Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Background growth, count noise and the first accepted year of a patent take-off rule

Andreas Peters

a.peters81@icloud.com

Rosenheim Technical University of Applied Sciences

Research Article

Keywords:

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10982539/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Background growth, count noise and the first accepted year of a patent take-off rule

Andreas Peters[1*†]

[1]Faculty of Applied Natural Sciences and Humanities, Rosenheim Technical University of Applied Sciences, Rosenheim, Germany.

Corresponding author(s). E-mail(s): a.peters81@icloud.com;
[†]ORCID 0009-0006-0059-1420

Abstract

Patent take-off rules assign retrospective years by combining growth, level, persistence and scale conditions. One fixed five-condition rule is examined here. For a permanent log-growth step $\boldsymbol{\delta}$ on a monotone background of rate $\boldsymbol{g}$, its level condition at offsets $\boldsymbol{d = -1, 0, 1}$ reduces at large counts to $\boldsymbol{4g + (d + 2)\delta \geq \log 1.5}$. Background growth can therefore advance acceptance relative to the imposed change although the growth contrast removes constant growth. On European Patent Office reporting-year counts for 35 fields and 175 geographies in 2011–2025 it accepts 201 events, where the level ratio has the smallest relative slack in 132 and the count floor in 52. Removing the ratio advances 56 first dates and adds 191 series, removing the floor advances 23 and adds 592, so both affect timing and membership in different proportions. Wider growth windows change the earliest-labelled geography in 24 of 33 fields with events under both definitions,; applicant and inventor attribution produce different calendars across five priority-year populations and a second field taxonomy. Fitted scenarios generate sizeable calendars: with parameters from 2011–2015 alone, 1,002 series carrying 183 observed events yield a scenario mean of 256.5, so calendar size is not evidence of discrete change, though no individual event is classified. Count perturbations reproduce an event's year in 45 per cent of replicates on the main panel and 41 on a second. Source version, window completion, pre-period growth and criterion margins make such labels interpretable; accuracy against independently dated onsets remains untested.

Keywords: Patent indicators, Technological take-off, Temporal measurement, Moving sums, European patent applications, Count data

MSC 2020: 62M10. JEL: C43, O34.

1 Introduction

Patent indicators are used to compare technological development across fields and countries, although their interpretation depends on counting conventions and on the relation between patenting and innovation (Griliches 1990; OECD 2009; Nagaoka et al. 2010). Indicators of new and fast-growing technologies combine growth with attributes such as novelty, coherence and persistence (Rotolo et al. 2015; Carley et al. 2018; Porter et al. 2019). Once an indicator also assigns a year, its construction becomes a dating problem. A rule using observations after the candidate year produces a retrospective label. Its first accepted year need not coincide with the beginning of the change that motivated the label.

One concrete example shows what the distinction costs. In the EPO reporting-year panel studied here, one fixed rule first labels Austria and Belgium in audio-visual technology in 2016. Increasing its post-growth window from three to four intervals changes that result to Switzerland in 2018. Across the 33 fields with events under both definitions, the set of earliest-labelled geographies changes in 24. Some changes concern entry into the labelled population; others change dates within it. Distinguishing the two is necessary before interpreting who came first.

In the worked instrument, a moving-sum growth contrast is combined with conditions on post-period growth, a median level ratio, positive annual changes and minimum counts. Moving-sum change-point methods supply the relevant linear contrast (Meier et al. 2021); first acceptance by the full conjunction also depends on nonlinear criteria and on chronological exclusions. A flat, noiseless benchmark isolates this mechanism but leaves open how it behaves on the growing, declining and irregular paths encountered in patent data.

RQ1. How do background growth, scale and variation change the first year accepted by the fixed rule relative to an imposed change?

RQ2. On observed patent-count panels, how close are accepted events to the individual thresholds, and how sensitive are their calendars to the rule's settings, counting conventions and stipulated count perturbations?

Four parts make up the contribution, and each carries its own reach. Analysis supplies the first. On the specified monotone path, the large-count median ratio contains four years of background growth. Its threshold can therefore be crossed earlier as prior growth rises, even though the growth contrast itself removes constant growth. Some nonlinear condition in the conjunction therefore changes the first accepted year, and the mechanism is identified rather than assumed.

Separating the calendar effects of individual conditions gives the second. Among the four continuous criteria, the level ratio and the count floor most often have the smallest relative slack. A chronological rescan with one condition removed gives a distinct comparison: removing the ratio advances 56 of the 201 original first dates and adds 191 series; removing the floor advances 23 dates and adds 592 series. Both removals affect timing and membership. In this panel, removing the ratio advances

more original first dates, whereas removing the floor adds more newly labelled series. A margin ranking alone does not establish these counterfactual calendar effects.

Attribution supplies the third. Across five related priority-year patent populations from one source, switching between applicant and inventor attribution changes fields' earliest-labelled geographies. The population-specific comparisons report their own denominators; together they show that counting conventions materially affect the calendar. They are not paired estimates of the window effects on the main reporting-year panel, and the aggregate data do not establish identity of the underlying document sets.

Scenario sensitivity supplies the fourth. Four fitted constant-mean-growth scenarios produce mean calendars of 103 to 194 events; one simulation interval contains the observed 201. Sizeable calendars can therefore arise from count variation around the stipulated mean paths. Because the scenarios are fitted to the same series used for comparison, the overlap is not a calibrated test of statistical indistinguishability.

One partial check against dated cases was attempted and returned nothing usable. Eight sources that date a shift in patenting were screened for a field–geography series with a comparable date inside the eligible calendar; twelve case and exclusion records are archived, and no strictly comparable case survived, the nearest candidate lacking an established mapping to a single field. Absence of a usable case in this screen is not evidence that none exists.

Each result has a defined scope. Conditional on the specified path and asymptotic in count scale, the analytical result reaches no further. Ablation describes this rule's conjunction on the observed panel and identifies no cause of a technological event, while fitted scenarios neither classify individual events as noise nor bound a false-discovery fraction. None of the empirical comparisons tests the conditional theorem or establishes historical dating accuracy.

Measurement research distinguishes an indicator's operational definition from evidence supporting its intended interpretation (Hand 1996; Adcock and Collier 2001; Borsboom et al. 2004; Kane 2013; Jacobs and Wallach 2021). Definition sensitivity has also been documented for alternative populations of highly cited researchers (Frietsch et al. 2025). For take-off calendars, the additional issue is temporal: an observation window and the background path jointly influence an assigned year. The present rule is a worked instrument inherited from an earlier workflow, not a published standard or a calibrated detector. Its numerical results apply to this rule; transfer to another indicator requires examining that indicator's own statistics and thresholds.

No independent chronology of technological onsets is available for the field–geography series. Accordingly, the empirical analysis assesses operating behaviour and sensitivity, rather than accuracy against technological history. The practical output is a richer event record containing the counting unit, source version, window-completion year, pre-period growth and criterion margins. Those quantities make a calendar interpretable without assigning it a validity that the available evidence cannot establish.

Table 1 The unchanged worked rule. Every condition must hold for an eligible candidate. Constants specify the instrument; they are not estimated optima.

Criterion	Acceptance condition
Post-period growth	$r_{ik}^{\text{post}} \geq c_1$, where $c_1 = 0.1133286853 \simeq \log(1.12)$
Growth increase	$D_{ik} \geq c_2$, where $c_2 = 0.1$
Level ratio	$m_{ik}^{\text{post}} / \max\{m_{ik}^{\text{pre}}, 1\} \geq 1.5$
Positive changes	At least two of $N_{ik} - N_{i,k-1}$, $N_{i,k+1} - N_{ik}$ and $N_{i,k+2} - N_{i,k+1}$ are strictly positive
Count floor	$m_{ik}^{\text{post}} \geq 20$

2 The rule, the counting unit and the three dates

2.1 A fixed five-criterion rule

Let $N_{it} \geq 0$ denote the annual count for series i in year t, and put $\ell_{it} = \log(1+N_{it})$ and $z_{it} = \ell_{it} - \ell_{i,t-1}$. In the empirical application i indexes a technology-field–geography pair. The pseudocount accommodates zeros without treating an absent observation as a zero. Calendar eligibility of candidate year k requires all eight observations from $k-5$ through $k+2$. Within that block define

$$r_{ik}^{\text{post}} = \frac{\ell_{i,k+2} - \ell_{i,k-1}}{3}, \qquad r_{ik}^{\text{pre}} = \frac{\ell_{i,k-1} - \ell_{i,k-5}}{4}, \qquad D_{ik} = r_{ik}^{\text{post}} - r_{ik}^{\text{pre}}, \quad (1)$$

and the level medians $m_{ik}^{\text{pre}} = \text{median}\{N_{i,k-5}, \ldots, N_{i,k-1}\}$ and $m_{ik}^{\text{post}} = \text{median}\{N_{ik}, N_{i,k+1}, N_{i,k+2}\}$. Candidates are evaluated chronologically. A year is accepted when the five conditions in Table 1 hold and no previous accepted event excludes it; acceptance at k excludes $k+1, \ldots, k+5$, and $k+6$ is eligible again. Because of this cooldown a candidate can satisfy all five inequalities without becoming an accepted event, and both statuses are kept apart throughout.

For a monotone path the pre- and post-period medians occur at $k-3$ and $k+1$, four years apart. Irregular paths need not have their medians at these times. Keeping the two apart separates the exact background result from descriptive calculations on observed counts. Positivity is a discrete condition, and chronological cooldown further separates qualification from acceptance.

Constants and windows are retained from the supplied earlier workflow; no empirical calibration of this exact conjunction is claimed. Published indicators of new technologies provide precedents for its components rather than for the complete rule. For example, Carley et al. (2018) compare base and active periods with growth and minimum-record conditions, while Porter et al. (2019) combine such information in a single score. Patent forecasting also uses growth curves and supervised methods (Daim et al. 2006; Kyebambe et al. 2017). The purpose here is to characterise the chosen instrument, not to establish that these alternatives share its dating formula.

2.2 Applications by field and applicant geography

Annual European patent applications form the primary panel for 35 WIPO technology fields and 175 reporting geographies over 2011–2025: 6,125 series and 91,875 counts. Its recorded source is the EPO Data Desk report released in March 2026, with an extraction date of 2 February 2026 (European Patent Office 2026). Counts cover direct European applications and international applications entering the European phase in the reporting year, attributed to the first-named applicant's residence (European Patent Office n.d.; Schmoch 2008).

Reporting year differs from priority year. Entry into the European phase is generally due within 31 months of the relevant filing or priority date (European Patent Office 2020). Mixing that route with direct European filings therefore combines different temporal conventions; a fixed correction in years would not recover invention dates. Route use also differs across applicants and fields (Frietsch et al. 2013). Priority-year and inventor-location counts provide a different approximation to inventive activity (OECD 2009; de Rassenfosse et al. 2013). Family definitions and office coverage further affect the unit of observation (Martínez 2011; Neuhäusler and Frietsch 2013; Dechezleprêtre et al. 2017; Frietsch and Schmoch 2010). Section 5.5 repeats the rule on five OECD priority-year populations under applicant and inventor attribution.

Applicant residence need not locate the inventors: ownership and inventive activity can be geographically separated (OECD 2009). Likewise, the 35 broad fields combine heterogeneous technologies, and their trajectories depend on classification choices (Benner and Waldfogel 2008; Lafond and Kim 2019). An accepted event consequently describes a particular reported aggregate, whose innovations may differ substantially in novelty and quality (Verhoeven et al. 2016; Higham et al. 2021).

2.3 Nominal year, window completion and source version

Each event has three temporal attributes. Its nominal year is k; its window is complete at $k+2$; and its counts belong to a specified source release. With observations for 2011–2025, eligible nominal years are 2016–2023. Truncating the current panel reproduces shorter windows, not earlier database releases or the information then available (Lafond and Kim 2019).

Chronological scanning begins in 2016 without a previous accepted event. Earlier events and their cooldown effects cannot be recovered from this panel. Thus a first label in 2016 is first within the observed scan, and cross-series comparisons inherit this left-censoring. Events whose windows would end after 2025 are ineligible; this mechanical boundary alone does not explain variation among the eligible years.

3 First acceptance on structured backgrounds

3.1 The moving-sum contrast and its response

Telescoping the endpoint differences in Eq. (1) gives

$$D_{ik} = \frac{1}{3}\sum_{h=0}^{2} z_{i,k+h} - \frac{1}{4}\sum_{h=1}^{4} z_{i,k-h}, \tag{2}$$

a zero-sum linear contrast of annual log growth. With window lengths $p = 3$ and $q = 4$ it coincides, up to normalisation and indexing, with the asymmetric moving-sum statistic of Meier et al. (2021, Eq. (13)); general-window identities and their derivations are in Supplementary Information **??**. The corresponding symmetric moving-sum construction is established in Eichinger and Kirch (2018). For a permanent increase $\Delta > 0$ in log growth beginning at τ, the incremental response at offset $d = k - \tau$ is $\Delta\psi(d)$ with

$$\big(\psi(-2), \psi(-1), \psi(0), \psi(1), \psi(2), \psi(3)\big) = \big(\tfrac{1}{3}, \tfrac{2}{3}, 1, \tfrac{3}{4}, \tfrac{1}{2}, \tfrac{1}{4}\big), \qquad \psi(d) = 0 \text{ otherwise}, \tag{3}$$

which counts the changed growth intervals entering the two windows. Two properties of the contrast carry over to every background considered below. A constant growth rate is removed exactly, so the contrast of a step on any background equals the contrast of the background plus $\Delta\psi(d)$. And a response before τ is retrospective window overlap, not foresight, because the post-window of candidate $\tau - 1$ already contains observations from τ onward. Linearity of the contrast transfers neither to the medians nor to the sign count nor to the sequential acceptance, which is where the dating results come from.

3.2 Flat baseline

Suppress the series index and take a flat pre-change level $N_t = N_0 \geq 20$ for $t < \tau$ with a permanent change from zero to $\Delta > 0$ in log growth, $N_t = (1+N_0)\exp\{\Delta(t-\tau+1)\}-1$ for $t \geq \tau$. With the finite-level threshold $Q(N_0) = \log\{(1.5N_0 + 1)/(N_0 + 1)\}$ and

$$T_{-1} = Q(N_0), \qquad T_0 = \frac{Q(N_0)}{2}, \qquad T_1 = \max\Big\{c_1, \frac{4c_2}{3}, \frac{Q(N_0)}{3}\Big\}, \tag{4}$$

the first accepted year is $\tau - 1$ for $\Delta \geq T_{-1}$, τ for $T_0 \leq \Delta < T_{-1}$, $\tau + 1$ for $T_1 \leq \Delta < T_0$, and there is no acceptance for $0 < \Delta < T_1$; no first acceptance occurs at $\tau - 2$, at $\tau + 2$ or later (Proposition **??**, proof in Supplementary Information **??**). Figure 1 shows the ladder for $N_0 = 100$. Read as a classification, the one-year advance for large steps is part of the definition; read as the start of an acceleration, it is an offset that requires disclosure. Either way it is a property of a path without growth before τ, and Sections 3.3 and 3.4 remove that assumption.

3.3 Constant background growth

Now let the background grow at a constant log rate g before and after the change, $\ell_t = L + g(t - \tau + 1) + \delta \max\{t - \tau + 1, 0\}$, with $A = e^L$ the pseudocount level at $\tau - 1$

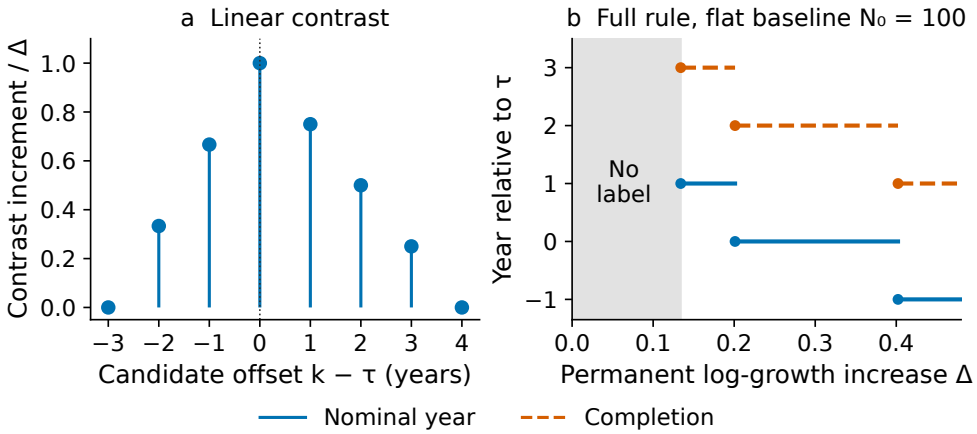


Fig. 1 Analytical time construction on a flat baseline, not observed patent data. Panel a gives the incremental response of the contrast to a unit change in growth. Panel b shows first acceptance and the last required observation year for the full rule at $N_0 = 100$. Each threshold belongs to the interval on its right. The contrast peaks at the change year, whereas first acceptance depends on the size of the change.

and $N_t = e^{\ell_t} - 1$ real-valued. The growth contrast of this path is exactly $\delta\psi(d)$ for every g, so the gap condition is untouched by g. Everything else moves. Because both windows now contain growth, the post-period growth rate becomes $g + \delta[d+3]_0^3/3$, and the level medians sit at the middle-time observations $\tau + d - 3$ and $\tau + d + 1$ whenever the path is monotone. Their ratio therefore contains four years of background growth plus $(d + 2)$ years of the added step.

Proposition 1 (First acceptance on a constant-growth background) *Let $g \geq 0$, $A - 1 \geq 20$ and $\delta > 0$ on the path above, with non-negative counts throughout every required block, complete calendar blocks for candidates $\tau - 2, \ldots, \tau + 3$, chronological evaluation and no inherited cooldown. Write E_d for the joint satisfaction of the five conditions at offset d. Then*

$$E_{-2} \iff g > 0,\ \delta \geq 3c_2,\ g + \tfrac{\delta}{3} \geq c_1,\ A - 1 \geq 1.5\max\{Ae^{-4g} - 1, 1\}, \tag{5}$$

$$E_{-1} \iff \tfrac{2\delta}{3} \geq c_2,\ g + \tfrac{2\delta}{3} \geq c_1,\ Ae^{g+\delta} - 1 \geq 1.5\max\{Ae^{-3g} - 1, 1\},\ Ae^{g+\delta} - 1 \geq 20, \tag{6}$$

$$E_0 \iff \delta \geq c_2,\ g + \delta \geq c_1,\ Ae^{2g+2\delta} - 1 \geq 1.5\max\{Ae^{-2g} - 1, 1\},\ Ae^{2g+2\delta} - 1 \geq 20, \tag{7}$$

$$E_1 \iff \tfrac{3\delta}{4} \geq c_2,\ g + \delta \geq c_1,\ Ae^{3g+3\delta} - 1 \geq 1.5\max\{Ae^{-g} - 1, 1\}, \tag{8}$$

the first accepted year is $\tau + \min\{d : E_d\}$ over $d \in \{-2, -1, 0, 1\}$, and no first acceptance occurs at $\tau + 2$ or later. If none of these conditions holds, there is no acceptance. Equations (6) *and* (7)*, and only these two, hold for every real g, including declining backgrounds, provided $g + \delta > 0$;* (5) *and* (8) *require $g \geq 0$.*

Supplementary Information **??** gives the proof, which evaluates the five conditions offset by offset; the floor holds because every relevant post-period median is at least

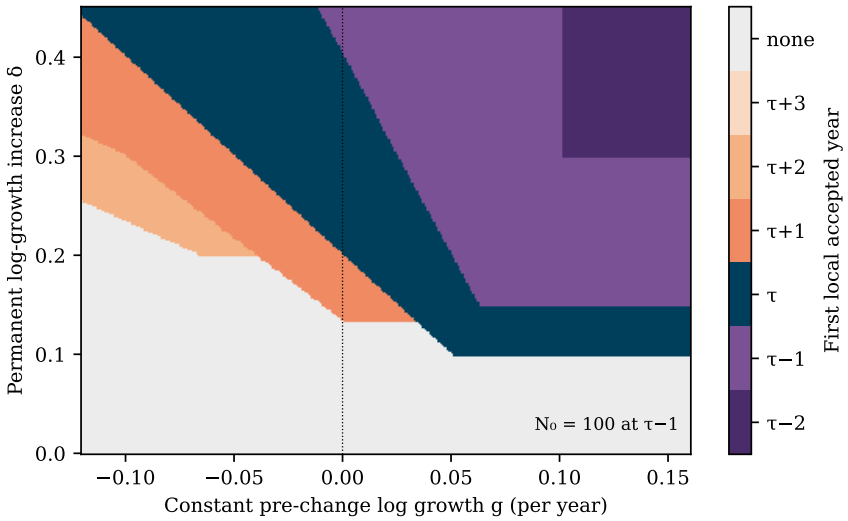


Fig. 2 First accepted year of the full rule on constant-growth backgrounds at level $A-1=100$, evaluated numerically from the defined continuous count path over a grid of pre-change growth g and added step δ. Along $g=0$ the flat ladder of Fig. 1 reappears. Regions are constructed benchmarks, not patent observations.

$A-1$, and the count of positive changes is two or three except at $\tau-2$ on a non-growing background. Its content is easier to see in the large-count limit, where the pseudocount can be dropped.

Corollary 1 (Large-count thresholds) *For $g \geq 0$ and $A \to \infty$, E_d holds for $d \in \{-1, 0, 1\}$ exactly when*

$$\delta \geq T_d(g) = \max\Big\{ \frac{c_2}{\psi(d)},\ \frac{3(c_1-g)}{[d+3]_0^3},\ \frac{\log 1.5 - 4g}{d+2} \Big\}, \tag{9}$$

and E_{-2} holds exactly when $g \geq \frac{1}{4}\log 1.5 \simeq 0.1014$ and $\delta \geq \max\{3c_2, 3(c_1-g)\}$. The level term is the only one that depends on the offset through $(d+2)$: it stops binding at the change year once $g \geq (\log 1.5 - 2c_2)/4 \simeq 0.0514$, and at $\tau-1$ once $g \geq (\log 1.5 - 1.5c_2)/4 \simeq 0.0639$.

Background log growth contributes $4g$ to the large-count ratio. The level condition at $d=0$ ceases to dominate the gap threshold at $g \simeq 0.0514$; at $d=-1$ the corresponding crossover is $g \simeq 0.0639$. Both are annual log rates, equivalent to ordinary growth of approximately 5.27 and 6.60 per cent. At sufficiently large counts and $g \geq 0.0639$, a step of $\delta = 0.15$ is first accepted at $\tau-1$. A two-year advance requires both stronger background growth and a step meeting $\delta \geq 0.30$. The finite-count regions in Figure 2 use the path itself. Declining backgrounds can have later first dates; their full rule is evaluated numerically, beyond the two declining-background equivalences stated in the proposition.

Supplementary Information **??** compares the finite-count conditions with the controlled trend backgrounds and states the numerical boundary treatment. The large-count corollary is restricted to its non-negative-growth domain; agreement outside that domain cannot be used to attribute discrepancies solely to pseudocounts.

3.4 Smooth acceleration

Smooth acceleration is a second departure from the flat case, and it needs no step at all. Let annual log growth rise linearly, $z_t = a + c\,(t - t_0)$ with $c > 0$, so that the log level is a quadratic in time.

Proposition 2 (Constant contrast under smooth acceleration) *On a path with $z_t = a + c(t - t_0)$ the contrast equals $D_k = 3.5\,c$ for every candidate k. Consequently the gap condition of the rule holds at every candidate when $c \geq c_2/3.5 \simeq 0.0286$ and at none when c is smaller; the first accepted year on such a path is determined by the growth, level and floor conditions together with the first eligible candidate of the scan, and not by any event on the path.*

Averaging z places the post-window centre at $k + 1$ and the pre-window mean at $k - 2.5$, so their difference is $3.5c$. For a flat start and $c = 0.03$, the illustrative scan in Supplementary Information **??** first accepts in 2016 at levels 20, 100 and 500. It shows how scan initialisation can matter under the stated path, without identifying the reason for an initial-year concentration in observed data.

3.5 Fluctuation, first passage and near-threshold decisions

Observed series are neither flat nor smooth. Write the log level as a specified background plus a deviation, $\ell_t = b_t + \varepsilon_t$. The contribution of the deviation to the contrast is exact and small in support,

$$D_k(\varepsilon) = \tfrac{1}{3}\varepsilon_{k+2} - \tfrac{7}{12}\varepsilon_{k-1} + \tfrac{1}{4}\varepsilon_{k-5}, \qquad |D_k(\varepsilon)| \leq \tfrac{7}{6}\,\eta \text{ if } |\varepsilon_t| \leq \eta, \tag{10}$$

and the post-period growth rate moves by at most $2\eta/3$. Independent level deviations of variance σ_ℓ^2 give the contrast variance $37\sigma_\ell^2/72$, independent growth deviations of variance σ^2 give $7\sigma^2/12$, and the two are not interchangeable, since differencing independent level errors produces dependent growth errors; both are properties of a stipulated disturbance rather than estimates from the panel (Supplementary Information **??**).

Both bounds concern the growth criteria. First acceptance combines correlated candidate windows, median ratios, signs and cooldown, so a small change in one component can alter a nominal year. Relative margins are therefore reported as diagnostics, with condition-removal scans distinguishing proximity from influence on the first date. Simulations in Section 6 evaluate the full rule under stipulated count laws.

4 Data and empirical design

4.1 Source, populations and what is recorded

According to the supplied preparation record, 35 EPO field exports contain 181 source labels each. Aggregates, the overlapping EU27 group and the provenance categories Unknown and Not Applicable are excluded. Three historical labels are retained only in a separate 178-geography specification, without reallocation to successor states. Of the 6,125 primary series, 2,971 are zero throughout. Explicit zeros remain observations; missing values are not converted to zeros. The archived panel and preparation code distinguish source-derived counts from subsequent transformations. An independent check against the current official workbook corroborates the reconciliation of 16,275 overlapping 2016–2025 cells, including the five recorded discrepancies (Supplementary Information **??** and **??**). Diagnostic substitution changes no accepted event. A fresh export of all 35 fields taken on 9 September 2026 reproduces every one of the 91,875 panel cells without exception, and rescanning it returns the same 268 qualifications and 201 accepted events. Being a later retrieval rather than the archived one, it carries no original checksum: none of its files carries the checksum recorded in the panel, so it corroborates the values while leaving the original file chain unrecovered. A separate export of all fields jointly exceeds the sum of the 35 field exports in 626 of 2,715 geography–year cells, which is a property of the source rather than of the panel and is reported here so that the agreement on the 91,875 cells is not read as an accounting identity.

All 49,000 eligible candidates retain the five criterion values and pass states, qualification before cooldown, acceptance after cooldown, and the nominal and completion years. Margins are recorded in each criterion's own units. For the four continuous criteria, relative slack is the value divided by its threshold minus one; the smallest identifies the tightest continuous criterion. Positivity is reported separately as two or three positive changes. Such a convention permits comparison of proportional slack but does not establish which condition caused the first accepted date. Removing one condition at a time and rescanning chronologically supplies a separate counterfactual diagnostic in Section 5.1. The 5,298 candidates with a pre-period median of at least 20 form an additional scale-restricted denominator. Neither denominator represents independent trials.

4.2 Definition sensitivity as membership and redating

Alternative specifications change one threshold, window or cooldown setting at a time, using the common complete calendar 2016–2022 and the same initial state. Event-set intersection and union give Jaccard agreement. Series-level comparisons separate unchanged first dates, earlier and later first dates, loss of all events and gain of a first event. Field-level comparisons use the set of geographies tied for the earliest label; within-field pair comparisons retain ties and distinguish preserved, reversed and newly created orderings. Every comparison states its common population. Supplementary Information **??** also reports a local two-mean segmentation of the seven growth intervals in an accepted window, used as a descriptive comparator.

4.3 Controlled increments on observed backgrounds

Known change dates are introduced in complete 13-level blocks $\tau - 7, \ldots, \tau + 5$, with anchors $\tau = 2018, 2019, 2020$. Selection requires a pre-period median and $N_{i,\tau-1}$ of at least 20, no full-rule qualification at offsets $-2, \ldots, +3$, and no accepted event in $\tau - 7, \ldots, \tau - 3$. Each selected block supplies three paired backgrounds: its observed log-count path, a constant at $\ell_{i,\tau-1}$, and a trend through that level using mean growth over the four preceding intervals. Adding $\delta \max\{t - \tau + 1, 0\}$ on the log scale for $\delta \in \{0, 0.10, 0.15, 0.20, 0.30, 0.40\}$ supplies the controlled change. The fixed rule scans offsets $-2, \ldots, +3$ with local cooldown reset. Fractional constructed values remain distinct from observed counts. Exact rational coefficients resolve constructed growth-threshold equalities; logarithms, exponentials and numerical path evaluation remain floating-point computations (Supplementary Information **??**).

Stratification by pre-period slope, scale and anchor year was added after the initial results. Slope classes have boundaries $-0.05, 0, 0.05, 0.10$. Paired comparisons of observed, trend and constant backgrounds describe how dates change when path components are replaced; they do not identify a causal share attributable to slope. A further paired comparison retains only the two growth criteria. Repeated series and overlapping windows receive no independence-based confidence intervals.

4.4 Count noise around known mean paths

Unselected simulations address labels produced without an imposed step. For $t = \tau - 10, \ldots, \tau + 10$ and $x = t - \tau + 1$, the mean is

$$\mu_t = N_0 \exp\{gx + \tfrac{1}{2}a(x + 9)^2 + \delta \max(x, 0)\}.$$

Backgrounds $(g, a) = (0, 0), (0.03, 0), (0, 0.01)$ are flat, growing and smoothly accelerating. Values $N_0 = 20, 100, 500$ are scale parameters: they equal $\mu_{\tau-1}$ only when $a = 0$; in the acceleration cells $\mu_{\tau-1} = N_0 e^{40.5a}$. Steps are $\delta = 0, 0.15, 0.30$. Counts are Poisson or negative binomial with variance $\mu + \alpha\mu^2$, where $\alpha = 0.03, 0.10$, generated through a gamma–Poisson mixture (Hausman et al. 1984; Cameron and Trivedi 2013). A descriptive moment estimate around log-linear paths for 670 observed series with mean at least 20 has median 0.013 and upper quartile 0.031. It motivates the scenarios without validating their stochastic law.

Each of 81 cells contains 2,000 seeded sequences. A local scan covers $-2, \ldots, +3$; a chronological scan covers $-5, \ldots, +8$ with cooldown. The growth-only comparator uses identical draws. Monte Carlo standard errors are at most 1.12 percentage points per cell. A separate latent AR(1) scenario uses standard deviation 0.15 and autocorrelation 0 or 0.6 on the growing mean at $N_0 = 100$. Level jumps, three-year growth phases and logistic transitions provide further deterministic and noisy change-shape checks (Supplementary Information **??** and **??**).

4.5 No-change scenarios, date stability and a whole-series comparator

Three calendar scenarios fit a log-linear Poisson mean to each non-empty series over 2011–2025, then draw counts with Poisson, pooled negative-binomial or series-specific negative-binomial dispersion. A fourth replaces the fit for an observed event series with a trend estimated before its accepted year; other series retain their full-period fits. Each scenario supplies 500 replicate panels and applies the unchanged chronological scan. Full-period fits may absorb actual changes; the fourth scenario depends on the observed event selection. Neither construction establishes a direction of bias or a formal bound on false discoveries.

At event level, 2,000 continuation draws estimate the frequency of qualification at the observed year using a pre-period trend and dispersion selected by the mean of the five local pre-window counts. The calendar continuation uses full-period mean-count classes, so these are distinct scenarios. Frequencies are pointwise diagnostics, uncorrected for selection and multiplicity; they are not calibrated hypothesis-test p values.

Date stability uses 2,000 negative-binomial perturbations around each observed path, with the original counts as means and dispersion classified by the five local pre-window counts. Rescanning records the first accepted year, agreement within one year and absence of any label. Parametric perturbation of that kind holds the path fixed and measures sensitivity to a stipulated count law (Efron and Tibshirani 1993; Davison and Hinkley 1997).

Fitting a continuous broken stick, $\ell_{it} = a + b_1(t - t_0) + b_2 \max\{0, t - t_0\}$, is fitted by least squares across knots $t_0 = 2013, \ldots, 2023$; the minimum residual sum of squares selects the knot and $t_0 + 1$ supplies the comparison year. Structural-break methods motivate this comparator (Bai and Perron 1998, 2003), but the uncorrected fixed-knot F reference is descriptive after knot selection. Its date is not independent ground truth.

4.6 Computation and verification

Python analyses retain the fixed thresholds and separate source-derived, constructed and simulated values. An independently written scan checks the primary candidates and specification variants. Supplementary Information **??** records boundary treatment; **??** adapts the GLOBAL bibliometric reporting checklist to this patent study (Ng et al. 2025). Illustrative series follow the ordering rules stated in their captions. Simulation seeds, parameter grids and analysis outputs are retained, with post-result additions identified separately from the original analysis.

5 Results on European patent-application counts

5.1 Accepted events and criterion margins

Of 49,000 eligible candidates, 268 qualify and 201 are accepted after cooldown, in 201 series, 37 geographies and 34 fields. Only 112 events have a pre-period median of at least 20; the remaining 89 reach the post-period floor from a lower base. Pre-period endpoint log growth ranges from -0.229 to 0.549, with quartiles -0.033, 0.013 and

Table 2 Tightest continuous criterion and distance from threshold for the 201 accepted events. The tightest criterion is the one with the smallest distance relative to its threshold; distances are in the unit of each condition (log growth per year for the growth conditions, ratio points for the level ratio, applications for the floor). Positivity is met with the minimum of two positive changes in 110 events and is not ranked. Source: EPO Data Desk report, extraction of 2 February 2026; calculations in the replication materials.

Criterion	Tightest in events	Distance from threshold, quantiles 0.10	0.25	0.50	0.75
Level ratio (≥ 1.5)	132	0.040	0.122	0.318	0.833
Count floor (≥ 20)	52	2	6	25	71
Growth increase (≥ 0.1)	15	0.022	0.066	0.137	0.239
Post-period growth ($\geq c_1$)	2	0.043	0.088	0.145	0.230

0.062; 78 events have negative growth. Figure **??** illustrates paths selected around quartiles within the 112-event subset meeting the pre-period scale restriction.

Among four continuous criteria, the median ratio has the smallest relative slack in 132 events, the count floor in 52, the growth gap in 15 and post-period growth in 2 (Table 2). Positivity is satisfied with exactly two increases in 110 events. Thus the margin ranking does not mean positivity was nonbinding or that the closest continuous threshold uniquely determined the date. Forty-eight events have a minimum relative slack below 0.05 and 87 below 0.10. Among 5,298 candidates meeting the pre-period scale restriction, 518 fail exactly one condition: 301 fail the ratio and 206 the gap. The counts describe proximity and exclusion under the stated normalisation.

Counterfactual rescanning establishes the distinct effect of removing a condition (Table 3). Omitting the ratio advances the first date in 56 baseline event series and adds 191 previously unlabelled series. Removing the floor instead adds 592 series while advancing only 23 of the original first dates. Positivity removal advances seven and adds eleven. No condition is necessary for a baseline series to receive some label after that condition is removed, but earlier acceptance can alter later cooldown decisions. Level and scale therefore matter to the calendar, while showing why relative slack and influence on the calendar must be reported separately.

5.2 Calendar dependence on definition and earliest geography

On the common 2016–2022 calendar the baseline accepts 192 events. Ratio thresholds of 1.25 and 1.75 produce 280 and 129 events, with Jaccard agreement 0.542 and 0.514. Windows $(4, 4)$ produce 138 and agreement 0.284. Redating and membership effects differ (Table 4): the lower ratio advances 26 existing first dates and adds 86 series, whereas the higher ratio delays 20 and removes 63. Wider windows advance 32, delay seven, remove 80 and add 25. Cooldown changes leave first dates and membership unchanged in this panel; they affect repeat events.

Among fields with events under both definitions, the earliest-labelled geography set changes in 24 of 33 fields under $(4, 4)$, in 19 of 34 under ratio 1.25 and in 17 of 32 under ratio 1.75. For pairs labelled under both definitions, $(4, 4)$ preserves 211 of 260

Table 3 Removing one condition and rescanning 2016–2023 chronologically. Baseline: 201 events in 201 series. Earlier: baseline event series whose first accepted year moves earlier. New: series with no baseline event that gain a label. All baseline event series remain labelled. Source: independent recalculation from the archived EPO panel.

Condition omitted	Events	Labelled series	Earlier	New
Post-period growth	202	202	1	1
Growth increase	280	268	21	67
Level ratio	405	392	56	191
Positive changes	214	212	7	11
Count floor	813	793	23	592

Table 4 Definition sensitivity on the common 2016–2022 calendar read as membership, redating and lead order. Baseline: 192 events in 192 series, 34 fields with an event, 614 pairs of labelled series within fields. Fields first: number of fields, among those with events under both definitions, whose set of earliest-labelled geographies changes. Lead order: pairs of series labelled under both definitions whose within-field order is preserved, flipped, dissolved into a tie or created from a tie. Source: EPO Data Desk report, extraction of 2 February 2026; calculations in the replication materials.

Definition	Events	Same	Earlier	Later	Lost	Gained	Fields	Pairs	Flip./diss./cr.
Acceleration threshold 0.075	205	185	7	0	0	13	5	614	6/10/12
Acceleration threshold 0.125	171	168	0	3	21	0	7	471	0/3/5
Median ratio 1.25	280	166	26	0	0	86	19	614	20/34/16
Median ratio 1.75	129	109	0	20	63	0	17	309	1/5/14
Post-growth log(1.16)	177	177	0	0	15	0	6	529	0/0/0
No cooldown	253	192	0	0	0	0	0	614	0/0/0
Growth windows $(3, 3)$	187	141	10	12	29	24	16	478	4/17/21
Growth windows $(4, 4)$	138	73	32	7	80	25	24	260	8/24/17

within-field orderings; the remaining pairs reverse order or change tie status. Every comparison here depends on both the specification and its labelled population.

Boundary effects warrant a separate check. Forty-eight of the 201 events fall in 2016, and 23 of 34 fields have their earliest label then. Earlier history and cooldown are unobserved. Removing 2016 on the common comparison calendar leaves 144 events, with the earliest-geography set still changing in 25 of 31 fields under $(4, 4)$. Excluding seven selected geographies where applicant attribution may reflect ownership location leaves 148 events on that calendar and changes 25 of 33 fields; excluding both leaves 111 and changes 22 of 30. The geography exclusion is a sensitivity scenario, not evidence that particular applications belong to holding entities (Supplementary Information **??**). The three historical labels add no events; substituting the five discrepant source cells also changes no acceptance.

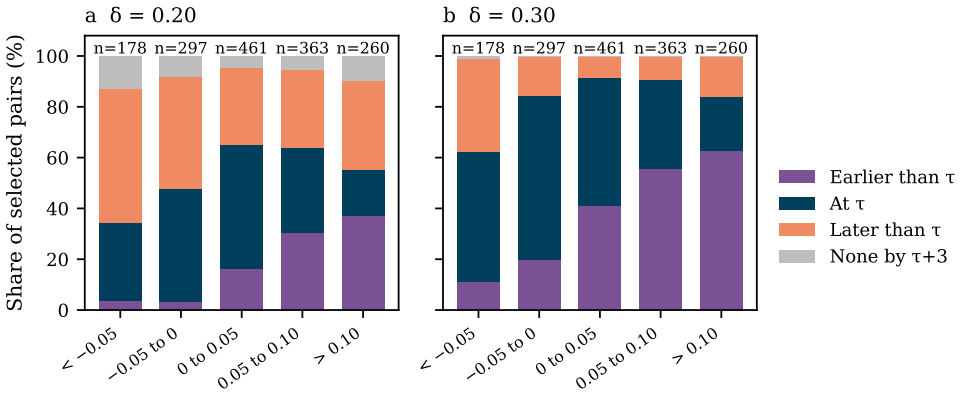


Fig. 3 First full-rule dates on observed backgrounds by pre-period slope, for added increments of 0.20 (panel a) and 0.30 (panel b). Slope classes use the endpoint mean log growth over the four intervals before τ, with boundaries −0.05, 0, 0.05 and 0.10. Stratifying in this way was added after the initial results. Frequencies describe the selected finite panel; overlapping windows are not independent trials. Source: EPO Data Desk report, extraction of 2 February 2026; calculations in the replication materials.

5.3 Known increments on observed backgrounds

Selection yields 1,559 series–anchor pairs from 547 series, 32 geographies and all 35 fields. With $\delta = 0.30$ every constant background is first accepted at τ. Observed backgrounds place 698 pairs there, 633 earlier and 222 later; six receive no accepted date in the local scan. Setting $\delta = 0.10$ leaves constant backgrounds unlabelled while 659 observed backgrounds are labelled (Supplementary Information **??**).

Returning to $\delta = 0.30$, the share dated early rises from 11 per cent in the lowest pre-period slope class to 63 per cent in the highest (Figure 3). Class boundaries are annual log-growth values −0.05 and 0.10. The gradient is consistent with the constant-growth mechanism, but irregular backgrounds also change medians and signs. For that step size the observed first date agrees with the fitted-trend date in 718 of 1,559 pairs and with the constant-background date in 698. Replacing a constant with a fitted trend therefore does not reproduce most of the variation in observed dates.

With observed backgrounds at $\delta = 0.20$, the full rule dates later than the growth-only rule in 972 pairs and suppresses a growth-only label in 80; 474 pairs share a finite first date. Across all 28,062 background–step comparisons, the full rule never has an earlier first date on the common local grid. Set inclusion before first acceptance gives that ordering, and it follows before first acceptance and is independently verified. The remaining criteria materially change both labels and timing, beyond what the growth contrast alone describes.

5.4 Fitted no-change scenarios and date stability

Four no-change scenarios produce materially different calendar sizes (Table 5). The Poisson trend scenario averages 103 events; pooled dispersion raises the mean to 172

Table 5 Accepted events under four no-change nulls and in the data. Each null draws counts around a fitted mean path with a constant log-growth rate; 500 replicates of the whole panel per null, unchanged chronological scan with cooldown. Interval: 2.5th to 97.5th percentile of the 500 replicate calendars. Last two columns: series receiving at least one label, among the 201 series that carry an observed event and among the 2,953 remaining non-empty series. Values describe fitted scenarios, not calibrated false-discovery bounds. Source: EPO Data Desk report, extraction of 2 February 2026; calculations in the replication materials.

Null model	Events	95% interval	2016	2023	Event series	Other series
Trend, Poisson counts	103	88–119	14	14	62	39
Trend, pooled dispersion	172	152–190	23	24	91	75
Trend, series moment dispersion	194	178–214	27	27	106	79
Pre-period continuation	140	124–157	19	20	59	75
Observed calendar	201		48	9	201	0

and series-specific dispersion to 194. Its simulation interval, 178–214, contains the observed 201. Continuing pre-period trends for event series produces approximately 140 events. Each is a calendar generated from fitted means and stipulated count laws. A mean of 140 does not imply that 140 of the observed events are explained by noise, and subtraction from 201 identifies neither real changes nor a lower bound on their number.

Full-period trends and residual dispersion can absorb departures from constant growth. Their consequences for acceptance are not monotone across the five conditions and cooldown. Pre-period continuation additionally uses the observed labels to select fitting windows; it is not an independently specified counterfactual. The scenarios show that sizeable calendars can arise without a discrete growth change, while their spread demonstrates model sensitivity.

A further variant removes the objection that a scenario is fitted to the years it is compared with. Level, slope and dispersion are estimated from 2011–2015 alone and extrapolated over 2016–2025, with the observed history to 2015 held fixed; only series whose 2011–2015 mean reaches five are admitted, which leaves 1,002 of the 6,125 series and 183 of the 201 observed events. Over 500 replicates the scenario calendar averages 256.5 events with a 2.5 to 97.5 percentile interval of 239 to 274, against those 183 observed in the same series. A companion variant that also simulates the pre-period gives 267.3 with an interval of 251 to 287. In this restricted population the scenario therefore produces more labels than the data, so calendar size is not evidence of discrete change here either. The comparison is confined to those 1,002 series, and estimating from 2011–2015 excludes changes before 2016 from the fitted parameters rather than from the data.

Observed events are concentrated toward the beginning of the eligible scan: 48 occur in 2016 and nine in 2023, compared with much flatter simulated profiles (Figure 4a). Left-censoring makes the initial peak difficult to interpret, but neither the mathematical acceleration example nor these scenarios establish its cause. The low count in 2023 is not an artefact of the scan boundary either. Because the rule uses no observation after $k + 2$, truncating the panel leaves every remaining decision unchanged, and rerunning it on panels ending in 2023, 2021 and 2019 reproduces the earlier calendar exactly. In those truncations the final eligible nominal year carries 26, 20 and 29 events, so a year

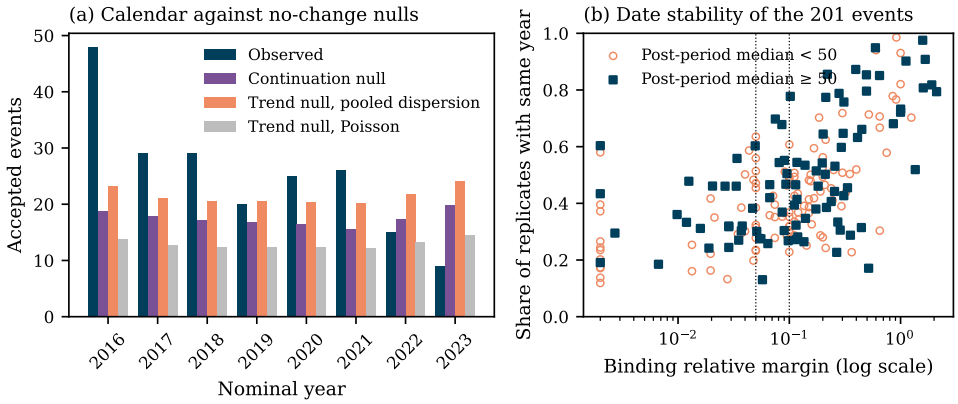


Fig. 4 Fitted calendar scenarios and event-level count perturbation. Panel a: accepted events by nominal year in the data and under three no-change nulls, means of 500 replicate calendars. Panel b: share of parametric bootstrap replicates reproducing the nominal year of each accepted event, against the minimum relative slack of the four continuous criteria, by post-period median. Margins of exactly zero are plotted at 0.002 for the logarithmic axis. Bootstrap frequencies describe the stipulated count law around each observed path, not sampling error of a population.

does not become sparse merely by lying at the edge, and the nine events of 2023 require another explanation. The reported totals of the source fall by 0.3 per cent from 2023 to 2024 and rise by 1.4 per cent to 2025, the two weakest annual changes of the panel apart from 2016 and 2020, which is consistent with the completion window of an event dated 2023 falling into years whose counts have not yet grown as earlier years did.

Event-level continuation draws give pointwise qualification frequencies below 0.05 for 133 events and below 0.10 for 148, with a median of 0.02. Fitted and selected as they are, those diagnostics use a different dispersion-class convention from the calendar continuation. They have no calibrated significance interpretation across the 25,232 candidate evaluations in the 3,154 non-empty series. Zero successes in 2,000 draws is a zero Monte Carlo estimate, not evidence that the probability is below 0.0005; the one-sided 95 per cent binomial upper bound is approximately 0.0015.

Count perturbations around the observed paths reproduce the nominal year in 45 per cent of replicates on average, a year within one in 63 per cent, and no event in 24 per cent. Events with minimum relative slack below 0.05 reproduce their year in 33 per cent of draws, compared with 40 per cent for slack between 0.05 and 0.10 and 53 per cent above 0.10 (Figure 4b). Such a within-panel association motivates reporting margins. It does not validate a universal cutoff for reliable dates.

Whole-series broken-stick dates agree exactly in 19 of 201 cases, within one year in 38 and within two years in 72. The fitted slope change is non-positive in 113 event series. Local qualification and the best single whole-series slope break therefore describe different features of the same paths. Neither supplies ground truth for the other.

As a descriptive benchmark, $4r^{\mathrm{pre}}$ reaches $\log 1.5$ in 33 events and half that value in 67. It is negative in 78 and zero in 11. On an irregular count path this is the contribution under a constant log-growth continuation, not an exact decomposition of its observed median ratio. Rank correlations of 0.34 with ratio slack and -0.25 with

Table 6 The unchanged rule on the reporting-year panel and on five priority-year populations of the same 35-field system. Series counts exclude series that are zero throughout. Continuous-criterion columns give the number of accepted events whose smallest relative distance is at that criterion. The last two columns give the percentage of events within 10 per cent of the closest continuous threshold and the percentage carrying the first year the scan can accept. Priority-year populations cover 2008–2022 and their calendar 2013–2020; the reporting-year panel covers 2011–2025 and its calendar 2016–2023, so the rows are compared in pattern, not in level. Source: EPO Data Desk report and OECD.STI.PIE DSD_PATENTS@DF_PATENTS_WIPO; calculations in the replication materials.

Population	Attribution	Series	Events	Tightest criterion				Margin	First
				Ratio	Floor	Gap	Growth	$< 10\%$	(%)
EPO Data Desk, reporting year	First applicant	3154	201	132	52	15	2	43	24
EPO applications	Inventor	3038	139	92	26	20	1	50	22
	Applicant	2797	164	113	31	20	0	38	16
USPTO applications	Inventor	3257	114	80	23	10	1	46	30
	Applicant	3169	130	94	22	13	1	44	25
PCT applications	Inventor	3224	156	107	29	19	1	37	18
	Applicant	3121	185	131	31	22	1	38	19
Triadic families	Inventor	2845	77	54	17	6	0	35	23
	Applicant	2717	90	62	22	6	0	42	28
IP5 families	Inventor	3265	107	69	25	13	0	43	18
	Applicant	3069	122	83	22	15	2	34	20

the rule-minus-broken-stick offset are descriptive associations rather than verification of the constant-growth theorem on those paths.

5.5 Other offices and counting conventions

OECD priority-year counts permit five further populations: EPO and USPTO applications, PCT applications, and triadic and IP5 patent families, each under inventor and applicant attribution (Organisation for Economic Co-operation and Development 2026). Each uses the same 35 WIPO fields and 101 reporting areas for 2008–2022, with eligible years 2013–2020. Fractional counts are retained as published. Office and family populations overlap and are not independent replications; family selection also changes coverage and cannot be assumed to remove every geographical bias (Martínez 2011; Frietsch and Schmoch 2010). A sixth population changes the field system rather than the document set. The same service publishes EPO priority-year counts for thirteen OECD technology domains, among them artificial intelligence, information technology and medical technology, giving 1,313 domain–geography series over the same areas and years. Those domains overlap by construction, so no share of unique patents can be formed from them; within one export the domain values exceed the corresponding totals in 198 of 3,030 geography–year cells, which is what overlapping assignment implies. Run unchanged on that taxonomy the rule accepts 78 events under inventor and 100 under applicant attribution, and the level ratio is again the tightest continuous criterion, in 50 and 55 of them, with the floor second in 16 and 30 and the post-period growth rate in none. Of the twelve domains labelled under both attributions, nine change their set of earliest-labelled geographies.

Across all ten OECD runs, the level ratio most often has the smallest continuous relative slack, followed by the floor. Under inventor attribution the ratio is tightest

in 92 of 139 EPO events, 80 of 114 USPTO events, 107 of 156 PCT events, 54 of 77 triadic-family events and 69 of 107 IP5-family events. Between 34 and 50 per cent of events have minimum relative slack below 0.10. The counterfactual rescan of Section 5.1 repeats in all ten runs with the same ordering: removing the level ratio advances more of the original first dates than removing any other condition, between 21 and 52 depending on the population, and removing the count floor adds more previously unlabelled series than any other removal, between 340 and 598. Both removals again affect timing and membership together, and the ordering is a property of the conjunction on these calendars rather than a statement about patenting.

Count perturbation also repeats. Applied to the EPO priority-year population under inventor attribution with 2,000 replicates per event, the bootstrap reproduces an event's nominal year in 41.0 per cent of replicate comparisons, comes within one year in 60.4 and returns no event at all in 24.9, against 45, 63 and 24 per cent on the reporting-year panel. The change-shape and no-change scenario results have not been repeated across the five populations.

Event and series counts must be distinguished where a series receives two labels. EPO inventor attribution yields 139 events in 138 series and applicant attribution 164 in 161. Of 94 series labelled under both, 70 have the same first year. Switching attribution changes the set of earliest-labelled geographies in 21 of 33 EPO fields, 28 of 32 USPTO fields, 25 of 34 PCT fields, 15 of 32 triadic-family fields and 24 of 34 IP5-family fields. The comparisons use the first accepted event in each series. Source keys and complete counts appear in Supplementary Information **??**. Counting convention is therefore material to the calendar even when the rule itself is unchanged.

6 Count variation with and without imposed changes

Simulations distinguish the absence of an imposed step from a flat mean (Table 7). On flat Poisson means, the chronological scan accepts at least one event in 9.6 per cent of sequences at $N_0 = 20$ and none of the 2,000 sequences at 100 or 500. At dispersion $\alpha = 0.03$, the shares are 27.1, 6.4 and 2.0 per cent. A constant log-growth rate of 0.03 raises them to 53.0, 23.4 and 17.0 per cent. All of them are frequencies for the specified fourteen-candidate scan; longer or differently initialised scans have different probabilities.

Acceleration requires a separate interpretation. With $a = 0.01$, its log-mean growth contrast is $3.5a = 0.035$, below the gap threshold 0.1. Direct evaluation of $\log(1 + \mu_t)$ also produces no deterministic qualification at the three scales tested. The high noisy acceptance shares therefore arise from variation around an accelerating mean, not from deterministic satisfaction of Proposition 2. Acceleration helps other conditions pass while variation can supply the remaining gap.

Comparing the full and growth-only rules shows how the additional conditions restrict acceptance under the same draws. Their effect varies with scale and dispersion; this comparison alone does not apportion protection among the ratio, positivity and floor. Results with an imposed step also spread across dates. For a flat mean at $N_0 = 100$, $\delta = 0.30$ and $\alpha = 0.03$, the local scan dates 57.9 per cent at τ, 21.1 per cent at $\tau - 1$ and 16.8 per cent at $\tau + 1$. At $\alpha = 0.10$, 9.9 per cent of chronological

Table 7 Sequences with at least one accepted event under the chronological scan $\tau - 5, \ldots, \tau + 8$ when the mean path contains no step ($\delta = 0$), in per cent of 2,000 sequences per cell, and the corresponding share under the two growth conditions alone. Mean paths: flat; growing at log rate 0.03; smoothly accelerating with $a = 0.01$. In the latter, N_0 is a scale parameter, not the mean at $\tau - 1$. Monte Carlo standard error at most 1.12 percentage points. Values are properties of the stipulated distributions. Source: simulations in the replication materials; seeds in the Supplementary Information.

Mean path	N_0	Full rule			Growth conditions only		
		Poisson	NB 0.03	NB 0.10	Poisson	NB 0.03	NB 0.10
Flat	20	9.6	27.1	50.8	88.6	96.5	99.4
Flat	100	0.0	6.4	42.0	11.1	84.1	99.1
Flat	500	0.0	2.0	35.0	0.0	75.4	98.7
Growth $g = 0.03$	20	33.3	53.0	71.0	95.9	99.0	99.8
Growth $g = 0.03$	100	0.2	23.4	64.0	34.7	94.8	99.7
Growth $g = 0.03$	500	0.0	17.0	64.3	0.1	90.6	99.7
Acceleration	20	95.6	99.1	99.6	100.0	100.0	100.0
Acceleration	100	46.0	97.7	98.9	78.3	100.0	100.0
Acceleration	500	1.2	97.2	99.5	5.2	100.0	100.0

scans first accept at $\tau - 3$ or earlier, before any window contains the imposed change. Those early labels can then exclude later candidates through cooldown. The limited AR(1) scenarios change the reported frequencies only modestly; they do not establish insensitivity to serial dependence generally (Supplementary Information **??**).

Change shape matters as well. A deterministic level jump on a flat mean fails the requirement for two positive annual changes. On a growing background the same jump can qualify. A three-year growth phase initially resembles a permanent increase within the available post-window; a gradual logistic doubling can fail the gap condition altogether. Complete shape-by-scale results appear in Supplementary Information **??**, with noisy offset distributions kept separate from deterministic outcomes. Hence the rule's accepted dates depend on the path shape and count law as well as on the location of an imposed change.

7 Discussion

Across constant-growth paths the exact result identifies a route by which background growth affects dating. On the stated monotone path, the median ratio accumulates growth over four years, whereas the moving-sum contrast removes constant growth. Their conjunction can therefore place the first accepted year before an imposed change. Exact thresholds depend on count scale; the large-count expression is an approximation, and an irregular observed path need not have its medians at the window centres.

Empirically, both proximity and counterfactual changes point toward the importance of level and scale. The ratio has the smallest continuous slack in 132 of 201 accepted events, but omitting it advances only 56 original first dates. Removing the floor primarily expands the labelled population. Those differences matter: a descriptive margin is useful information about an accepted point, while a chronological ablation

answers what changes when a condition is absent. Neither establishes a unique cause of an observed technological event.

For cross-series comparisons, the large membership and first-date changes under alternative windows and counting conventions require explicit reporting. Small differences in nominal years may be specification-sensitive, yet no universal rule follows that a particular slope difference invalidates every one-year lead. Whether two labels change order must be checked on their own paths. Likewise, margin and scale flags derived from this panel remain exploratory; their association with simulated date stability does not validate a general reliability threshold.

Large simulated calendars under fitted no-change scenarios show why label counts alone are insufficient evidence of discrete changes. Their variation across fitted count laws matters as much. A comparison of 201 observed events with a scenario mean near 140 cannot assign particular events to noise or estimate a false-discovery fraction. Pointwise continuation frequencies also inherit fitted parameters, selection and overlapping windows. A formal testing procedure would need valid null statistics and selection-aware calibration before applying multiplicity control.

Transfer beyond the chosen rule should follow its components carefully. For monotone exponential paths and odd, fixed windows, median ratios retain the growth accumulated between the window centres. Other statistics, adaptive windows, learned combinations and different change shapes require their own derivations. Published approaches to technology emergence and maturity provide relevant comparisons (Breitzman and Thomas 2015; Lezama-Nicolás et al. 2018; Leydesdorff 2015; Liang et al. 2021; Jiang et al. 2024; Ebadi et al. 2026); the numerical offsets established here are not results for those methods. Statistical change-point procedures offer another route under their own assumptions (Killick et al. 2012; Fryzlewicz 2014; Cho 2016; Eichinger and Kirch 2018; Barigozzi et al. 2018; Truong et al. 2020; Barigozzi et al. 2026). The broken-stick comparison in this study describes disagreement without treating either output as the correct historical date.

Several limits remain decisive. No independent onset chronology is observed. The main panel is a current reporting-year snapshot; the five OECD populations share a broad field system and overlap in underlying applications. Finer classifications, earlier data releases and alternative family definitions remain untested. The perturbation anchors include 2020, without identifying common pandemic effects. Count distributions and dispersion classes are stipulated, and the date perturbation holds observed paths fixed. The adopted constants are not calibrated to technological truth. In the language of indicator assessment (Daraio et al. 2025), this supplies evidence about operational sensitivity while leaving historical validity unresolved.

8 Conclusion

RQ1. On the specified constant-growth path, first acceptance depends on the conjunction of a growth contrast with levels, scale and positive changes. Background growth enters the median ratio over four years and can advance the label relative to an imposed growth step. Irregular paths and count variation replace exact offsets with path-dependent outcomes.

RQ2. The EPO panel yields 201 accepted events and marked calendar sensitivity. The ratio and floor usually have the smallest continuous relative slack, but their removal affects timing

and membership differently. Wider growth windows change the earliest-labelled geography in 24 of the 33 fields with events under both definitions. OECD applicant and inventor conventions also produce different first-year calendars. Simulated count perturbations reproduce the observed year in 45 per cent of draws on average, conditional on the chosen law.

These results identify how the rule's construction shapes its calendar. A first accepted year is the first eligible candidate at which all five conditions hold simultaneously in the chronological scan, subject to the cooldown rule. It is not the first crossing of a uniquely determining condition. The analytical result isolates how background growth enters the level ratio; the empirical analyses establish additional sensitivities within their stated designs. Removing the ratio advances more original first dates, while removing the floor adds more newly labelled series in the main panel. Both conditions affect both outcomes. Across related patent populations, applicant and inventor attribution also change the earliest-labelled geographies. Fitted scenarios produce sizeable calendars under constant mean growth, and in the variant whose parameters come from 2011–2015 alone the scenario mean of 256.5 events exceeds the 183 observed in the same 1,002 series; neither establishes statistical indistinguishability nor identifies the origin of an individual event. The mechanism, ablation and sensitivity comparisons provide distinct evidence; none supplies an independent chronology of technological change.

Any useful event record therefore retains the counting unit and attribution rule, source version, nominal year k, window-completion year $k + 2$, pre-period endpoint growth, every criterion value and pass state, and cooldown history. The minimum relative slack across the four continuous criteria should be accompanied by the discrete positivity count, not presented as a unique determinant of the date. Specification comparisons should distinguish lost and gained series from redating within retained series and show ties at the first eligible year.

Such reporting permits a reader to inspect which comparisons survive changes in definition and data convention. It cannot turn an assigned year into a validated technological onset. Establishing that interpretation requires an independent chronology and an evaluation designed around it; until then, the calendar remains a transparent screening output whose sensitivity can be measured and disclosed.

Data and code availability

All data analysed here are aggregate public counts. The European Patent Office counts come from the Data Desk report cited in Section 4.1; the five priority-year populations and the thirteen-domain taxonomy come from the OECD dataflow cited there, retrieved with the keys and checksums recorded in the replication archive. That archive contains the normalized count panels, the raw source exports, the analysis code, the derived event records and the computational outputs, and is deposited in a public repository under an anonymous replication archive; the digital object identifier is supplied with the submission. It does not contain the original field-export workbooks or parts of the source-ingestion record from the first preparation of the panel; Supplementary Information **??** separates reproduction from the normalized panel from reconstruction of that original chain. A later full export of all 35 fields reproduces every panel cell,

without carrying the checksums recorded at the time. No application-level records are used.

Statements and Declarations

Funding. No funding was received for this research.

Competing interests. The author is a practising German and European patent attorney and a partner in a patent law firm, and lectures on patent and trademark law at Rosenheim Technical University of Applied Sciences. No client, firm or third party commissioned, funded or reviewed this work, and no party has a financial interest in its findings.

Author contributions. The author is solely responsible for the design, the analysis, the interpretation and the text.

Ethics approval. Not applicable; the study uses aggregate public statistics and involves no human participants or animals.

References

Adcock R, Collier D (2001) Measurement validity: A shared standard for qualitative and quantitative research. American Political Science Review 95(3):529–546. https://doi.org/10.1017/S0003055401003100

Bai J, Perron P (1998) Estimating and testing linear models with multiple structural changes. Econometrica 66(1):47–78. https://doi.org/10.2307/2998540

Bai J, Perron P (2003) Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18(1):1–22. https://doi.org/10.1002/jae.659

Barigozzi M, Cho H, Fryzlewicz P (2018) Simultaneous multiple change-point and factor analysis for high-dimensional time series. Journal of Econometrics 206(1):187–225. https://doi.org/10.1016/j.jeconom.2018.05.003

Barigozzi M, Cho H, Trapani L (2026) Moving sum procedure for multiple change point detection in large factor models. Journal of Time Series Analysis 47(3):450–464. https://doi.org/10.1111/jtsa.70028, published online 27 October 2025

Benner M, Waldfogel J (2008) Close to you? bias and precision in patent-based measures of technological proximity. Research Policy 37(9):1556–1567. https://doi.org/10.1016/j.respol.2008.05.011

Borsboom D, Mellenbergh GJ, van Heerden J (2004) The concept of validity. Psychological Review 111(4):1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Breitzman A, Thomas P (2015) The emerging clusters model: A tool for identifying emerging technologies across multiple patent systems. Research Policy 44(1):195–205. https://doi.org/10.1016/j.respol.2014.06.006

Cameron AC, Trivedi PK (2013) Regression Analysis of Count Data, 2nd edn. Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9781139013567

Carley SF, Newman NC, Porter AL, et al (2018) An indicator of technical emergence. Scientometrics 115(1):35–49. https://doi.org/10.1007/s11192-018-2654-5

Cho H (2016) Change-point detection in panel data via double CUSUM statistic. Electronic Journal of Statistics 10(2):2000–2038. https://doi.org/10.1214/16-EJS1155

Daim TU, Rueda G, Martin H, et al (2006) Forecasting emerging technologies: Use of bibliometrics and patent analysis. Technological Forecasting and Social Change 73(8):981–1012. https://doi.org/10.1016/j.techfore.2006.04.004

Daraio C, Glänzel W, Gorraiz J (2025) A responsible framework for an appropriate bibliometric-based research assessment. In: Proceedings of the 20th International Conference on Scientometrics and Informetrics (ISSI 2025), pp 6–15, https://doi.org/10.51408/issi2025_160

Davison AC, Hinkley DV (1997) Bootstrap Methods and Their Application. Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9780511802843

Dechezleprêtre A, Ménière Y, Mohnen M (2017) International patent families: From application strategies to statistical indicators. Scientometrics 111(2):793–828. https://doi.org/10.1007/s11192-017-2311-4

Ebadi A, Auger A, Gauthier Y (2026) WISDOM: An AI-powered framework for emerging research detection using weak signal analysis and advanced topic modelling. Journal of Informetrics 20(1):101759. https://doi.org/10.1016/j.joi.2025.101759

Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap. Chapman & Hall, New York, https://doi.org/10.1201/9780429246593

Eichinger B, Kirch C (2018) A MOSUM procedure for the estimation of multiple random change points. Bernoulli 24(1):526–564. https://doi.org/10.3150/16-BEJ887

European Patent Office (2020) European patent convention, 16th edition, implementing regulations, rule 159: The european patent office as a designated or elected office. requirements for entry into the european phase. URL https://www.epo.org/en/legal/epc/2020/r159.html, text in force on 1 July 2020, accessed 8 September 2026

European Patent Office (2026) European patent applications and grants. EPO Data Desk, URL https://datadesk.epo.org/dashboards/european-patent-applications-and-grants-2026-03, published 23 March 2026. Accessed 8 September 2026

European Patent Office (n.d.) Clarification on data sources. URL https://www.epo.org/en/about-us/statistics/clarification-data-sources, accessed 8 September 2026

Frietsch R, Schmoch U (2010) Transnational patents and international markets. Scientometrics 82(1):185–200. https://doi.org/10.1007/s11192-009-0082-2

Frietsch R, Neuhäusler P, Rothengatter O (2013) Which road to take? filing routes to the european patent office. World Patent Information 35(1):8–19. https://doi.org/10.1016/j.wpi.2012.09.011

Frietsch R, Gruber S, Bornmann L (2025) The definition of highly cited researchers: the effect of different approaches on the empirical outcome. Scientometrics 130(2):881–907. https://doi.org/10.1007/s11192-024-05158-1

Fryzlewicz P (2014) Wild binary segmentation for multiple change-point detection. The Annals of Statistics 42(6):2243–2281. https://doi.org/10.1214/14-AOS1245

Griliches Z (1990) Patent statistics as economic indicators: A survey. Journal of Economic Literature 28(4):1661–1707. URL https://ideas.repec.org/a/aea/jeclit/v28y1990i4p1661-1707.html

Hand DJ (1996) Statistics and the theory of measurement. Journal of the Royal Statistical Society: Series A (Statistics in Society) 159(3):445–492. https://doi.org/10.2307/2983326

Hausman J, Hall BH, Griliches Z (1984) Econometric models for count data with an application to the patents–r&d relationship. Econometrica 52(4):909–938. https://doi.org/10.2307/1911191

Higham K, de Rassenfosse G, Jaffe AB (2021) Patent quality: Towards a systematic framework for analysis and measurement. Research Policy 50(4):104215. https://doi.org/10.1016/j.respol.2021.104215

Jacobs AZ, Wallach H (2021) Measurement and fairness. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery, New York, pp 375–385, https://doi.org/10.1145/3442188.3445901

Jiang M, Yang S, Gao Q (2024) Multidimensional indicators to identify emerging technologies: Perspective of technological knowledge flow. Journal of Informetrics 18(1):101483. https://doi.org/10.1016/j.joi.2023.101483

Kane MT (2013) Validating the interpretations and uses of test scores. Journal of Educational Measurement 50(1):1–73. https://doi.org/10.1111/jedm.12000

Killick R, Fearnhead P, Eckley IA (2012) Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association 107(500):1590–1598. https://doi.org/10.1080/01621459.2012.737745

Kyebambe MN, Cheng G, Huang Y, et al (2017) Forecasting emerging technologies: A supervised learning approach through patent analysis. Technological Forecasting and Social Change 125:236–244. https://doi.org/10.1016/j.techfore.2017.08.002

Lafond F, Kim D (2019) Long-run dynamics of the U.S. patent classification system. Journal of Evolutionary Economics 29(2):631–664. https://doi.org/10.1007/s00191-018-0603-3

Leydesdorff L (2015) Can technology life-cycles be indicated by diversity in patent classifications? the crucial role of variety. Scientometrics 105:1441–1451. https://doi.org/10.1007/s11192-015-1639-x, URL https://link.springer.com/article/10.1007/s11192-015-1639-x

Lezama-Nicolás R, Rodríguez-Salvador M, Río-Belver R, et al (2018) A bibliometric method for assessing technological maturity: the case of additive manufacturing. Scientometrics 117:1425–1452. https://doi.org/10.1007/s11192-018-2941-1, URL https://link.springer.com/article/10.1007/s11192-018-2941-1

Liang Z, Mao J, Lu K, et al (2021) Combining deep neural network and bibliometric indicator for emerging research topic prediction. Information Processing & Management 58(5):102611. https://doi.org/10.1016/j.ipm.2021.102611

Martínez C (2011) Patent families: When do different definitions really matter? Scientometrics 86(1):39–63. https://doi.org/10.1007/s11192-010-0251-3

Meier A, Kirch C, Cho H (2021) mosum: A package for moving sums in change-point analysis. Journal of Statistical Software 97(8):1–42. https://doi.org/10.18637/jss.v097.i08, URL https://www.jstatsoft.org/article/view/v097i08

Nagaoka S, Motohashi K, Goto A (2010) Patent statistics as an innovation indicator. In: Hall BH, Rosenberg N (eds) Handbook of the Economics of Innovation, vol 2. Elsevier, Amsterdam, p 1083–1127, https://doi.org/10.1016/S0169-7218(10)02009-5

Neuhäusler P, Frietsch R (2013) Patent families as macro level patent value indicators: applying weights to account for market differences. Scientometrics 96(1):27–49. https://doi.org/10.1007/s11192-012-0870-y

Ng JY, Liu H, Masood M, et al (2025) Guidance list for reporting bibliometric analyses (GLOBAL): A two-round modified delphi study. In: Proceedings of the 20th International Conference on Scientometrics and Informetrics (ISSI 2025), pp 837–861, https://doi.org/10.51408/issi2025_069

OECD (2009) OECD Patent Statistics Manual. OECD Publishing, Paris, https://doi.org/10.1787/9789264056442-en

Organisation for Economic Co-operation and Development (2026) Patents by WIPO technology domains. OECD Data Explorer, dataflow OECD.STI.PIE,

DSD_PATENTS@DF_PATENTS_WIPO, version 1.0, retrieved 9 September 2026 in SDMX-CSV via https://sdmx.oecd.org/public/rest/data/

Porter AL, Garner J, Carley SF, et al (2019) Emergence scoring to identify frontier r&d topics and key players. Technological Forecasting and Social Change 146:628–643. https://doi.org/10.1016/j.techfore.2018.04.016

de Rassenfosse G, Dernis H, Guellec D, et al (2013) The worldwide count of priority patents: A new indicator of inventive activity. Research Policy 42(3):720–737. https://doi.org/10.1016/j.respol.2012.11.002

Rotolo D, Hicks D, Martin BR (2015) What is an emerging technology? Research Policy 44(10):1827–1843. https://doi.org/10.1016/j.respol.2015.06.006

Schmoch U (2008) Concept of a technology classification for country comparisons. Final report to the World Intellectual Property Organisation (WIPO). Tech. rep., Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, uRL https://wipo.int/documents/2948119/3215563/wipo_ipc_technology.pdf, accessed 8 September 2026

Truong C, Oudre L, Vayatis N (2020) Selective review of offline change point detection methods. Signal Processing 167:107299. https://doi.org/10.1016/j.sigpro.2019.107299

Verhoeven D, Bakker J, Veugelers R (2016) Measuring technological novelty with patent-based indicators. Research Policy 45(3):707–723. https://doi.org/10.1016/j.respol.2015.11.010

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 02SupplementaryInformation.pdf