapping blocks of the specified width. The seasonal dummy structure in the baseline model was adjusted to match each aggregation level, and the full pipeline applied at each frequency. Detection delays in periods were converted to months by multiplying by the aggregation factor.

Results are summarised in Figures 18–19 and Table 2. Overall, the 12-month aggregation yielded the shortest detection delays across all case studies in which alarms were detected. In contrast, the monthly series failed to produce any alarms, while the 2-month and 3-month aggregations generally led to longer detection delays. The performance of the 4-month and 6-month aggregations was more case-dependent: although they resulted in long delays for the oral contraceptives and economic crisis case, they generated substantially shorter delays in the post-2010 case. Thus, the results suggest that annual aggregation provides the most stable early-warning performance among the tested sampling frequencies. This counterintuitive finding captures the noise-reduction benefit of aggregation: summing observations over longer windows averages out short-run variation, allowing the underlying trend change to accumulate more clearly in the CUSUM statistic. However, annual aggregation comes at a cost: detection delays can only be measured in whole years, so any change occurring within a given year is dated to that year as a whole.

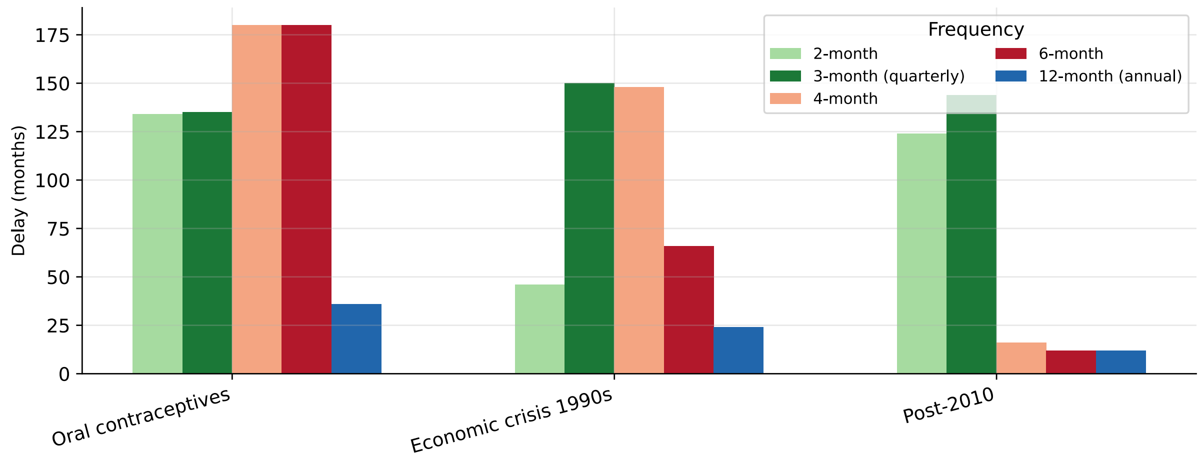

Figure 18: Detection delay in months as a function of sampling frequency, for the three main case studies. Bars are shown only where an alarm was detected; the 12-month (annual) aggregation gives the shortest delay in every case.

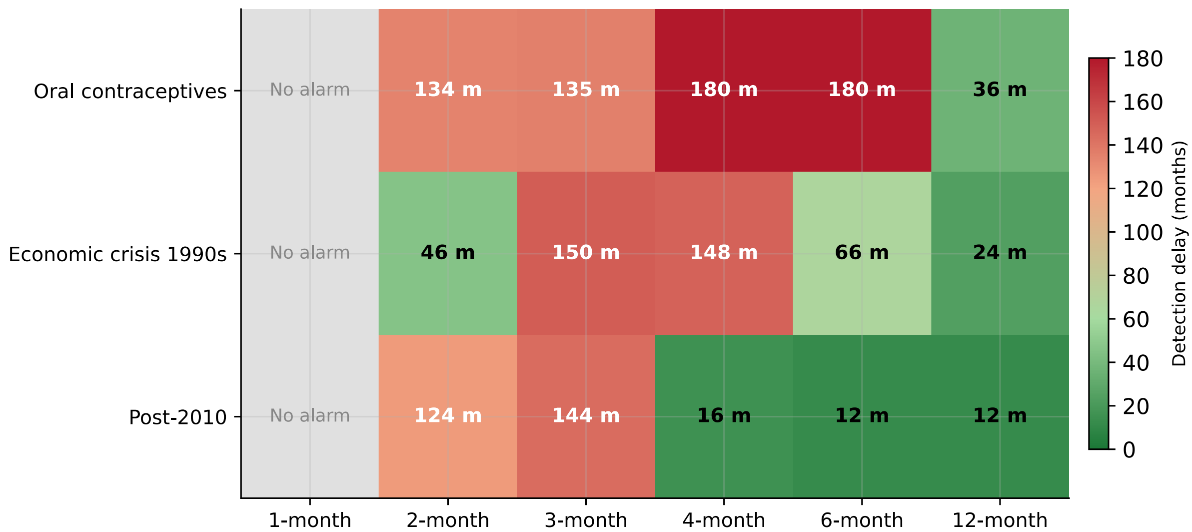

Figure 19: Detection delay in months by sampling frequency and case study. Grey cells indicate no alarm; greener cells mark shorter delays, with annual aggregation consistently the quickest to detect.

Table 2: Sampling frequency analysis results. Periods is the number of observations at each aggregation level, and detection delay is reported both in periods and converted to months. A dash indicates no alarm.

Case study	Frequency	Periods	Alarm	Delay (periods)	Delay (months)	h
Oral contraceptives	1-month	361	–	–	–	3.561
Oral contraceptives	2-month	181	1976-03-01	67	134	3.397
Oral contraceptives	3-month (qtr)	121	1976-04-01	45	135	3.289
Oral contraceptives	4-month	91	1980-01-01	45	180	3.336
Oral contraceptives	6-month	61	1980-01-01	30	180	3.196
Oral contraceptives	12-month (ann.)	31	1968-01-01	3	36	3.277
Economic crisis 1990s	1-month	361	–	–	–	3.558
Economic crisis 1990s	2-month	181	1994-11-01	23	46	3.398
Economic crisis 1990s	3-month (qtr)	121	2003-07-01	50	150	3.289
Economic crisis 1990s	4-month	91	2003-05-01	37	148	3.334
Economic crisis 1990s	6-month	61	1996-07-01	11	66	3.184
Economic crisis 1990s	12-month (ann.)	31	1993-01-01	2	24	3.277
Post-2010	1-month	360	–	–	–	3.561
Post-2010	2-month	180	2020-05-01	62	124	3.397
Post-2010	3-month (qtr)	120	2022-01-01	48	144	3.278
Post-2010	4-month	90	2011-05-01	4	16	3.365
Post-2010	6-month	60	2011-01-01	2	12	3.206
Post-2010	12-month (ann.)	30	2011-01-01	1	12	3.252

3.7 Power Analysis

To characterise the sensitivity of the CUSUM procedure, a simulation-based power analysis was conducted for TFR in the 1990s crisis. This combination was selected as a representative case where CUSUM successfully detected the break (delay = 2 years), allowing the power curve to characterise the procedure's sensitivity around a meaningful operating point; the MDC values reported in Table 1 provide the corresponding sensitivity assessment for the remaining 44 indicator–case study combinations. For each value of the slope change parameter $\delta \in \{0, 0.01, \ldots, 0.08\}$, 500 synthetic AR(1) series were generated with the estimated parameters $\hat{\phi}$ and $\hat{\sigma}$ from the empirical run, with a slope break of magnitude δ introduced at the break point, and the full pipeline applied using the calibrated h. Power was estimated as the proportion of replications in which a detection occurred at or after the break.

Figure 20 presents the resulting power curve and mean detection delay. Adopting the conventional 80% threshold for adequate statistical power, the procedure reaches this level only for slope changes substantially larger than the smallest observed in the data, indicating that gradual declines near the empirical noise floor may go undetected for several years even when they are real. The mean detection delay decreases with increasing δ but remains above three years even for the largest slope changes examined here, providing a lower bound on how quickly the procedure can be expected to react in practice. Power curves for indicators with substantially higher MDC, such as monthly births (MDC $\approx$ 319–352) or tempo-adjusted ASFRs at younger ages (MDC > 6), would lie below the present curve across the entire δ-range examined here, consistent with their failure to trigger CUSUM alarms in the empirical analysis.

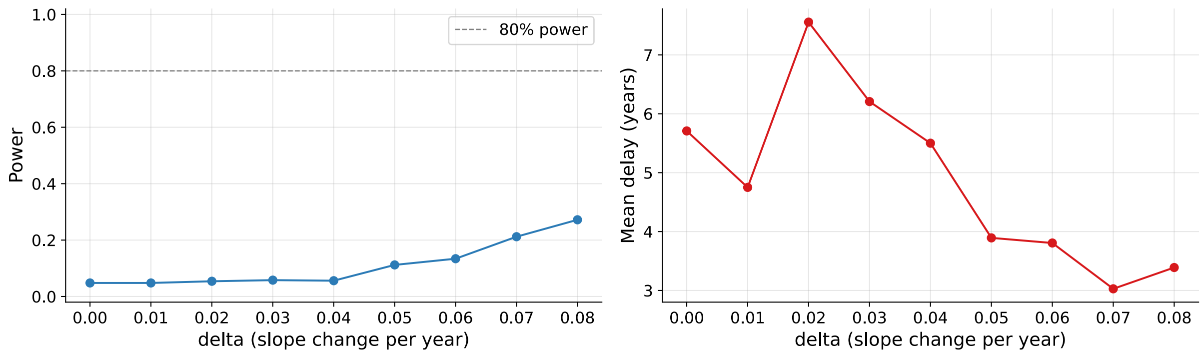

Figure 20: Estimated power (left) and mean detection delay (right) as a function of slope change magnitude δ, TFR, economic crisis 1990s. The dashed line indicates the 80% power level, reached only for slope changes well above those seen in the data; mean delay stays above three years throughout.

3.8 Offline Structural Break Methods

Three offline methods were applied to contextualise the CUSUM results: the Bai–Perron procedure, the slope change test, and the Chow test, all described in Section 2.4. For the slope change and Chow tests the known reference date is used as the break date. Bai–Perron searches for breaks without using the reference date, so any breaks identified near the reference date emerge from the data alone.

Slope change test results are shown in Figures 21–29, and Bai–Perron results in Figures 30–35. Complete numerical results are in Table 3.

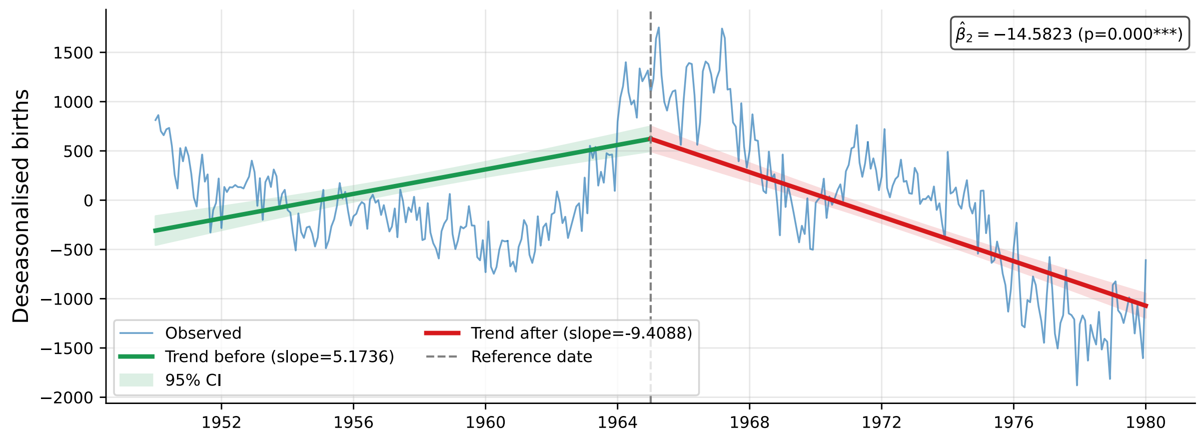

Figure 21: Slope change test on deseasonalised monthly births, oral contraceptives case study.

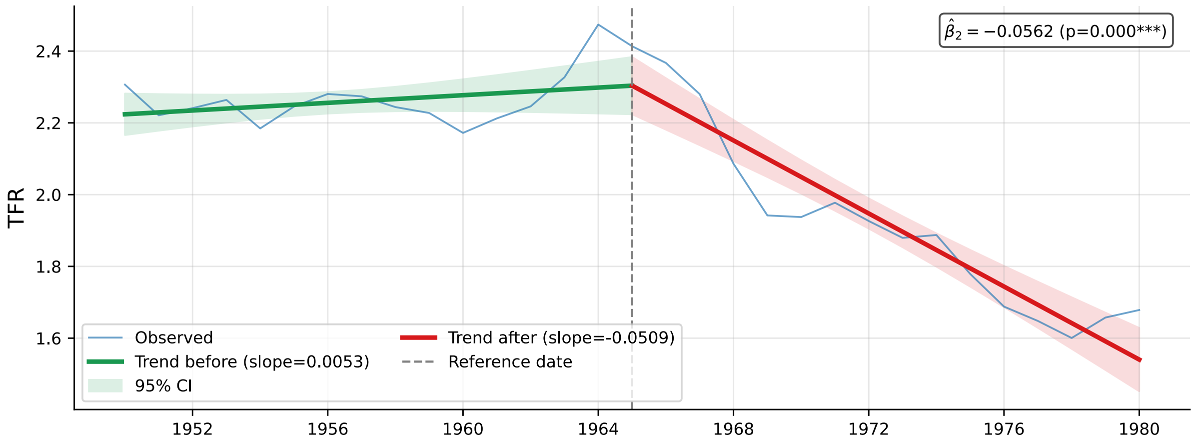

Figure 22: Slope change test on TFR, oral contraceptives case study.

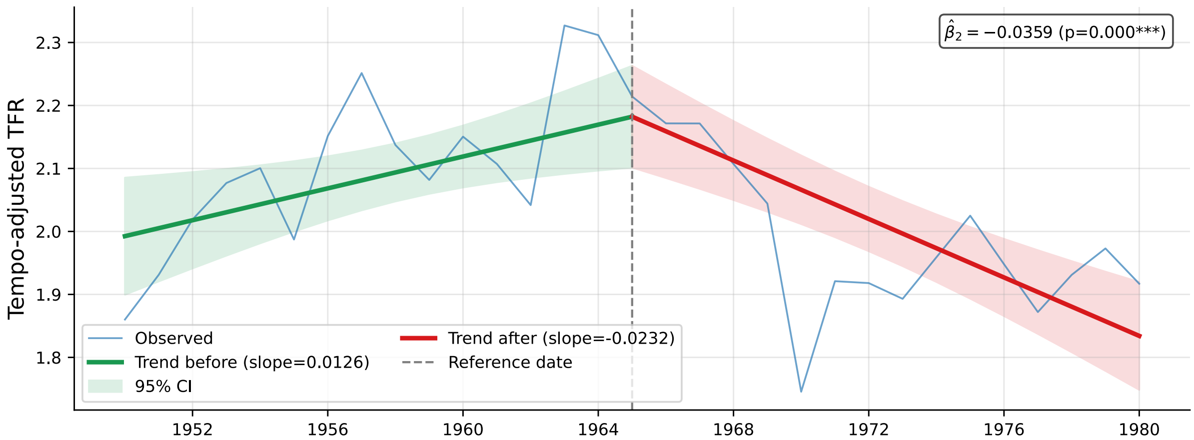

Figure 23: Slope change test on tempo-adjusted TFR, oral contraceptives case study.

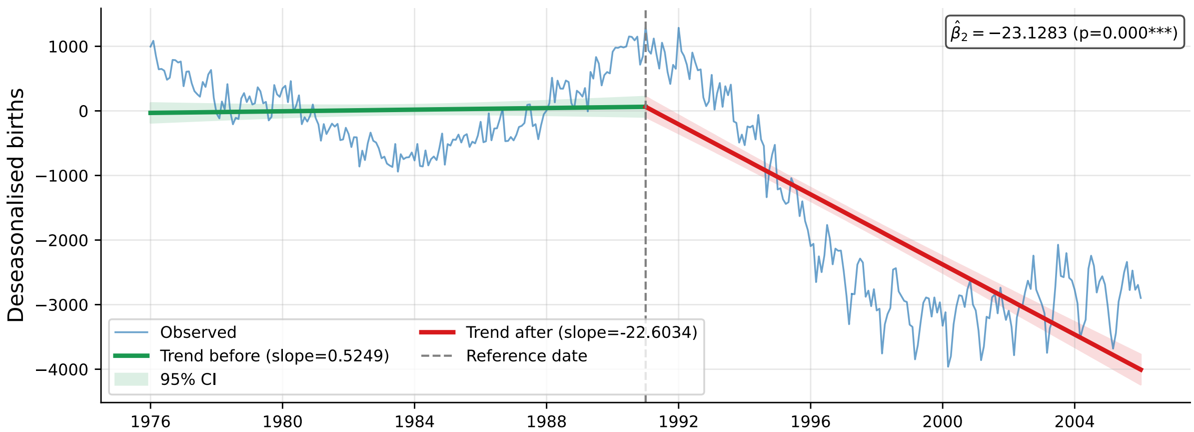

Figure 24: Slope change test on deseasonalised monthly births, economic crisis case study.

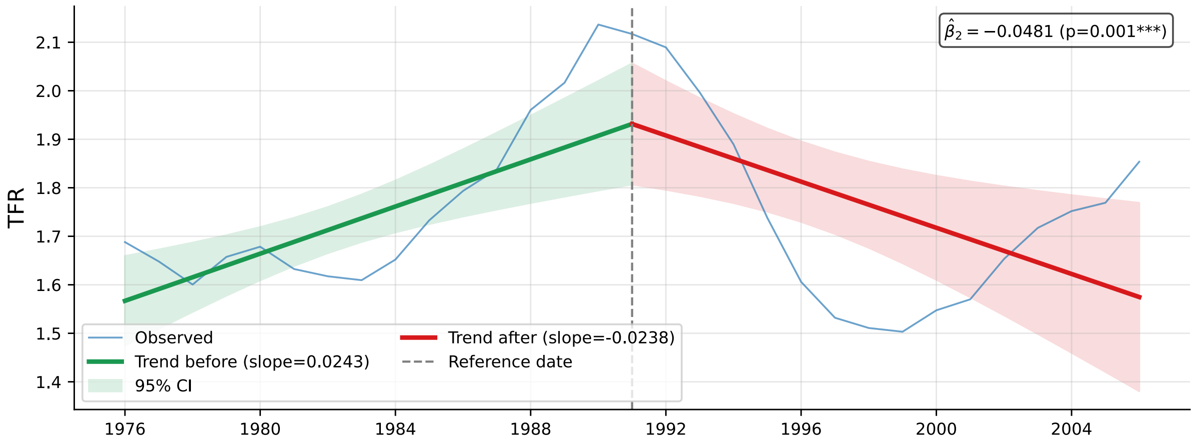

Figure 25: Slope change test on TFR, economic crisis case study.

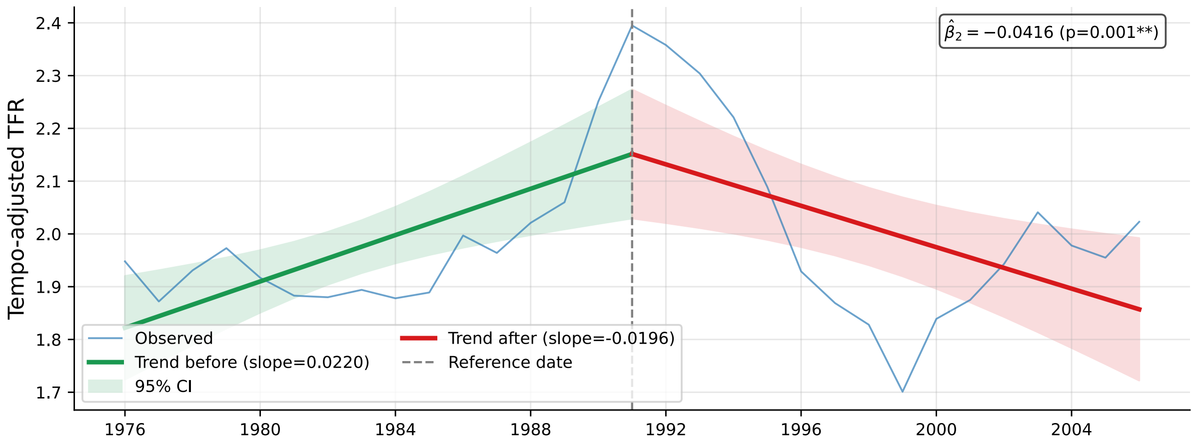

Figure 26: Slope change test on tempo-adjusted TFR, economic crisis case study.

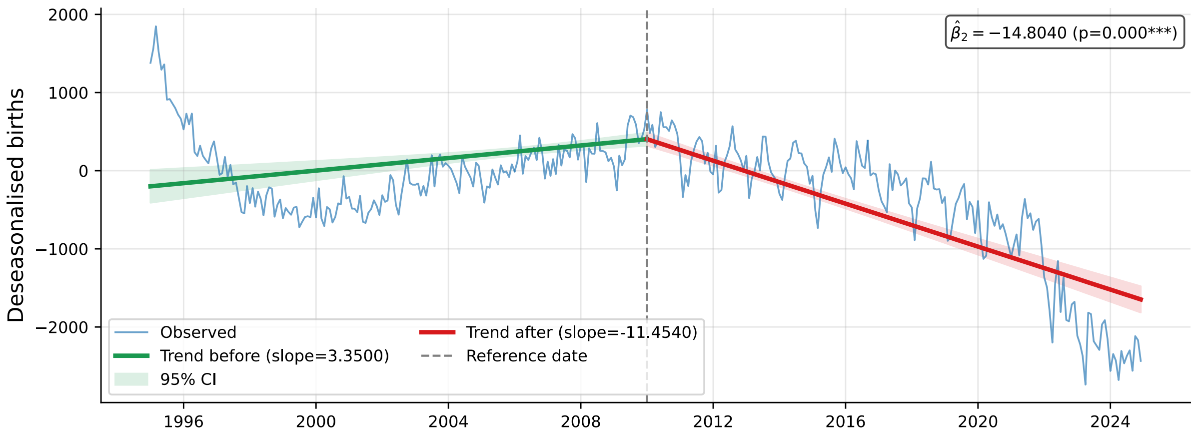

Figure 27: Slope change test on deseasonalised monthly births, post-2010 case study.

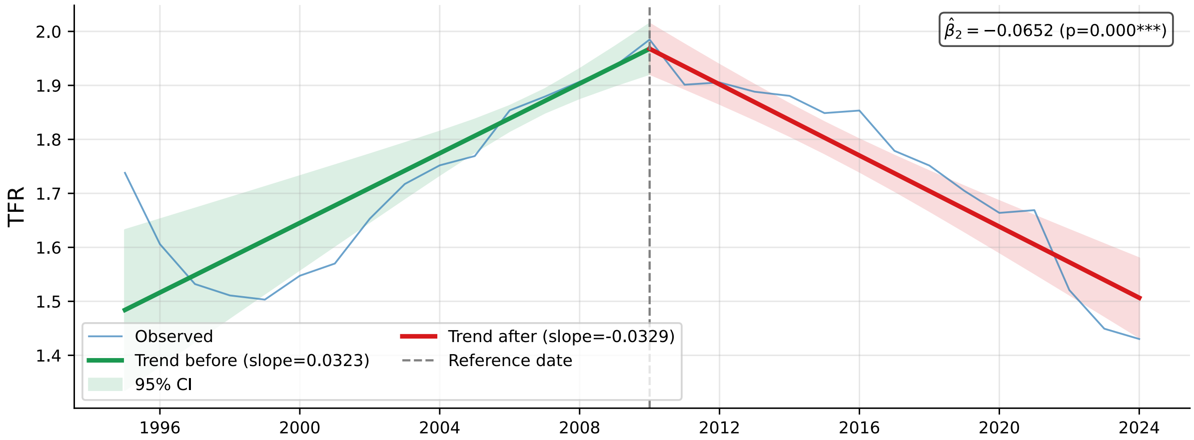

Figure 28: Slope change test on TFR, post-2010 case study.

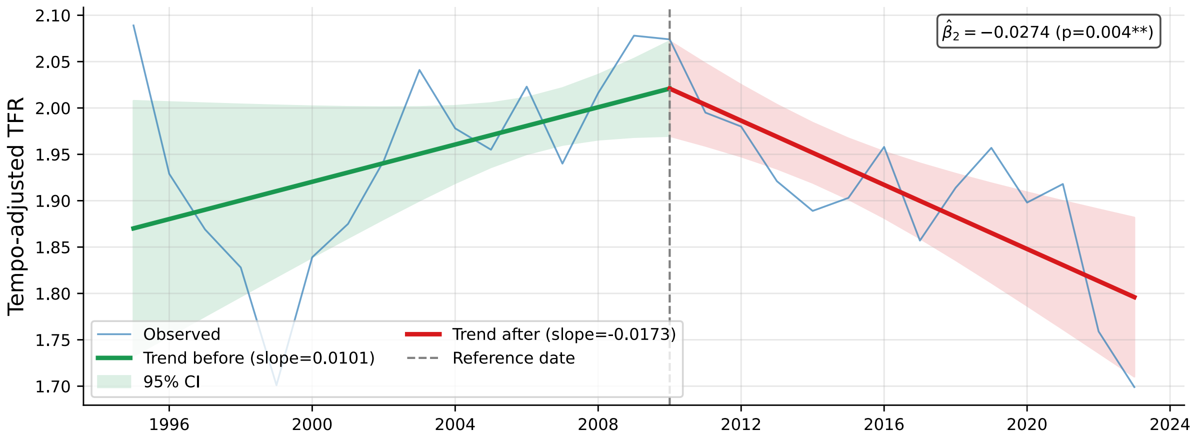

Figure 29: Slope change test on tempo-adjusted TFR, post-2010 case study.

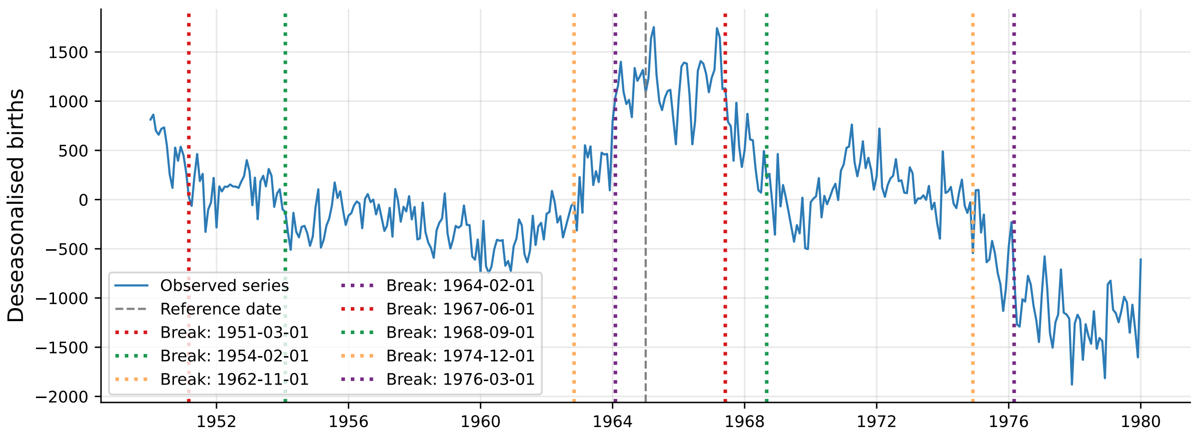

Figure 30: Bai–Perron break detection on deseasonalised monthly births, oral contraceptives case study. Vertical lines mark detected break dates.

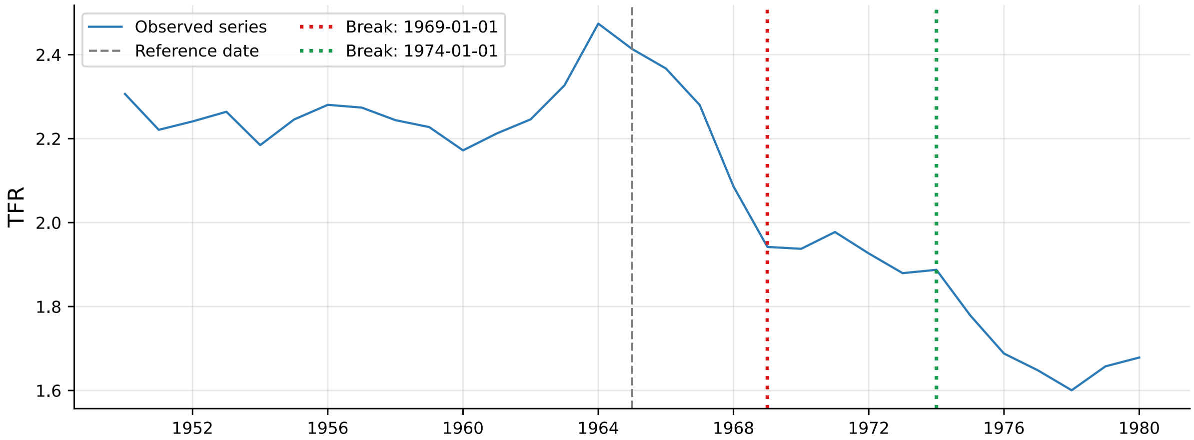

Figure 31: Bai–Perron break detection on TFR, oral contraceptives case study.

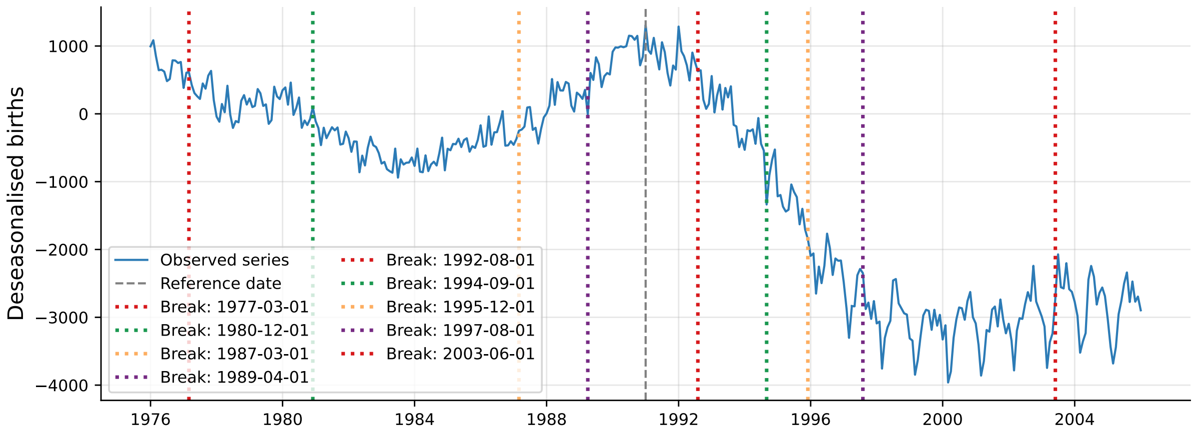

Figure 32: Bai–Perron break detection on deseasonalised monthly births, economic crisis case study.

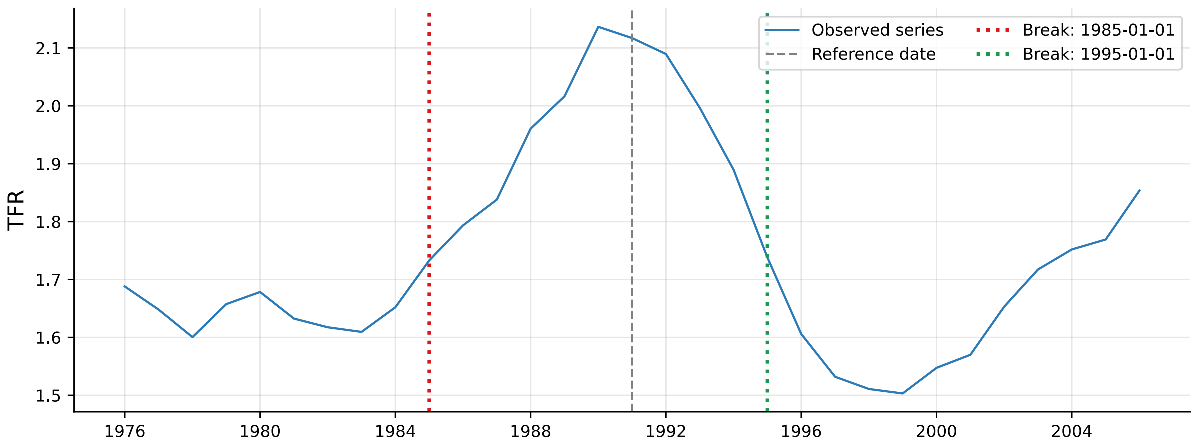

Figure 33: Bai–Perron break detection on TFR, economic crisis case study.

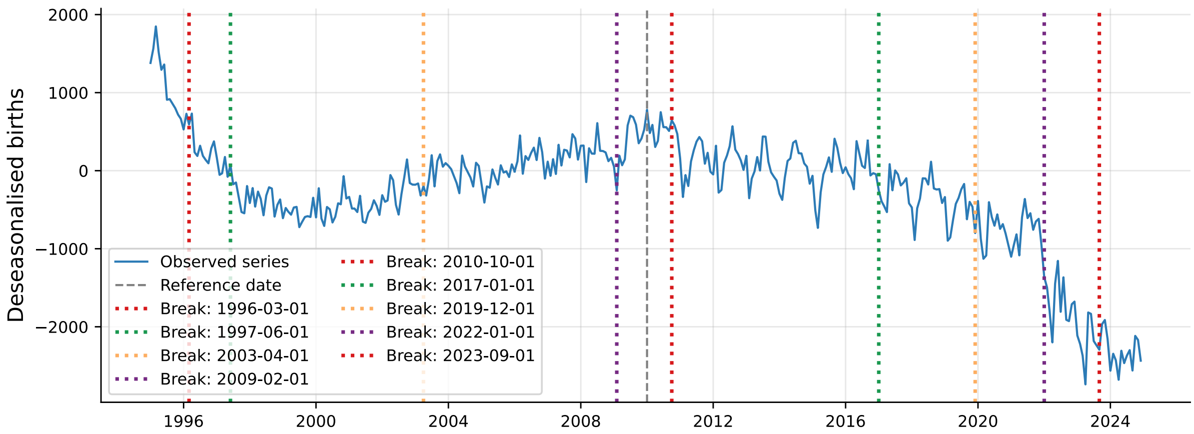

Figure 34: Bai–Perron break detection on deseasonalised monthly births, post-2010 case study.

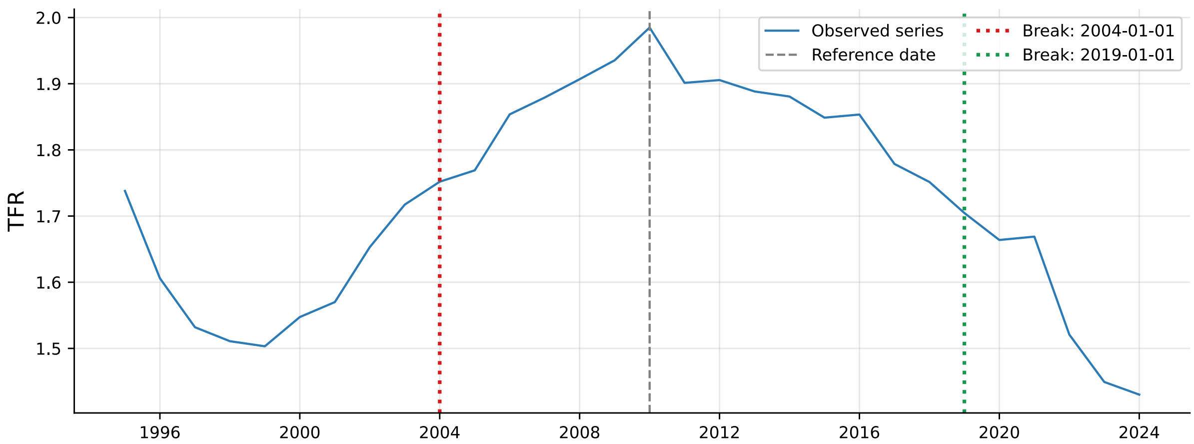

Figure 35: Bai–Perron break detection on TFR, post-2010 case study.

3.9 Cross-Validation with Offline Methods

Figure 36 presents a significance heatmap across all 45 case study–indicator combinations for all four methods. The slope change and Chow tests match each other well, registering structural change in roughly 50–100% of indicators per case study (Figure 37). CUSUM produces alarms in 40–65% of combinations. These detection rates are not directly comparable: the offline tests use the full sample and address the retrospective question of whether a structural change occurred, while CUSUM operates sequentially under a controlled false alarm rate of 5% and addresses the prospective question of when such a change would first have been detectable.

Agreement between CUSUM and Bai–Perron on the *timing* of breaks is summarised in Figure 38. Where both methods register a result, the CUSUM alarm date and the nearest Bai–Perron break typically fall within 12–24 months of each other. This is a notable consistency given the fundamental difference in operating principles: CUSUM is sequential and prospective, Bai–Perron retrospective and full-sample. Figure 39 shows that Bai–Perron breaks cluster near the reference dates for the economic crisis and post-2010 case studies, providing independent data-driven confirmation that these episodes coincide with structural changes in the fertility series.

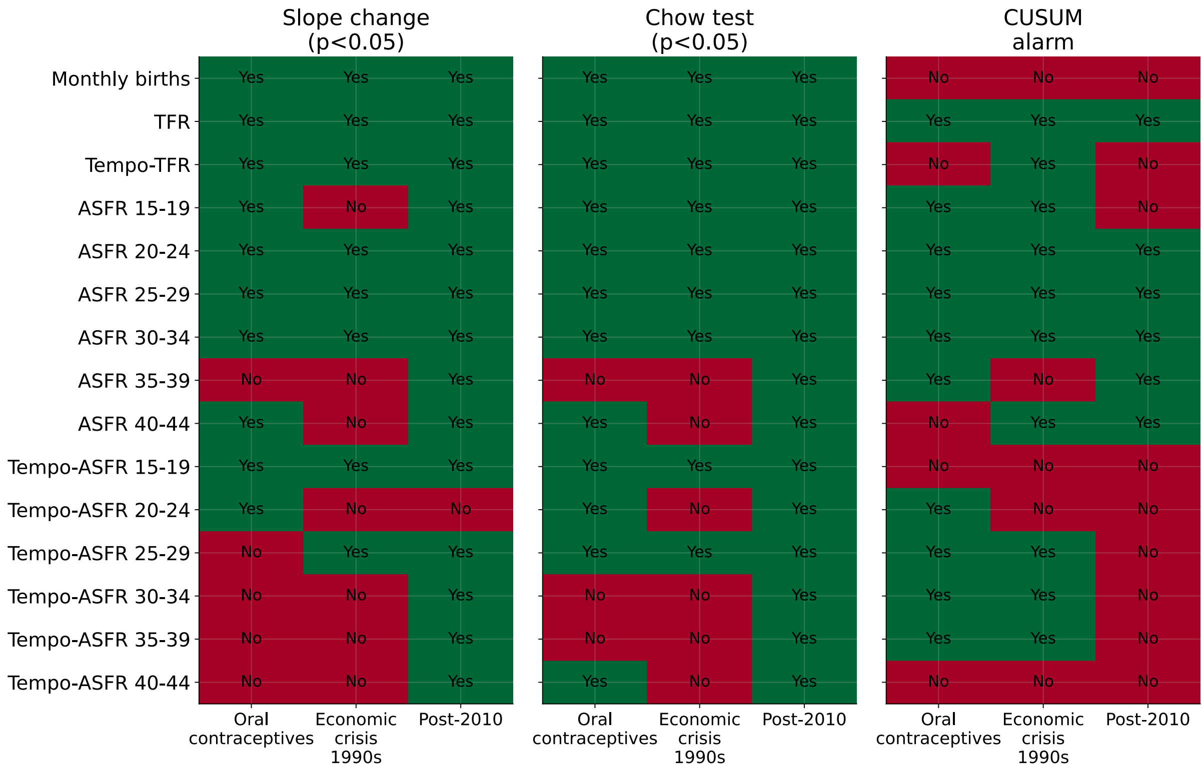

Figure 36: Significance heatmap across all 45 case study–indicator combinations. Left panel: slope change test ($p < 0.05$). Centre panel: Chow test ($p < 0.05$). Right panel: CUSUM alarm. Green = detected or significant; red = not detected.

Table 3: Summary of results: CUSUM, Bai–Perron, slope change test, and Chow test. An asterisk denotes significance at the 5% level ($p < 0.05$).

Case study	Indicator	CUSUM alarm	CUSUM delay	Bai-Perron breaks	Slope change	Chow test
Oral contraceptives	Monthly births	–	–	1951-03, 1954-02, 1962-11, 1964-02, 1967-06, 1968-09, 1974-12, 1976-03	−14.58* (p=0.000)	$F = 248.10$* (p=0.000)
Oral contraceptives	TFR	1968-01	3	1969-01, 1974-01	−0.056* (p=0.000)	$F = 38.88$* (p=0.000)
Oral contraceptives	Tempo-TFR	–	–	none	−0.036* (p=0.000)	$F = 13.52$* (p=0.000)
Oral contraceptives	ASFR 15–19	1968-01	3	1964-01, 1969-01, 1974-01	−2.571* (p=0.000)	$F = 81.99$* (p=0.000)
Oral contraceptives	ASFR 20–24	1968-01	3	1954-01, 1969-01, 1974-01	−3.745* (p=0.000)	$F = 56.60$* (p=0.000)
Oral contraceptives	ASFR 25–29	1968-01	3	1959-01, 1969-01	−3.846* (p=0.000)	$F = 50.04$* (p=0.000)
Oral contraceptives	ASFR 30–34	1965-01	0	1964-01, 1969-01	−1.214* (p=0.019)	$F = 6.93$* (p=0.004)
Oral contraceptives	ASFR 35–39	1969-01	4	1954-01, 1964-01, 1969-01	−0.053 (p=0.864)	$F = 0.14$ (p=0.869)
Oral contraceptives	ASFR 40–44	–	–	1954-01, 1964-01, 1969-01	0.162* (p=0.041)	$F = 4.80$* (p=0.016)
Oral contraceptives	Tempo-ASFR 15–19	–	–	1959-01, 1964-01, 1969-01, 1974-01	−2.245* (p=0.000)	$F = 56.49$* (p=0.000)
Oral contraceptives	Tempo-ASFR 20–24	1980-01	15	1954-01, 1969-01	−2.099* (p=0.002)	$F = 10.60$* (p=0.000)
Oral contraceptives	Tempo-ASFR 25–29	1970-01	5	1954-01, 1969-01, 1974-01	−1.699 (p=0.098)	$F = 6.03$* (p=0.007)
Oral contraceptives	Tempo-ASFR 30–34	1979-01	14	1969-01, 1974-01	−0.230 (p=0.740)	$F = 3.15$ (p=0.059)
Oral contraceptives	Tempo-ASFR 35–39	1979-01	14	1959-01, 1964-01, 1969-01	0.143 (p=0.688)	$F = 1.87$ (p=0.174)
Oral contraceptives	Tempo-ASFR 40–44	–	–	1954-01, 1964-01, 1969-01	0.132 (p=0.155)	$F = 4.93$* (p=0.015)
Economic crisis 1990s	Monthly births	–	–	1977-03, 1980-12, 1987-03, 1989-04, 1992-08, 1994-09, 1995-12, 1997-08, 2003-06	−23.13* (p=0.000)	$F = 133.24$* (p=0.000)
Economic crisis 1990s	TFR	1993-01	2	1985-01, 1995-01	−0.048* (p=0.001)	$F = 8.39$* (p=0.001)
Economic crisis 1990s	Tempo-TFR	1996-01	5	1985-01, 1990-01, 1995-01, 2000-01	−0.042* (p=0.001)	$F = 8.81$* (p=0.001)
Economic crisis 1990s	ASFR 15–19	1991-01	0	1980-01, 1995-01	0.339 (p=0.149)	$F = 3.65$* (p=0.040)
Economic crisis 1990s	ASFR 20–24	1993-01	2	1980-01, 1995-01	−2.142* (p=0.006)	$F = 5.83$* (p=0.008)
Economic crisis 1990s	ASFR 25–29	1995-01	4	1985-01, 1995-01	−5.362* (p=0.000)	$F = 28.36$* (p=0.000)
Economic crisis 1990s	ASFR 30–34	1993-01	2	1985-01, 2000-01	−2.110* (p=0.009)	$F = 6.15$* (p=0.006)
Economic crisis 1990s	ASFR 35–39	–	–	1985-01, 2000-01	−0.370 (p=0.257)	$F = 1.29$ (p=0.292)
Economic crisis 1990s	ASFR 40–44	1996-01	5	1985-01, 1990-01, 2000-01	0.025 (p=0.635)	$F = 1.09$ (p=0.350)
Economic crisis 1990s	Tempo-ASFR 15–19	–	–	1980-01, 1995-01	0.576* (p=0.036)	$F = 9.64$* (p=0.001)
Economic crisis 1990s	Tempo-ASFR 20–24	–	–	1980-01, 1995-01	−1.334 (p=0.152)	$F = 3.13$ (p=0.060)
Economic crisis 1990s	Tempo-ASFR 25–29	1997-01	6	1990-01, 1995-01	−4.653* (p=0.000)	$F = 12.60$* (p=0.000)
Economic crisis 1990s	Tempo-ASFR 30–34	1995-01	4	1985-01, 2000-01	−1.610 (p=0.082)	$F = 1.77$ (p=0.189)
Economic crisis 1990s	Tempo-ASFR 35–39	1997-01	6	1985-01, 1990-01, 2000-01	−0.161 (p=0.674)	$F = 0.59$ (p=0.561)
Economic crisis 1990s	Tempo-ASFR 40–44	–	–	1980-01, 1990-01, 2000-01	0.071 (p=0.252)	$F = 0.83$ (p=0.448)
Post-2010	Monthly births	–	–	1996-03, 1997-06, 2003-04, 2009-02, 2010-10, 2017-01, 2019-12, 2022-01, 2023-09	−14.80* (p=0.000)	$F = 164.61$* (p=0.000)
Post-2010	TFR	2011-01	1	2004-01, 2019-01	−0.065* (p=0.000)	$F = 56.91$* (p=0.000)
Post-2010	Tempo-TFR	–	–	2019-01	−0.027* (p=0.004)	$F = 6.79$* (p=0.004)
Post-2010	ASFR 15–19	–	–	1999-01, 2009-01, 2014-01, 2019-01	−0.157* (p=0.000)	$F = 15.18$* (p=0.000)
Post-2010	ASFR 20–24	2020-01	10	1999-01, 2014-01, 2019-01	−1.240* (p=0.006)	$F = 12.03$* (p=0.000)
Post-2010	ASFR 25–29	2022-01	12	2019-01	−2.693* (p=0.000)	$F = 23.15$* (p=0.000)
Post-2010	ASFR 30–34	2011-01	1	2004-01, 2019-01	−5.702* (p=0.000)	$F = 115.38$* (p=0.000)
Post-2010	ASFR 35–39	2014-01	4	2004-01	−2.869* (p=0.000)	$F = 127.70$* (p=0.000)
Post-2010	ASFR 40–44	2023-01	13	2004-01, 2009-01	−0.394* (p=0.000)	$F = 43.25$* (p=0.000)
Post-2010	Tempo-ASFR 15–19	–	–	1999-01, 2009-01, 2019-01	−0.103* (p=0.003)	$F = 7.77$* (p=0.002)
Post-2010	Tempo-ASFR 20–24	–	–	1999-01, 2014-01, 2019-01	−0.762 (p=0.084)	$F = 4.84$* (p=0.016)
Post-2010	Tempo-ASFR 25–29	–	–	2014-01, 2019-01	−1.587* (p=0.020)	$F = 4.45$* (p=0.022)
Post-2010	Tempo-ASFR 30–34	–	–	1999-01	−4.788* (p=0.000)	$F = 23.02$* (p=0.000)
Post-2010	Tempo-ASFR 35–39	–	–	2004-01	−2.432* (p=0.000)	$F = 26.30$* (p=0.000)
Post-2010	Tempo-ASFR 40–44	–	–	2004-01, 2009-01	−0.288* (p=0.000)	$F = 7.28$* (p=0.003)

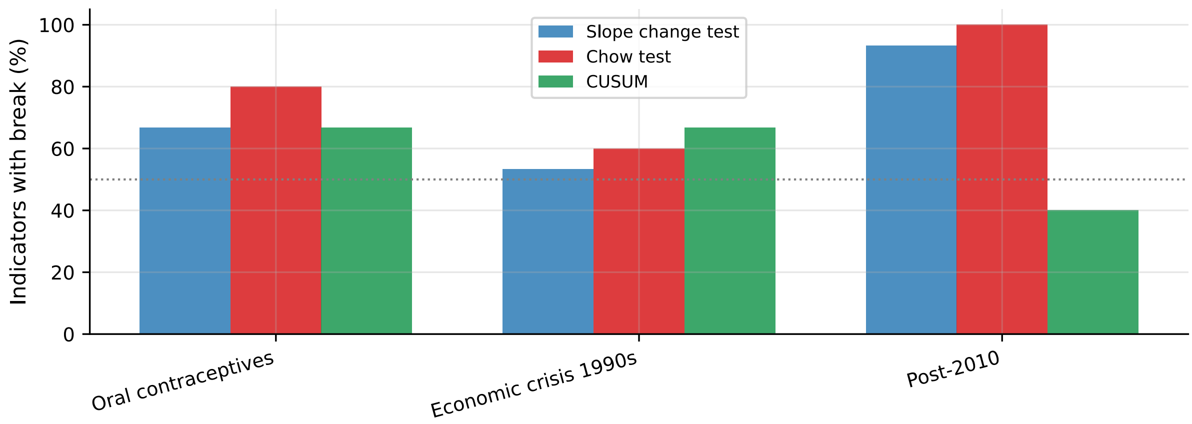

Figure 37: Percentage of indicators with a registered structural change per method and case study.

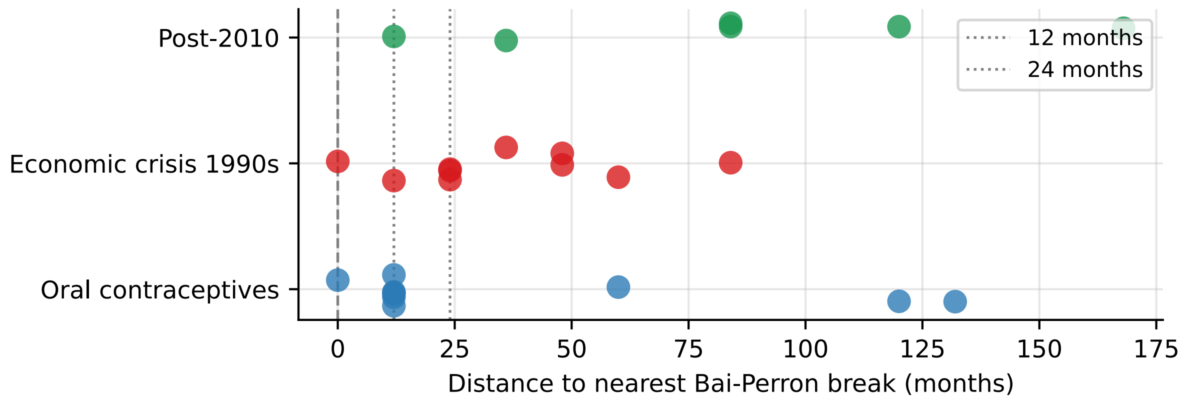

Figure 38: Months between each CUSUM alarm and the nearest Bai–Perron break, for all indicator–case study pairs where both methods produced a result. Points near zero indicate timing agreement.

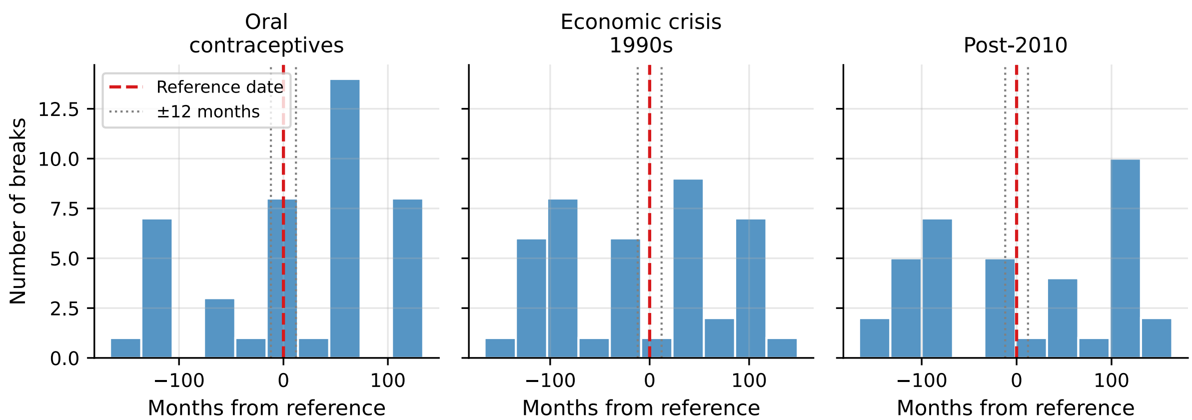

Figure 39: Distribution of distances (in months) between each Bai–Perron break and the reference date, by case study. Breaks clustered near zero provide data-driven confirmation of structural change near the event.

4 Discussion

4.1 Prospective detection versus retrospective confirmation

A central finding of this thesis is the systematic gap between what can be confirmed in hindsight and what was statistically detectable as the change unfolded. The slope change and Chow tests, applied with knowledge of the full sample, identify significant structural breaks in roughly 50–100% of indicator–case study combinations. CUSUM, constrained to process data sequentially under a controlled false alarm rate, detects in only 40–65% of the same combinations. This gap is not a failure of CUSUM but a quantification of a genuine statistical limitation: ruling out ordinary fluctuation as the data arrive requires more evidence than confirming a break in hindsight.

The detection delays themselves, ranging from zero to fifteen years across indicators and episodes, give concrete meaning to this limitation. Even in the 1990s economic crisis, arguably the sharpest and most abrupt fertility shock in postwar Swedish history, TFR required two years to accumulate sufficient evidence, and tempo-adjusted TFR required five. A demographer monitoring fertility prospectively in 1991 would have faced genuine uncertainty about whether an emerging decline was real or transitory for at least two years after the break had occurred. For more gradual episodes such as the post-2010 decline, this uncertainty extends to a decade or more for younger age groups.

4.2 Indicator choice and the tempo–quantum distinction

The comparison between observed and tempo-adjusted indicators across all three case studies reveals a consistent and substantively important pattern. Observed indicators detect earlier in every case study and for nearly every age group. Tempo-adjusted indicators either detect considerably later or not at all. This difference carries demographic meaning beyond the purely statistical.

For the 1960s contraceptive episode, neither the tempo-adjusted TFR nor any tempo-adjusted ASFR at younger ages produced a timely alarm, suggesting that the TFR decline observed after 1965 was in large part a timing artefact: women were postponing births rather than permanently reducing their intended family size, which is exactly what the Second Demographic Transition literature would predict for this period (Lesthaeghe 2010; van de Kaa 1987). The few tempo-adjusted alarms that did appear came at intermediate ages and with delays of five to fifteen years, suggesting a slow and partial quantum effect embedded within a larger timing shift.

For the 1990s economic crisis, the picture is qualitatively different. The tempo-adjusted TFR did produce an alarm, albeit five years after the observed TFR. This indicates that the crisis generated a genuine quantum reduction in fertility, which agrees with B. Hoem (2000) and Andersson (2000), who document falling first-birth rates and strong pro-cyclical income effects, not merely a postponement of births that would later be recuperated.

The post-2010 decline most closely resembles the contraceptive episode in this respect: tempo-adjusted TFR produced no alarm, and no tempo-adjusted ASFR detected within the window. This suggests that postponement is a major contributing mechanism. It is in line with Hellstrand et al. (2021), who show that completed cohort fertility has only recently begun to decline for the youngest cohorts, and with Ohlsson-Wijk and Andersson (2022), who document that the decline is concentrated at younger ages within each parity. At the same time, the rapid alarm on observed ASFR 30–34, the age group least susceptible to further postponement, adds nuance: some genuine reduction in birth rates at peak childbearing ages does appear to be occurring, even if it cannot yet be confirmed in tempo-adjusted series.

4.3 The failure of monthly births

Monthly births produced no alarm at their original resolution, and annual aggregation gave the most stable detection; intermediate frequencies detected unevenly. The sampling frequency analysis clarifies why: the monthly birth series carries a signal-to-noise ratio that is simply too low for CUSUM to accumulate sufficient evidence against any of the three historical breaks within a fifteen-year window. This is not because births are uninformative about underlying fertility. The slope change and Chow tests confirm highly significant breaks in deseasonalised monthly births for all three episodes. The reason is that month-to-month variation around any trend is large relative to the gradual slope changes of interest. Annual aggregation resolves this by averaging out short-run noise, reducing the effective variance of the monitoring series and allowing the cumulative signal to build more rapidly.

The implication for demographic surveillance is practical: a monitoring system based on raw monthly births will be systematically late relative to one based on annualised fertility rates, even though the underlying data are the same. TFR, by standardising for population structure and aggregating over twelve months of births per woman-year, acts as an effective noise filter that monthly counts cannot replicate.

4.4 Age-group heterogeneity and what it reveals

The variation in detection delay across age groups, particularly pronounced in the post-2010 case study, is among the most substantively informative findings of the analysis. In that case study, the contrast between a one-year delay for ASFR 30–34 and delays of ten to twelve years (or no detection) for ASFR 20–24 and 25–29 is difficult to explain except by reference to postponement. If the decline at younger ages were driven by a genuine reduction in intended births, one would expect a reasonably uniform CUSUM signal across age groups, similar to the pattern in the 1990s crisis. Instead, the pattern seen is precisely what a postponement mechanism predicts: the age groups from which births are being transferred show weak or absent signals, while the destination age group (30–34) shows a strong and early signal as expected births arrive at older ages.

For the 1990s crisis, the age pattern is different. The earliest alarm came from ASFR 15–19 (delay = 0), followed closely by ASFR 20–24 and 30–34. This broad simultaneous response across age groups matches an economic shock that reduced fertility propensity across the board, without a systematic directional shift in timing, precisely the mechanism documented by B. Hoem (2000) and Sobotka, Skirbekk, et al. (2011).

4.5 Cross-validation with offline methods and previous research

The offline methods, Bai–Perron, slope change test, and Chow test, consistently identify structural change in the fertility series during each case-study window. This supports the interpretation that the episodes coincide with real shifts rather than artefacts of the CUSUM specification, although the precise location of Bai–Perron breaks does not always align exactly with the chosen reference dates. The temporal agreement between CUSUM alarms and the nearest Bai–Perron breaks (typically within 12–24 months) is reassuring: a window of one to two years is small relative to the multi-year detection delays observed across the case studies, and well within the natural granularity of annual fertility data. The two methods, operating under very different assumptions and using different amounts of data, broadly agree on when each break occurred.

The main value added by CUSUM relative to these offline benchmarks is not accuracy but timing: the question it answers is not "did a break occur?" but "when would an observer following the data sequentially have had sufficient evidence to conclude that a break had occurred?" The gap between Bai–Perron break dates and CUSUM alarm dates, sometimes several years, quantifies precisely the surveillance lag that demographic monitoring systems must contend with.

4.6 Limitations

Several limitations should be acknowledged. The analysis is not designed for causal identification: CUSUM alarms and structural break dates indicate *when* a series departed from its prior trend, but not *why*. The assignment of reference dates to historical events necessarily involves some arbitrariness, particularly for episodes that do not map onto a single institutional event, such as the gradual diffusion of oral contraceptives or the post-2010 decline, where the choice of 2010 represents a stylised onset rather than a discrete trigger.

The cross-validation with offline methods in Section 3.9 should be interpreted as triangulation rather than as a formal performance comparison. Because the true break points in real fertility data are unknown, statements about which method is "better" at detecting changes cannot be made directly from the empirical analysis. Establishing relative power, false alarm rates, and detection delays would require a simulation study with known break points and magnitudes; this is left for future work.

The power analysis in Section 3.7 is restricted to TFR in the 1990s crisis as a representative case where CUSUM successfully detected the break. MDC values for the remaining 44 indicator–case study combinations are reported in Table 1 and provide the corresponding sensitivity assessment, but a systematic power study across all combinations is not undertaken here.

Several methodological choices warrant explicit acknowledgement. Results may be sensitive to the choice of reference window length; sensitivity to this choice is reported in Appendix B. The CUSUM reference value $k = 0.5$ follows the conventional choice for detecting one-standard-deviation shifts but was not separately optimised for the slope-change setting after differencing. The Bongaarts–Feeney adjustment applied separately to each age-group ASFR extends a method originally derived for aggregate TFR; while internally consistent, this is a non-standard application that future work could refine.

The analysis is also constrained by data availability. Annual TFR and ASFR are available from 1950, but the tempo-adjusted TFR from the HFD begins only in 1971, requiring a self-computed spliced estimate for the oral contraceptives case study. While Figure 5 confirms that the self-computed estimate and the HFD series are closely consistent over their overlap period, the pre-1971 extension introduces additional measurement uncertainty. Finally, the Bongaarts–Feeney tempo adjustment, while standard and widely used, has known limitations discussed in Section 2.1: it captures only mean shifts in the timing schedule and may not fully account for changes in the variance of the age distribution (Kohler and Philipov 2001).

5 Conclusion

This thesis has applied sequential change detection to Swedish fertility data from 1950 to 2025, asking a question that existing demographic research rarely poses directly: not why fertility declined, but *when* each decline first became statistically detectable as the change unfolded. The answer depends critically on which indicator is monitored, and this dependence is itself demographically informative.

Across three historical episodes, the diffusion of oral contraceptives (1965), the Swedish economic crisis (1991), and the post-2010 fertility decline, several findings emerge with consistency. First, observed TFR is the single most reliable early-warning indicator, detecting in all three case studies with delays of one to three years. Monthly births, despite their high temporal resolution, detect promptly only once aggregated to the annual level; intermediate frequencies detect unevenly, because the short-run noise in this series overwhelms any gradual trend change. Second, annual aggregation of birth counts gives the shortest detection delay in every case, highlighting that signal-to-noise ratio, not temporal granularity, is the binding constraint on surveillance timeliness. Third, tempo-adjusted indicators consistently detect later than their observed counterparts or not at all, a pattern that, taken across all three episodes, provides a structured statistical lens onto the tempo–quantum distinction that is central to demographic theory.

The substantive interpretation of this third finding differs across episodes. For the 1960s decline and the post-2010 decline, the absence of CUSUM alarms in tempo-adjusted series points toward postponement as a major contributing mechanism: the observed TFR fell in part because births were shifted to later ages rather than only because women reduced their intended family size. For the 1990s crisis, the tempo-adjusted TFR did produce an alarm, with a longer delay but detectably, indicating a genuine quantum reduction in line with the pro-cyclical income and employment mechanisms documented in the demographic literature. The age-group pattern of detection delays reinforces this interpretation: the 1990s crisis generated broad simultaneous alarms across age groups, while the post-2010 decline produced rapid detection at age 30–34 but very long delays at younger ages, the signature of a postponement-dominated process.

The methodological contribution of this thesis is to demonstrate that CUSUM, adapted for slope changes in autocorrelated demographic time series, provides a principled framework for analysing the prospective detectability of fertility declines and quantifying the lag between an underlying change and the moment it becomes statistically distinguishable from ordinary fluctuation. The cross-validation with offline methods, Bai–Perron, slope change test, and Chow test, shows that structural breaks confirmed in hindsight may have been statistically invisible for two to fifteen years after the true break date, depending on the indicator and the nature of the shock. This quantification of the surveillance lag has direct implications for the design of early warning systems in national statistical offices: a monitoring dashboard based on annual TFR and age-group-specific rates will systematically outperform one based on monthly birth counts, and the inclusion of tempo-adjusted series alongside observed series adds qualitative information about the nature of any detected change.

Future work could extend the framework in several directions. A natural next step is to move from the fixed-window analysis used here to a fully prospective implementation with rolling reference windows, in which the baseline regression and threshold calibration update sequentially as new annual data arrive. Such an implementation would more closely match how prospective monitoring is actually carried out in practice and would make it possible to evaluate detection timing without conditioning on known reference dates. A second extension would apply the same pipeline to the other Nordic countries, allowing the synchronised post-2010 decline documented by Nordic Statistics (2024) and Hellstrand et al. (2021) to be examined through the lens of prospective detection: did the decline become detectable simultaneously across countries, or did Sweden's signal emerge earlier or later than its neighbours? A third extension would replace the AR(1) pre-whitening with a richer ARMA or state-space model, potentially improving both power and calibration accuracy for the threshold. The AR(1) specification was chosen here for its simplicity and tractability in the Monte Carlo calibration step; a richer model would capture more of the residual dependence structure but at the cost of additional parameters to estimate from short reference windows, with the attendant risk of overfitting. Finally, a formal simulation study with known break points would allow a direct performance comparison between CUSUM and the offline methods used here, complementing the cross-validation provided in Section 3.9.

References

Andersson, G. (2000). "The Impact of Labour-Force Participation on Childbearing Behaviour: Pro-Cyclical Fertility in Sweden during the 1980s and the 1990s". In: *European Journal of Population* 16.4, pp. 293–333. DOI: 10.1023/A:1006454909642.

— (2008). "Childbearing Trends in Europe". In: *Demographic Research* 19, pp. 1–30. DOI: 10.4054/DemRes.2008.19.1.

Bai, J. and P. Perron (1998). "Estimating and Testing Linear Models with Multiple Structural Changes". In: *Econometrica* 66.1, pp. 47–78. DOI: 10.2307/2998540.

— (2003). "Computation and Analysis of Multiple Structural Change Models". In: *Journal of Applied Econometrics* 18.1, pp. 1–22. DOI: 10.1002/jae.659.

Bailey, M. J. (2006). "More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Supply". In: *Quarterly Journal of Economics* 121.1, pp. 289–320. DOI: 10.1093/qje/121.1.289.

Basseville, M. and I. V. Nikiforov (1993). *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs, NJ: Prentice Hall. ISBN: 0-13-126780-9.

Billari, F. C. and H.-P. Kohler (2004). "Patterns of Low and Lowest-Low Fertility in Europe". In: *Population Studies* 58.2, pp. 161–176. DOI: 10.1080/0032472042000213695.

Bongaarts, J. and G. Feeney (1998). "On the Quantum and Tempo of Fertility". In: *Population and Development Review* 24.2, pp. 271–291. DOI: 10.2307/2807974.

Bongaarts, J. and T. Sobotka (2012). "A Demographic Explanation for the Recent Rise in European Fertility". In: *Population and Development Review* 38.1, pp. 83–120. DOI: 10.1111/j.1728-4457.2012.00473.x.

Brown, R. L., J. Durbin, and J. M. Evans (1975). "Techniques for Testing the Constancy of Regression Relationships Over Time". In: *Journal of the Royal Statistical Society: Series B* 37.2, pp. 149–192. DOI: 10.1111/j.2517-6161.1975.tb01532.x.

Bujard, M. and G. Andersson (2024). "Fertility Declines Near the End of the COVID-19 Pandemic: Evidence of the 2022 Birth Declines in Germany and Sweden". In: *European Journal of Population* 40.1, pp. 1–21. DOI: 10.1007/s10680-023-09689-w.

Cantalini, S., S. Ohlsson-Wijk, and G. Andersson (2024). "Cohabitation and Marriage Formation in Times of Fertility Decline: The Case of Sweden in the Twenty-First Century". In: *European Journal of Population* 40.1, pp. 1–26. DOI: 10.1007/s10680-024-09703-9.

Chow, G. C. (1960). "Tests of Equality Between Sets of Coefficients in Two Linear Regressions". In: *Econometrica* 28.3, pp. 591–605. DOI: 10.2307/1910133.

Chu, C.-S. J., M. Stinchcombe, and H. White (1996). "Monitoring Structural Change". In: *Econometrica* 64.5, pp. 1045–1065. DOI: 10.2307/2171955.

Comolli, C. L., G. Neyer, G. Andersson, L. Dommermuth, P. Fallesen, M. Jalovaara, Á. K. Jónsson, M. Kolk, and T. Lappegård (2021). "Beyond the Economic Gaze: Childbearing During and After Recessions in the Nordic Countries". In: *European Journal of Population* 37.2, pp. 473–520. DOI: 10.1007/s10680-020-09570-0.

Davidson, R. and J. G. MacKinnon (2004). *Econometric Theory and Methods*. New York: Oxford University Press.

Englund, P. (1999). "The Swedish Banking Crisis: Roots and Consequences". In: *Oxford Review of Economic Policy* 15.3, pp. 80–97. DOI: 10.1093/oxrep/15.3.80.

Esping-Andersen, G. (1990). *The Three Worlds of Welfare Capitalism*. Princeton, NJ: Princeton University Press.

Gauthier, A. H. (2007). “The Impact of Family Policies on Fertility in Industrialized Countries: A Review of the Literature”. In: *Population Research and Policy Review* 26.3, pp. 323–346. DOI: 10.1007/s11113-007-9033-x.

Goldin, C. and L. F. Katz (2002). “The Power of the Pill: Oral Contraceptives and Women’s Career and Marriage Decisions”. In: *Journal of Political Economy* 110.4, pp. 730–770. DOI: 10.1086/340778.

Hawkins, D. M. and D. H. Olwell (1998). *Cumulative Sum Charts and Charting for Quality Improvement*. New York: Springer. DOI: 10.1007/978-1-4612-1686-5.

Hellstrand, J., J. Nisén, V. Miranda, P. Fallesen, L. Dommermuth, and M. Myrskylä (2021). “Not Just Later, but Fewer: Novel Trends in Cohort Fertility in the Nordic Countries”. In: *Demography* 58.4, pp. 1373–1399. DOI: 10.1215/00703370-9373618.

Hoem, B. (2000). “Entry into Motherhood in Sweden: The Influence of Economic Factors on the Rise and Fall in Fertility, 1986–1997”. In: *Demographic Research* 2.4, pp. 1–28. DOI: 10.4054/DemRes.2000.2.4.

Hoem, J. M. (1993). “Public Policy as the Fuel of Fertility: Effects of a Policy Reform on the Pace of Childbearing in Sweden in the 1980s”. In: *Acta Sociologica* 36.1, pp. 19–31. DOI: 10.1177/000169939303600102.

Human Fertility Database (2026). *Sweden: Period Summary Indicators*. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). Accessed 2026. URL: https://www.humanfertility.org.

Johnson, R. A. and M. Bagshaw (1974). “The Effect of Serial Correlation on the Performance of CUSUM Tests”. In: *Technometrics* 16.1, pp. 103–112. DOI: 10.1080/00401706.1974.10489147.

Jonung, L., J. Kiander, and P. Vartia (2009). “The Great Financial Crisis in Finland and Sweden: The Dynamics of Boom, Bust and Recovery, 1985–2000”. In: *The Great Financial Crisis in Finland and Sweden*. Ed. by L. Jonung, J. Kiander, and P. Vartia. Cheltenham: Edward Elgar.

Junod, S. W. and L. Marks (2002). “Women’s Trials: The Approval of the First Oral Contraceptive Pill in the United States and Great Britain”. In: *Journal of the History of Medicine and Allied Sciences* 57.2, pp. 117–160. DOI: 10.1093/jhmas/57.2.117.

Killick, R., P. Fearnhead, and I. A. Eckley (2012). “Optimal Detection of Changepoints With a Linear Computational Cost”. In: *Journal of the American Statistical Association* 107.500, pp. 1590–1598. DOI: 10.1080/01621459.2012.737745.

Kohler, H.-P. and D. Philipov (2001). “Variance Effects in the Bongaarts-Feeney Formula”. In: *Demography* 38.1, pp. 1–16. DOI: 10.1353/dem.2001.0004.

Künsch, H. R. (1989). “The Jackknife and the Bootstrap for General Stationary Observations”. In: *The Annals of Statistics* 17.3, pp. 1217–1241. DOI: 10.1214/aos/1176347265.

Lam, D. A. and J. A. Miron (1991). “Seasonality of Births in Human Populations”. In: *Social Biology* 38.1–2, pp. 51–78. DOI: 10.1080/19485565.1991.9988773.

Lam, D. A., J. A. Miron, and A. Riley (1994). “Modeling Seasonality in Fecundability, Conceptions, and Births”. In: *Demography* 31.2, pp. 321–346. DOI: 10.2307/2061888.

Lesthaeghe, R. (2010). “The Unfolding Story of the Second Demographic Transition”. In: *Population and Development Review* 36.2, pp. 211–251. DOI: 10.1111/j.1728-4457.2010.00328.x.

Lesthaeghe, R. and D. J. van de Kaa (1986). “Twee demografische transities?” In: *Bevolking: groei en krimp*. Ed. by D. J. van de Kaa and R. Lesthaeghe. Deventer: Van Loghum Slaterus, pp. 9–24.

Lorden, G. (1971). "Procedures for Reacting to a Change in Distribution". In: *The Annals of Mathematical Statistics* 42.6, pp. 1897–1908. DOI: 10.1214/aoms/1177693055.

Matysiak, A., T. Sobotka, and D. Vignoli (2021). "The Great Recession and Fertility in Europe: A Sub-National Analysis". In: *European Journal of Population* 37.1, pp. 29–64. DOI: 10.1007/s10680-020-09556-y.

Montgomery, D. C. (2009). *Introduction to Statistical Quality Control*. 6th ed. Hoboken, NJ: Wiley. ISBN: 978-0-470-16992-6.

Moustakides, G. V. (1986). "Optimal Stopping Times for Detecting Changes in Distributions". In: *The Annals of Statistics* 14.4, pp. 1379–1387. DOI: 10.1214/aos/1176350164.

Neyer, G., G. Andersson, J. Dahlberg, S. Ohlsson-Wijk, L. Andersson, and S. Billingsley (2022). "Fertility Decline, Fertility Reversal and Changing Childbearing Considerations in Sweden". In: *Stockholm Research Reports in Demography* 2022.8. Working paper. DOI: 10.17045/sthlmuni.19698442.v2.

Ní Bhrolcháin, M. (2011). "Tempo and the TFR". In: *Demography* 48.3, pp. 841–861. DOI: 10.1007/s13524-011-0033-4.

Nordic Statistics (2024). *Record Low Fertility in the Nordics*. Nordic Statistics Database. URL: https://www.nordicstatistics.org/news/record-low-fertility-in-the-nordics.

Ohlsson-Wijk, S. and G. Andersson (2022). "Disentangling the Swedish Fertility Decline of the 2010s". In: *Demographic Research* 47.12, pp. 345–358. DOI: 10.4054/DemRes.2022.47.12.

— (2025). "Swedish Fertility Developments Before, During and After the COVID-19 Pandemic". In: *European Journal of Population* 41.1, pp. 1–19. DOI: 10.1007/s10680-025-09744-8.

Oláh, L. S. and E. M. Bernhardt (2008). "Sweden: Combining Childbearing and Gender Equality". In: *Demographic Research* 19.28, pp. 1105–1144. DOI: 10.4054/DemRes.2008.19.28.

Page, E. S. (1954). "Continuous Inspection Schemes". In: *Biometrika* 41.1–2, pp. 100–115. DOI: 10.1093/biomet/41.1-2.100.

Politis, D. N. and J. P. Romano (1994). "The Stationary Bootstrap". In: *Journal of the American Statistical Association* 89.428, pp. 1303–1313. DOI: 10.1080/01621459.1994.10476882.

Preston, S. H., P. Heuveline, and M. Guillot (2001). *Demography: Measuring and Modeling Population Processes*. Malden, MA: Blackwell Publishers. ISBN: 978-1-55786-451-2.

RFSU (2026). *Elise Ottesen-Jensen — the pioneer behind RFSU*. RFSU official history. Accessed 2026. URL: https://www.rfsu.com/en/inspiration/about-rfsu/elise-ottesen-jensen-founder-rfsu/.

Rindfuss, R. R., D. K. Guilkey, S. P. Morgan, and Ø. Kravdal (2010). "Child-Care Availability and Fertility in Norway". In: *Population and Development Review* 36.4, pp. 725–748. DOI: 10.1111/j.1728-4457.2010.00355.x.

Ryder, N. B. (1964). "The Process of Demographic Translation". In: *Demography* 1.1, pp. 74–82. DOI: 10.1007/BF03208446.

Schmid, W. (1997). "CUSUM Control Schemes for Gaussian Processes". In: *Statistical Papers* 38, pp. 191–217. DOI: 10.1007/BF02925223.

Sobotka, T. (2004). "Is Lowest-Low Fertility in Europe Explained by the Postponement of Childbearing?" In: *Population and Development Review* 30.2, pp. 195–220. DOI: 10.1111/j.1728-4457.2004.00010.x.

Sobotka, T. and W. Lutz (2011). "Misleading Policy Messages Derived from the Period TFR: Should We Stop Using It?" In: *Comparative Population Studies* 35.3, pp. 637–664. DOI: 10.4232/10.CPoS-2010-15en.

Sobotka, T., V. Skirbekk, and D. Philipov (2011). "Economic Recession and Fertility in the Developed World". In: *Population and Development Review* 37.2, pp. 267–306. DOI: 10.1111/j.1728-4457.2011.00411.x.

Sonesson, C. and D. Bock (2003). "A Review and Discussion of Prospective Statistical Surveillance in Public Health". In: *Journal of the Royal Statistical Society: Series A* 166.1, pp. 5–21. DOI: 10.1111/1467-985X.00256.

Statistics Sweden (SCB) (2026). *Live Births (Monthly) and Fertility Indicators (Annual), Sweden*. SCB Statistical Database, series coverage 1900–2025. Accessed 2026. URL: https://www.scb.se.

Steiner, S. H., R. J. Cook, V. T. Farewell, and T. Treasure (2000). "Monitoring Surgical Performance Using Risk-Adjusted Cumulative Sum Charts". In: *Biostatistics* 1.4, pp. 441–452. DOI: 10.1093/biostatistics/1.4.441.

Sveriges Riksbank (1992). *Interest Rate 500% — the Krona Floats*. Riksbank Historical Timeline. URL: https://www.riksbank.se/en-gb/about-the-riksbank/history/historical-timeline/1900-1999/interest-rate-500--the-krona-floats/.

van de Kaa, D. J. (1987). "Europe's Second Demographic Transition". In: *Population Bulletin* 42.1, pp. 1–59.

Westoff, C. F. and N. B. Ryder (1977). *The Contraceptive Revolution*. Princeton, NJ: Princeton University Press.

Woodall, W. H. (2006). "The Use of Control Charts in Health-Care and Public-Health Surveillance". In: *Journal of Quality Technology* 38.2, pp. 89–104. DOI: 10.1080/00224065.2006.11918593.

Zeileis, A. (2006). "Implementing a Class of Structural Change Tests: An Econometric Computing Approach". In: *Computational Statistics & Data Analysis* 50.11, pp. 2987–3008. DOI: 10.1016/j.csda.2005.07.001.

List of Figures

List of Tables

A CUSUM algorithm

This appendix presents the full CUSUM detection pipeline applied to every indicator and case study in the thesis. The procedure transforms slope changes in the original series into approximate mean shifts that the standard CUSUM statistic can detect efficiently. Each step is described in prose in Section 3.3; the pseudocode below summarises the full sequence of operations.

Algorithm 1 CUSUM detection pipeline for slope changes in fertility series

Input: time series y_t, reference end index τ, seasonal structure (number of seasons s), reference value $k = 0.5$, significance level $\alpha = 0.05$
Output: alarm date t_a (or `None`)

Step 1: Baseline regression on reference period.
Fit OLS regression $y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{s-1} \gamma_j D_{j,t} + \varepsilon_t$ on $t \in [1, \tau]$, where $D_{j,t}$ are seasonal dummy variables.
Compute fitted values $\hat{y}_t$ for the full window.
Compute residuals $e_t = y_t - \hat{y}_t$.

Step 2: AR(1) pre-whitening.
Fit AR(1) model on $\{e_t\}_{t=1}^{\tau}$ to estimate $\hat{\phi}$.
Apply Ljung–Box test to AR(1) innovations to verify whiteness.
Filter full residual series: $u_t = e_t - \hat{\phi}\, e_{t-1}$.

Step 3: First-difference transformation.
Compute $z_t = u_t - u_{t-1}$ to convert slope changes into approximate mean shifts.
Standardise z_t using reference-period mean and standard deviation.

Step 4: Threshold calibration via Monte Carlo simulation.
Generate $N = 1{,}000$ synthetic series of length T from the fitted AR(1) model under H_0 (no change).
Apply Steps 1–3 to each synthetic series.
Compute $M^{(i)} = \max_t \max(C_t^{+,(i)}, C_t^{-,(i)})$ for each replication i.
Set threshold h as the $(1-\alpha)$ quantile of $\{M^{(i)}\}_{i=1}^{N}$.

Step 5: Two-sided CUSUM detection on observed series.
Initialise $C_0^+ = C_0^- = 0$.
for $t = 1, \ldots, T$ **do**
    $C_t^+ \leftarrow \max\big(0,\ C_{t-1}^+ + z_t - k\big)$
    $C_t^- \leftarrow \max\big(0,\ C_{t-1}^- - z_t - k\big)$
    if $C_t^+ > h$ or $C_t^- > h$ **then**
        return $t_a = t$
    end if
end for
return `None` ▷ no alarm within window

The pipeline is implemented in Python using `numpy`, `statsmodels` for AR(1) estimation, and `scipy.stats` for the Ljung–Box test. The full implementation, including data preprocessing, threshold calibration, and figure generation, is available from the authors on request.

B Sensitivity to reference window length

To assess the sensitivity of the main results to the choice of reference window length, the CUSUM pipeline was re-run on TFR and tempo-adjusted TFR for all three case studies using symmetric windows of ±10, ±12, ±15 (baseline), and ±20 years. Results are reported in Table 4.

For TFR, alarm dates are stable across window lengths from ±10 to ±15 years: the oral contraceptives episode is detected in 1967–1968 (delay 2–3 years), the 1990s economic crisis in 1992–1993 (delay 1–2 years), and the post-2010 decline consistently in 2011 (delay 1 year). Detection delays vary by at most one year within this range, confirming that the baseline choice of ±15 years does not drive the main findings.

At ±20 years, detection deteriorates in two cases: the post-2010 TFR alarm shifts from 2011 to 2022 (delay 12 years), and the 1990s TFR delay increases from 2 to 4 years. This is attributable to the longer reference period capturing earlier structural changes in the fertility series, which inflates the estimated AR(1) coefficient ($\hat{\phi} > 0.92$ in both cases) and causes the pre-whitening step to remove a larger share of the true signal. This finding supports the use of a reference window short enough to be approximately homogeneous with respect to the prevailing fertility regime.

For tempo-adjusted TFR, the qualitative pattern is unchanged across all window lengths: no alarm is produced for the oral contraceptives or post-2010 episodes regardless of window choice, while the 1990s crisis is detected with delays of 4–5 years at ±10 to ±15 but not at ±20. The robustness of the non-detection result for tempo-adjusted TFR in Cases I and III strengthens the interpretation that these declines are predominantly driven by postponement rather than quantum change.

Overall, the main conclusions of the analysis, that observed TFR detects within 1–3 years, that tempo-adjusted TFR fails to detect in the contraceptive and post-2010 episodes, and that monthly births never detect, are not sensitive to the window length within the ±10 to ±15 range.

Table 4: Robustness check: CUSUM results under varying window lengths. The baseline specification uses ±15 years; results are shown for ±10, ±12, ±15, and ±20 years. A dash indicates no alarm within the window.

Case study	Indicator	Window	ref_len	Alarm date	Delay (yr)
Oral contraceptives	TFR	±10	10	1967-01-01	2
Oral contraceptives	Tempo-TFR	±10	10	–	–
Oral contraceptives	TFR	±12	12	1967-01-01	2
Oral contraceptives	Tempo-TFR	±12	12	–	–
Oral contraceptives	TFR	±15	15	1968-01-01	3
Oral contraceptives	Tempo-TFR	±15	15	–	–
Oral contraceptives	TFR	±20	20	1968-01-01	3
Oral contraceptives	Tempo-TFR	±20	20	–	–
Economic crisis 1990s	TFR	±10	10	1992-01-01	1
Economic crisis 1990s	Tempo-TFR	±10	10	1995-01-01	4
Economic crisis 1990s	TFR	±12	12	1993-01-01	2
Economic crisis 1990s	Tempo-TFR	±12	12	1995-01-01	4
Economic crisis 1990s	TFR	±15	15	1993-01-01	2
Economic crisis 1990s	Tempo-TFR	±15	15	1996-01-01	5
Economic crisis 1990s	TFR	±20	20	1995-01-01	4
Economic crisis 1990s	Tempo-TFR	±20	20	–	–
Post-2010	TFR	±10	10	2011-01-01	1
Post-2010	Tempo-TFR	±10	10	–	–
Post-2010	TFR	±12	12	2011-01-01	1
Post-2010	Tempo-TFR	±12	12	–	–
Post-2010	TFR	±15	15	2011-01-01	1
Post-2010	Tempo-TFR	±15	15	–	–
Post-2010	TFR	±20	20	2022-01-01	12
Post-2010	Tempo-TFR	±20	20	–	–

Master’s Theses in Mathematical Sciences 2026:E57
ISSN 1404-6342

LUTFMS-3556-2026

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lu.se/