Centre for Economic Performance

Discussion Paper

ISSN 2042-2695

No. 2234
November 2026

Why surveys can hide rising inequality (and how to correct for it)

Charlie Harrison
Alberto Prati

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

UKRI Economic and Social Research Council

Abstract
Self-reported ratings are widely used in social surveys. Yet, changes in how respondents interpret rating scales over time can make trends in inequalities difficult to assess. In this paper, we develop the V-LINE, a new method that uses panel data to estimate and correct for rescaling. We validate our method using ordinal income categories, showing that it accurately recovers changes in dispersion of the underlying income distribution. Then, using more than 1.2 million observations from four national panel studies, we shed light on two empirical puzzles: the apparent decline of life satisfaction inequality despite rising wealth and income inequalities; the declining dispersion of self-reported ideology despite rising elite political polarisation. We show that, in Germany, life satisfaction inequality has increased substantially since the 1980s, as did ideological polarisation in the UK since 2015. Conversely, we detect no hidden inequality rises in Australia and Switzerland.

Keywords: happiness inequality; political polarisation; rescaling; life satisfaction; political ideology; Brexit
JEL codes: I31; D72; C83

This paper was produced as part of the Wellbeing Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

ChatGPT was used to help rephrase some sentences and proofread the draft. Claude was used to assist with coding and debugging, editing the manuscript, and proofreading the draft.

Charlie Harrison, University of Oxford and the London School of Economics. Alberto Prati, University College London, University of Oxford, and the Centre for Economic Performance at the London School of Economics.

Published by
Centre for Economic Performance
London School of Economic and Political Science
Houghton Street
London WC2A 2AE

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means without the prior permission in writing of the publisher nor be issued to the public or circulated in any form other than that in which it is published.

Requests for permission to reproduce any article or part of the Working Paper should be sent to the editor at the above address.

© C. Harrison and A. Prati, submitted 2026.

1 Introduction

Ordinal self-reported data are increasingly used in large-scale surveys to measure important constructs – like feelings and opinions – that would otherwise be unobservable. This is typically done by asking people to place themselves on a numerical scale, say, from 0 to 10. The analysis of the responses over time can inform not only about changes in the mean sentiment in the population, but also about changes in its dispersion.

In this paper, we focus on the dispersion of two self-reported ratings which are widely measured in surveys and often studied in economics. One is life satisfaction, derived from questions like "Overall, how satisfied are you with your life?" on a scale 0 to 10. Self-reported wellbeing ratings have become increasingly common in economics research (e.g., Perez-Truglia, 2020; Benjamin et al., 2024; Vivalt et al., 2026), and their dispersion has been used to learn about the evolution of happiness inequality (Kalmijn and Veenhoven, 2005; Stevenson and Wolfers, 2008; Clark et al., 2012; Helliwell et al., 2016; Goff et al., 2018). The second measure is self-reported political ideology, based on a survey question where respondents are asked to place themselves on an 11-point Left–to-Right scale. In political economy, self-reported Left–Right positioning is regularly used to measure respondents' political ideology (e.g., Gethin et al., 2022; Enke et al., 2023; Chinoy et al., 2026), and its dispersion for the level of political polarisation (Dalton, 2008; Dow, 2011; Ezrow, 2007; Moral and Best, 2023; Bernaerts et al., 2023).

Here, we document two paradoxes of the same form: self-reported ratings fail to diverge when we would expect them to. The first paradox is known: happiness inequality has not increased even in countries and epochs during which material inequalities have risen (Stevenson and Wolfers, 2008; Clark et al., 2012). The second paradox is less familiar: in the UK since Brexit, public ideological polarisation has not increased even though elite polarisation has risen sharply. Similar findings have been reported for the United States (Fiorina and Abrams, 2008; Hill and Tausanovitch, 2015). We complement these inequality paradoxes with two novel empirical puzzles: for both constructs, declines in the cross-sectional dispersion of ratings are accompanied by a concomitant decline in response volatility, with panel respondents moving less and less across scale rungs, even after controlling for panel conditioning.

We show that these seemingly independent puzzles can be explained by the same mechanism affecting self-reported ordinal data, i.e. *rescaling*. By "rescaling", we mean the change over time in how the response scales are interpreted. Unlike physical measurements, subjective ratings are not expressed in fixed, externally defined units. Instead, respondents evaluate their position relative to subjective anchors corresponding to their interpretations of the scale endpoints. As a result, identical numerical responses need not represent the same underlying quantity across periods, since the internal scale against which evaluations are made may vary. For instance, the perceived distance between the far Right and the moderate center (10/10 vs 5/10 on the political ideology scale) might be different when measured today vs a decade ago, shifting alongside changes in the positions of political parties. Similarly, the meaning of the best possible life (10/10 on the life satisfaction scale) might have changed across decades, along with changes in material standards, particularly at the

top of the income distribution. The dispersion of the underlying construct is scale-dependent, and is therefore not directly comparable if rescaling occurs. In particular, we show that if the internal scale *stretches* over time because the meaning of the endpoints gets further apart, then survey responses would *underestimate* the true divergence in the underlying construct.

The main contribution of our paper is to develop a new method that estimates scale stretching, and allows us to correct dispersion estimates from longitudinal ordinal data. The key innovation is to use volatility - that is, within-person variations - to learn about rescaling. We call this method V-LINE (volatility-corrected LINE). The underlying intuition is that, if rescaling can occur, comparisons over time require a conversion factor, much like comparing weights expressed in kilograms and pounds: if average annual volatility is 1 pound in one period and 0.5 kg in another, the implied conversion factor is about 2:1 (insofar as short-term latent volatility is approximately constant). The V-LINE applies this same logic to subjective ratings.

When we apply this method to several longitudinal datasets, gathering over 1.2 million observations from over 150,000 respondents, the aforementioned empirical puzzles disappear. We find that in Germany, between 1984 and 2019, while the dispersion of life satisfaction ratings has *declined*, the V-LINE shows a 17% *increase* in latent happiness inequality when correcting for rescaling. This is in line with the rise of material inequality in Germany, where the top-10-to-bottom-50 income ratio rose by about 40 per cent over the same period (World Inequality Database, 2026). In the UK, between 2015 and 2024, while there has been a puzzling *decline* in the dispersion of self-reported political ideology, the V-LINE shows that the latent standard deviation has *increased* by about 27%, thus implying that the ideological distance between voters has increased substantially. This is consistent with rises in party polarisation as assessed by experts (Rovny et al., 2025), and with voters' own perceptions (Ipsos, 2025).

Yet, not all ordinal data hide divergences. When applied to life satisfaction in Australia between 2001 and 2019, the V-LINE shows a flat trend in happiness inequality - slightly declining, if anything. Similarly, in Switzerland, we find no latent increase in life satisfaction inequality, nor in political polarisation.

Crucially, as a proof of concept, we test our method on an economic variable for which we can observe the latent quantity, i.e., income. We proceed by partitioning the income distribution into 11 equally spaced income bands, and show that, as the support stretches because of the rise of top incomes, the dispersion recorded by standard deviation of the ordinal categories mechanically declines. When we apply the V-LINE to the categorical income bands, we show that we recover the value of the latent quantity remarkably accurately.

No correction method is assumption-free, and the V-LINE is no exception. We assume that (conditional) latent volatility is constant across periods, and that scale stretching (and shrinking) keeps categories equidistant. However, we show that the trends we document are robust to plausible violations of these assumptions, and are not an artefact of attrition bias or panel conditioning.

There is no obvious way of checking how well each measure – the V-LINE or the uncorrected ratings – tracks changes in latent inequality. Yet, the ability of either measure to account for a variety of empirical puzzles provides a benchmark to assess its credibility. Although rival theories

can explain each uncorrected pattern in isolation, the V-LINE offers a single unifying mechanism that accounts for them jointly. Overall, when a measure that is theoretically vulnerable to rescaling displays puzzling changes in inequality alongside corresponding changes in response volatility, there are good reasons to use the V-LINE as a correction.

Our paper contributes to the literature on survey-item measurement, with specific relevance for the measurement of subjective wellbeing and of ideological polarisation. On the one hand, it contributes to the growing research strand discussing the implications of rescaling for self-reported satisfaction data (Kaiser, 2022; Fabian, 2022; Prati and Senik, 2026; Harrison, 2025). Arguably, the paper closest to ours is Prati and Senik (2026), which proposes a method to correct life satisfaction data for rescaling. The V-LINE complements that method in some important ways: (i) it does not require a question about retrospective satisfaction (which is not always present in surveys); (ii) it allows us to identify the stretching parameter without having to fix the other model parameter (i.e., the shifter); (iii) it allows us to correct dispersion statistics without heroic assumptions on the structure of the memory error.

On the other hand, it contributes to the literature on measuring subjective political ratings in general, and ideological polarisation in particular. For a long time, political scientists have corrected scales for differential item functioning using anchoring vignettes (King et al., 2004) or Aldrich–McKelvey scaling (Aldrich and McKelvey, 1977). Hare et al. (2015) find that corrected placements are more polarised than raw ones – the same direction of correction as ours, identified across respondents rather than across time. Because both methods need survey items that are rarely repeated across waves, they are applied within a single cross-section, while the dispersion of self-placements is routinely compared over time (Ezrow, 2007; Moral and Best, 2023; Bernaerts et al., 2023). The V-LINE, instead, offers a way to correct trends in ideological polarisation over time, and without the need for calibrating questions.

The remainder of the paper is organised as follows. Section 2 describes the V-LINE method formally, and spells out its underlying assumptions. Section 3 discusses the data. Section 4 applies this framework to income, where the true latent is directly observable and the V-LINE estimate can be checked against known trends. Section 5 applies it to life satisfaction and political ideology. Section 6 discusses the implications of our findings for estimating means. Section 7 concludes.

2 Model

This section formalises the mechanics of hidden divergences; when nominal and latent trends in inequality diverge; and how the V-LINE estimator can help recover trends in latent inequality.

2.1 Setup

This paper presents applications in three domains: life satisfaction (Section 5.1), political ideology (Section 5.2), and income (Section 4). The first two are psychological variables which are measured

with similar survey questions: "how satisfied are you with your life?" (0-10); "where would you place yourself [politically] from 0 (left) to 10 (right)?". We assume both have some latent quantity (q_{it}) - what we are interested in - but we only observe an individual's survey response at a particular time (R_{it}). Our third case, income, has an observable q_{it} (household income). We construct artificial ordinal responses (R_{it}) and map individuals into a 0–10 scale based on their income band. The income case is a proof of concept that allows us to validate our method. Our method extends to any ordinal category that can be modelled as the expression of a latent variable.

Table 1: Latent Quantities and Reported Variables

Case	**Latent (q_{it})**	**Reported (R_{it})**	**Research Question**
Life satisfaction	Wellbeing	Life satisfaction	Has happiness inequality increased?
Political ideology	Political ideology	Left–Right self-placement	Has political polarisation increased?
Income	Household income	Income category	Does the V-LINE work?

2.2 Standard Assumptions

To study dispersion from self-reported data, what is typically done is to (i) take responses R_{it} (e.g. life satisfaction and political ideology); (ii) label those responses in their rank order (e.g., 0 to 10); (iii) look at some measures of dispersion (e.g., standard deviation); (iv) interpret the dispersion statistics as directly diagnostic of the dispersion of the latent q_{it}. For this process to work, we require some standard assumptions about the latent q_{it} and its mapping to observed responses R_{it}. Let us call this mapping the "reporting function" (Oswald, 2008).

We start by assuming that each rating measures a latent quantity q_{it} on a single axis: latent wellbeing for life satisfaction, and a latent ideological position for left–right self-placement. This assumption is contested but remains standard in both literatures.[1]

Assumption 0 (Existence). Responses refer to a measurable underlying latent state:

$$\exists\, q_{it} \in \mathbb{R} \quad \text{such that} \quad R_{it} = f_{it}(q_{it}). \tag{1}$$

where $f_{it}(\cdot)$ is the reporting function.

We assume that the difference between response categories is constant in terms of the underlying mental state q_{it}. For instance, the difference between a 5 and a 6, and between a 6 and a 7 on the

[1]The existence of a single latent quantity is a workhorse assumption for both constructs (life satisfaction and political ideology), but is not unanimous. Wellbeing can be thought of as a unidimensional construct that evaluates various dimensions, but it can also be considered as inherently multidimensional, and there is an ongoing debate about whether single-item measures can adequately capture it (VanderWeele and Johnson, 2025; Helliwell et al., 2026). Similarly, the left–right self placement, albeit very common, can be contested as an imperfect representation of underlying political ideology, which is best represented along multiple axes (Johnston and Ollerenshaw, 2020). In this paper, we do not challenge Assumption 0.

same 0-to-10 scale should be the same in terms of q_{it}. This corresponds to assuming that the reporting function is linear. Formally:

Assumption 1 (Linearity). Each individual's reporting function is affine and positive:

$$R_{it} = d_{it}\, q_{it} + k_{it}, \qquad d_{it} > 0 \tag{2}$$

where d_{it} is a *stretching* parameter for individual i in wave t, and k_{it} is a *shifting* (location) parameter. Strict positivity of d_{it} ensures a higher reported value always corresponds to a higher latent quantity: more wellbeing or more right-wing.

Next, we require that the underlying mental state among different respondents who report the same rung is the same. For instance, if two respondents both choose 5 out of 10, they should be referring to the same thing. That is, we require the reporting function to be interpersonally comparable.

Assumption 2 (Scale-Use Homogeneity). The reporting function is common across individuals within a wave: d_{it} and k_{it} carry no i subscript. We write $d_t \equiv d_{it}$ and $k_t \equiv k_{it}$.

Together, Assumptions 1 and 2 give a single reporting function shared by the whole population:[2]

$$R_{it} = d_t\, q_{it} + k_t, \qquad d_t > 0 \tag{3}$$

Typically, the reporting function is also assumed to be constant over time. In this paper, we relax this assumption and allow instead for the reporting parameters to vary across periods. That is, we allow for the possibility that *rescaling* occurs.

2.3 Hidden Divergences Are Possible

Here, we show that hidden divergences are possible as soon as we relax the assumption that the internal scale is fixed over time. Dispersion measures depend on the scale of measurement and are therefore not directly comparable over time if rescaling occurs. We take the standard deviation as our statistic of interest, since it is the conventional dispersion measure of happiness scales (Kalmijn and Veenhoven, 2005; Stevenson and Wolfers, 2008; Goff et al., 2018) and left–right self-placement (Ezrow, 2007; Dalton, 2008; Dow, 2011; Moral and Best, 2023; Bernaerts et al., 2023).[3] Scale

[2] Assumptions 1 and 2 are (like every assumption) subject to potential violations. Their plausibility and the consequences of their violations have been studied elsewhere (Kaiser and Lepinteur, 2025; Benjamin et al., 2026; Seré and Decancq, 2025), and are not of primary interest in this paper. Appendix E shows that our results are robust to plausible violations of linearity.

[3] Alternative methods to measure dispersion in ordinal data exist (see Dutta and Foster, 2013; Jenkins, 2021). Some of these measures are scale-invariant, in the sense that they are robust to the numerical labels assigned to the response categories, provided their ordering is preserved. However, this does not make them robust to changes over time in the mapping between latent quantities and reported responses. In particular, if the latent distribution and the thresholds stretch together, ordinal statistics of the response distribution are unchanged even if latent dispersion has risen. These methods therefore accommodate non-linear transformations of the numerical labels within a period, but cannot detect the kind of rescaling we study here.

dependence is apparent from the definition of a standard deviation applied to equation (3):

$$SD(R_t) = d_t \cdot SD(q_t). \tag{4}$$

Consider the ratio between reported inequality in two waves:

$$\frac{SD(R_t)}{SD(R_{t-1})} = \frac{d_t}{d_{t-1}} \cdot \frac{SD(q_t)}{SD(q_{t-1})} \tag{5}$$

This identity alone shows that a comparison of $SD(R_t)$ across waves carries no information about the sign of the underlying trend in $SD(q_t)$, absent further restrictions on the stretching parameter d_t.

We define "hidden divergences" or "hidden rising inequalities" as cases where latent dispersion is rising ($SD(q_t) > SD(q_{t-1})$) while nominal dispersion is falling ($SD(R_t) < SD(R_{t-1})$). Rearranging equation (5) gives this condition purely in terms of scale stretching and reported dispersions:

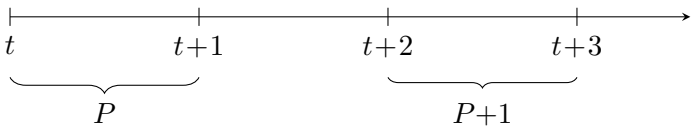
> **Proposition 1 (A theoretical condition for hidden divergences).** *A hidden divergence occurs between $t-1$ and t iff*
>
> $$\frac{d_t}{d_{t-1}} \; < \; \frac{SD(R_t)}{SD(R_{t-1})} \; < \; 1 \tag{6}$$

The right inequality is $SD(R_t) < SD(R_{t-1})$; the left, by equation (5), is $SD(q_t) > SD(q_{t-1})$. Since d_t, d_{t-1} are unrestricted beyond positivity (Assumption 1), the left inequality can always be satisfied.

Equation (6) is key. It shows that an increase in latent inequality can be reversed in nominal terms when the scale stretches faster than the increase in inequality. It depends on one unobservable quantity: the true stretching ratio d_t/d_{t-1}. Assumptions 3–4, described below, supply a model to estimate that ratio.

2.4 Identifying Assumptions

Let $\mathcal{P} = \{P_1, \ldots, P_K\}$ denote a partition of the survey waves into non-overlapping periods (two-year windows for SOEP, HILDA, and SHP, and election cycles for BES). Each period contains several wave observations t, but every wave belongs to exactly one period.



We require a single reporting function, d_P, k_P, within each period; d_P and k_P are free to differ in $P+1$.

Assumption 3 (No Within-Period Rescaling). Reporting parameters are constant within each period:

$$d_t = d_P, k_t = k_P, \qquad \forall\, t \in P,\ P \in \mathcal{P}. \tag{7}$$

It is worth noting that Assumption 3 substantially weakens the assumption the existing literature makes implicitly: that the reporting function is constant *across all periods*. Testing that assumption, rather than maintaining it, is the point of this paper.

Finally, recovering an estimate for scale stretching requires a crucial assumption about latent volatility.

Assumption 4 (Conditional Homoskedasticity). Latent between-wave volatility is constant, once we account for observable differences in covariates.

$$E\big[\,|q_{i,t+1} - q_{it}| \;\big|\; X_{it}\,\big] = b_0 \quad \forall t \tag{8}$$

Assumption 4 is the key assumption doing the identifying work: latent quantities are, on average, no more or less volatile today than in the past, once observables X_{it} such as age are accounted for. It is what allows us to estimate changes in reporting behaviour.

Because periods are defined as non-overlapping sets of adjacent waves, a constant that holds for every wave also holds for every period. Holding the covariate distribution fixed across periods, average latent volatility is therefore the same in every period:

$$\overline{|\Delta q_P|} = b_0 \qquad \forall P. \tag{9}$$

We discuss the plausibility of all four assumptions in more detail, and the sensitivity of our results to violations, in Section 5.3.6.

2.5 From Volatility to the Stretching Parameter

Our goal is to isolate d_P – the stretching parameter – using the volatility of reported ratings. The within-period volatility in reported ratings is:[4]

$$|R_{i,t+1} - R_{it}| \equiv |\Delta R_{it}| \tag{10}$$

Assumption 3 requires that a single reporting function is used within each period P, so the shifter parameter k_P cancels out (equation (2)):

$$|\Delta R_{iP}| = |\Delta q_{iP} \cdot d_P| \tag{11}$$

Since d_P is a positive constant within P (Assumptions 1–2), it separates cleanly in an average across individuals i:[5]

$$\overline{|\Delta R_P|} = d_P \cdot \overline{|\Delta q_P|} \qquad \text{and} \qquad d_P = \frac{\overline{|\Delta R_P|}}{\overline{|\Delta q_P|}}. \tag{12}$$

[4]Here showing unconditional volatility for ease of notation. In practice we use the period coefficients of the regression models in Section 3, rather than raw period means.

[5]Assumption 1 gives $|d_P| = d_P$. Assumption 2 ensures d_P is constant across i, so no covariance term arises when averaging the right side of (11).

This holds at any period, including a reference period $P = 0$, so:

$$\frac{d_P}{d_{P=0}} = \frac{\overline{|\Delta R_P|}}{\overline{|\Delta R_{P=0}|}} \div \frac{\overline{|\Delta q_P|}}{\overline{|\Delta q_{P=0}|}} \tag{13}$$

From Assumption 4, the divisor collapses to:

$$\frac{\overline{|\Delta q_P|}}{\overline{|\Delta q_{P=0}|}} = \frac{b_0}{b_0} = 1 \tag{14}$$

Substituting into Equation (13), and setting the numéraire $d_{P=0} \equiv 1$, yields:

$$\widehat{d}_P = \frac{\overline{|\Delta R_P|}}{\overline{|\Delta R_{P=0}|}} \tag{15}$$

Proposition 2. *The scale stretching parameter in period P, relative to the reference period, equals the indexed volatility of reported values in P.*

We call the estimator $\widehat{d}_P$ the **V-LINE** (Volatility-corrected LINE). The V-LINE is a relative of the M-LINE (Memory-corrected LINE), another rescaling estimator from Prati and Senik (2026).

2.6 From Stretching to Inequality/Polarisation

Since d_P carries no i subscript, it factors cleanly out of a cross-sectional variance:

$$SD(R_P) = d_P \cdot SD(q_P) \quad \Longrightarrow \quad \widehat{SD(q_P)} = \frac{SD(R_P)}{\widehat{d}_P} \tag{16}$$

Proposition 3. *The standard deviation of the latent quantities equals the standard deviation of the self-reported ratings, divided by the estimated stretching parameter.*

Propositions 2 and 3 require no additional assumption about the value of k_P, which cancels out of year-to-year volatility and of cross-sectional standard deviation.

Revisiting Proposition 1: when will the V-LINE detect a hidden divergence? From Proposition 1, a hidden divergence occurs whenever:

$$\frac{d_P}{d_{P-1}} \; < \; \frac{SD(R_P)}{SD(R_{P-1})} \; < \; 1 \tag{17}$$

By Proposition 2, $\widehat{d}_P/\widehat{d}_{P-1}$ recovers this left ratio.[6] From equation (15):

$$\frac{\widehat{d}_P}{\widehat{d}_{P-1}} = \frac{\overline{|\Delta R_P|}}{\overline{|\Delta R_{P-1}|}} \tag{18}$$

[6] We also replace t in Proposition 1 with P to match our estimation of scale stretching.

Substituting this in gives a condition in observables:

Proposition 4 (An empirical condition for hidden divergences). *A hidden divergence occurred between $P-1$ and P iff*

$$\frac{\overline{|\Delta R_P|}}{\overline{|\Delta R_{P-1}|}} < \frac{SD(R_P)}{SD(R_{P-1})} < 1 \tag{19}$$

Hidden divergences occur when response volatility falls *faster* than response dispersion.

3 Data and estimation

This paper draws on four major longitudinal surveys from Western advanced democracies: the German Socio-Economic Panel (SOEP), the Household, Income and Labour Dynamics in Australia survey (HILDA), the Swiss Household Panel (SHP), and the British Election Study Internet Panel (BES). These data cover more than 150,000 individuals across four decades of survey research. All four panels are unbalanced and periodically refresh their samples. We apply cross-sectional survey weights at the wave level in all regressions. In the BES, we use cumulative panel weights covering each set of election waves.

3.1 Life Satisfaction

Life satisfaction is measured in SOEP, HILDA, and SHP. Respondents answer on an 11-point scale running from 0 to 10. A typical wording appears in SOEP: "How satisfied are you with your life, all things considered?" The question has appeared in every SOEP wave since 1984, making SOEP the longest-running panel with life satisfaction data; SHP added it in its second year (2000) and HILDA at its inception (2001). Table 2 reports summary statistics. The mean of life satisfaction has remained flat in each country; the cross-sectional dispersion (SD) has fallen. The three surveys are fielded annually. We restrict our happiness data to pre-COVID periods (up to 2019), and consider a separate COVID-19 period analysis in Appendix D.

3.2 Political ideology among the public

Self-reported political ideology is measured in BES and SHP. In both surveys, respondents place themselves on an 11-point left–right scale. The BES wording is: "In politics people sometimes talk of left and right. Where would you place yourself on the following scale?" The endpoints are labelled 'left' (0) and 'right' (10).

For Switzerland, our main analysis covers 1999–2023, the full span of the SHP sample, and volatility is measured between survey years.

Table 2: Life Satisfaction Levels and Within-Person Volatility

Measure	Germany	Australia	Switzerland
Individuals	65,363	25,633	18,377
Observations	573,229	234,880	133,380
Years with LS	1984–2019	2001–2019	2000–2019
Life Satisfaction (LS)	7.10 (1.80)	7.89 (1.48)	8.06 (1.40)
First Year: LS	7.42 (2.14)	7.94 (1.68)	8.19 (1.48)
Last Year: LS	7.58 (1.65)	7.93 (1.45)	8.07 (1.33)
$\lvert\Delta\text{LS}\rvert$	1.09 (1.22)	0.83 (1.01)	0.78 (0.97)
First Year: $\lvert\Delta\text{LS}\rvert$	1.56 (1.54)	1.03 (1.19)	0.89 (1.07)
Last Year: $\lvert\Delta\text{LS}\rvert$	0.96 (1.15)	0.77 (0.96)	0.71 (0.91)

Note: LS = Life Satisfaction (0–10 scale). $\lvert\Delta\text{LS}\rvert$ = absolute year-to-year change in life satisfaction. Individuals and Observations refer to the Model 1 sample (no controls). $\lvert\Delta\text{LS}\rvert$ statistics use this sample; levels of LS use all valid responses. All statistics are unweighted; figures and regressions use survey weights. Values shown as Mean (SD).

The British Election Study (BES) Internet Panel follows a different rhythm: beginning in February 2014, it runs at least one wave per year, with additional waves clustered around national elections. It is a non-probability opt-in sample, but includes representative cross-sectional weights. This time frame includes five major electoral events: the General Elections of 2015, 2017, 2019, and 2024, and the 2016 Brexit Referendum. Each electoral event is bracketed by a set of three survey waves that follow the same respondents:

- **A** : fielded first, starting approximately two months before polling day.
- **B** : fielded closer to polling day, capturing attitudes near the end of the campaign.
- **C** : fielded immediately after the vote, capturing attitudes once the outcome is known.

Our main BES analysis uses these three waves – A, B and C – for each of the five elections, for a total of 15 waves. We only consider adjacent waves, and construct two measures of volatility in political ideology for each election, i.e., the average change between wave A and B, and between wave B and C. Table 8 (Appendix C) sets out the precise temporal structure. In Appendix C, we also construct a separate annual measure, and use it to replicate our analyses.

Table 3 presents summary statistics for self-placement. Political ideology averages close to 5 out of 10, which is also the modal response. The average has changed little over time, while dispersion has fallen. Note that *levels* of political volatility are higher in Switzerland, reflecting the longer time frame over which it is measured.

Table 3: Political Attitudes Levels and Within-Person Volatility

Measure	Switzerland (SHP)	UK (BES)
Individuals	17,068	45,535
Observations	130,524	142,487
Years with LR	1999–2023	2015–2024
Left-Right (LR)	4.83 (2.14)	4.94 (2.40)
First Period: LR	4.76 (2.25)	4.93 (2.50)
Last Period: LR	4.77 (2.23)	4.80 (2.28)
$\lvert\Delta\text{LR}\rvert$	0.84 (1.17)	0.60 (0.94)
First Period: $\lvert\Delta\text{LR}\rvert$	1.04 (1.32)	0.69 (1.01)
Last Period: $\lvert\Delta\text{LR}\rvert$	0.72 (1.08)	0.47 (0.80)

Note: LR = Left-Right self-placement (0–10 scale). $|\Delta\text{LR}|$ = absolute change in left-right position. Individuals and Observations refer to the Model 1 sample (no controls). $|\Delta\text{LR}|$ statistics use this sample; levels of LR use all valid responses. Switzerland: yearly waves. UK: volatility within each election cycle (General Elections of 2015, 2017, 2019 and 2024, and the 2016 Brexit Referendum). All statistics are unweighted; figures and regressions use survey weights. Values shown as Mean (SD).

3.3 External Benchmarks

Each application compares dispersion in a survey of the public with an external benchmark. These benchmarks are used only to illustrate trends, and different measures of inequality and polarisation would of course be possible.[7]

Material inequality. We compare trends in happiness inequality with trends in income and wealth inequality from the World Inequality Database (2026). For each country, we use the ratio of the average pre-tax national income of the top 10% to that of the bottom 50%, and the same ratio for net personal wealth. Both series cover adults aged 20 and over, with income and wealth split equally between spouses. Section 4 uses a different measure, the standard deviation of income, purely as a diagnostic for the V-LINE (not as a measure of economic inequality).

Elite polarisation. We compare mass polarisation with elite polarisation, measured by Dalton (2008)'s index: the vote-share-weighted standard deviation of party positions, normalised by the scale's half-range. Party positions come from the Chapel Hill Expert Survey (CHES), in which country experts place each party on 0–10 scales (Rovny et al., 2025).[8] We use the general left–right scale alongside the economic and cultural scales; for the latter two, experts are first given a short

[7] Material inequality could also be calculated from consumption data, and several alternative inequality ratios (e.g., P90/P10) and indexes (e.g., Gini or Theil) exist. For elite polarisation, party positions could also be derived, for instance, from manifesto text (Lehmann et al., 2024) or legislative voting records (Hix et al., 2006).

[8] We implicitly assume that party positions in the CHES are measured on a stable scale over time. CHES relies on a common international community of political scientists who assess parties across different countries, making it less likely that the interpretation of the scale is anchored to country-specific political contexts. Besides, cross-sectional vignette experiments show that CHES experts display very little scale use heterogeneity (Bakker et al., 2014), unlike the general population.

description of what each pole represents (Appendix B.1). Each party's position is the unweighted mean across experts for a given country–year. While CHES does not prescribe a minimum number of experts per observation, a comparable source (V-DEM) recommends at least five (Lindberg et al., 2022); our minimum is seven (Table 7).

For the UK, we use the CHES trend file for the 2014, 2019 and 2024 waves. Switzerland is absent from the trend file, so we combine the individual 2010, 2014 and 2019 waves with official election data.[9] Following CHES coding, we weight each party by its vote share in the national election immediately preceding the survey. We construct mass-level latent dispersion in two-year periods and link each elite-polarisation estimate to the corresponding period, indexed by its first year. For the UK, CHES's 2014 wave is linked to the 2015 BES election.

3.4 Empirical Model

In this paper, we analyze trends in volatility across time. To do so, we estimate a model of the form:

$$|\Delta R_{it}| = \alpha + \boldsymbol{\beta} \cdot \text{Period}_P + \varepsilon_{it} \tag{20}$$

with $|\Delta R_{it}|$ referring to our key variable: the within-person volatility of the rating, calculated as the absolute change between consecutive waves t and $t+1$.

Period_P refers to two-year periods for every dataset, aside from BES, where it refers to election cycles.[10] Our preferred specification adds two sets of controls:

$$|\Delta R_{it}| = \alpha + \boldsymbol{\beta} \cdot \text{Period}_P + \boldsymbol{\gamma} \cdot \mathbf{X}_{it} + \delta \cdot \text{Years in Panel}_{it} + \varepsilon_{it} \tag{21}$$

First, X_{it} includes demographic controls. In all datasets, we control for gender and age group (under 25, 25–34, ..., 75+). We also control for migration background in Germany and Australia, but not the UK or Switzerland due to data restrictions.[11]

Second, *Years in Panel*$_{it}$ is the cumulative number of valid responses individual i has given to R_{it} in waves up until t. This addresses panel conditioning effects — the possibility that respondents become less volatile simply through habituation to the survey instrument, for instance by anchoring to a remembered previous response or settling on a habitual rating out of convenience. Panel tenure is a common control in panel studies with self-reported outcomes (Frijters and Beatton, 2012; Kubiszewski et al., 2020). We specify panel tenure as a set of dummy variables, and also report results using a linear trend. Period, age, and panel conditioning effects are jointly identified because people enter the panels at different times and at different ages.

UK specifications control for volatility type (A to B, or B to C; see Section 3.2), except for the

[9] Swiss Federal Statistical Office, *Nationalratswahlen: Zeitreihen*, opendata.swiss.

[10] We estimate volatility for every pair of consecutive waves. For annual panels, this could be 1984–1985, 1985–1986, ... For the V-LINE, we then keep only changes between the two waves of the same period (e.g., 1984–1985, 1986–1987, ...) ensuring each wave t belongs to exactly one period P (Assumption 3). Shifting the grouping by one year gives similar trends (Appendix Figure 10). For the BES, both within-election changes (A to B, B to C) fall inside one period by construction.

[11] There is no migration background question in BES. SHP only began asking about migration background in 2004.

model without controls. Our preferred UK model also controls for the number of days until the next survey, since the average temporal gap between surveys differs slightly across election cycles. Appendix C gives more details.

4 Proof of Concept: Income

We start our empirical analysis by focusing on a continuous non-latent variable, i.e. income. This gives us a test environment, where we can compare the V-LINE estimations of latent inequality to the actual data.

We construct income categories, numbered 0 to 10 with equal width in absolute income space, separately for each year: in each year's sample, after minor trimming,[12] we identify the highest and lowest positive household incomes and divide the range between them into 11 equally sized bands. For example, if household incomes ranged between £1,000 and £111,000, the resulting categories would be: Category 0, £1,000–£10,999; Category 1, £11,000–£20,999; and so on.[13] We use disposable household income adjusted for CPI inflation. We construct this variable for Australia (HILDA), Germany (SOEP), and Switzerland (SHP).

4.1 Hidden Divergences, Demonstrated

The standard deviation of income rose across all three countries over our sample period.[14] Does the standard deviation of our constructed income category capture this rise? Figure 1a compares the two: the standard deviation of actual household income (green) against the standard deviation of the income category (blue). In Germany, the blue line is almost exactly flat; in Switzerland, it increases then decreases back to its starting level; in Australia, it falls. Across the three countries, the blue lines lie substantially below the green line throughout. Figure 1a displays a "hidden divergence" as we defined it earlier: rising inequality that is invisible in the ordinal measure.

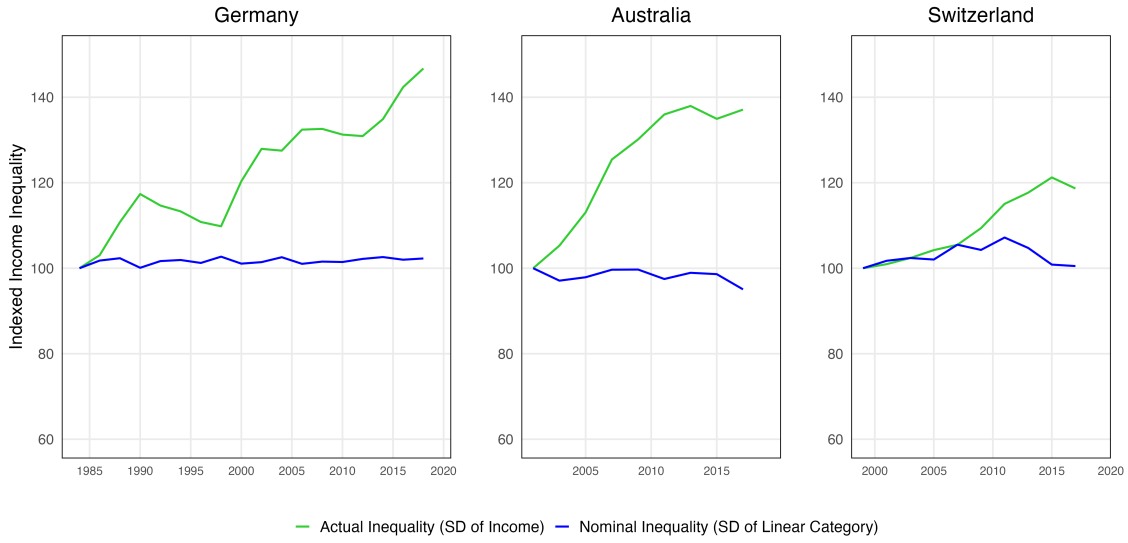
The *mechanism* is also the one Section 2 describes. Bottom incomes grew little relative to top incomes (Figure 4).[15] The range of the empirical distribution has increased. Therefore, the 11 categories have *stretched*, with the gap between each income category growing.

[12] We trim the top and bottom 2.5% of household incomes. This is for two reasons. First, without trimming, a small number of very high incomes would compress the useful range, forcing most of the population into a handful of bottom categories; trimming preserves enough inter-category volatility to be informative. Second, incomes are right-skewed within every category, so the mean income each rung represents would otherwise be markedly less equidistant, even though the category boundaries themselves are – a violation of a version of Assumption 1 (Linearity); trimming limits this. To alleviate both problems, we could have defined more categories instead. We chose eleven categories, rather than more, for direct comparability with the eleven-point life satisfaction scale.

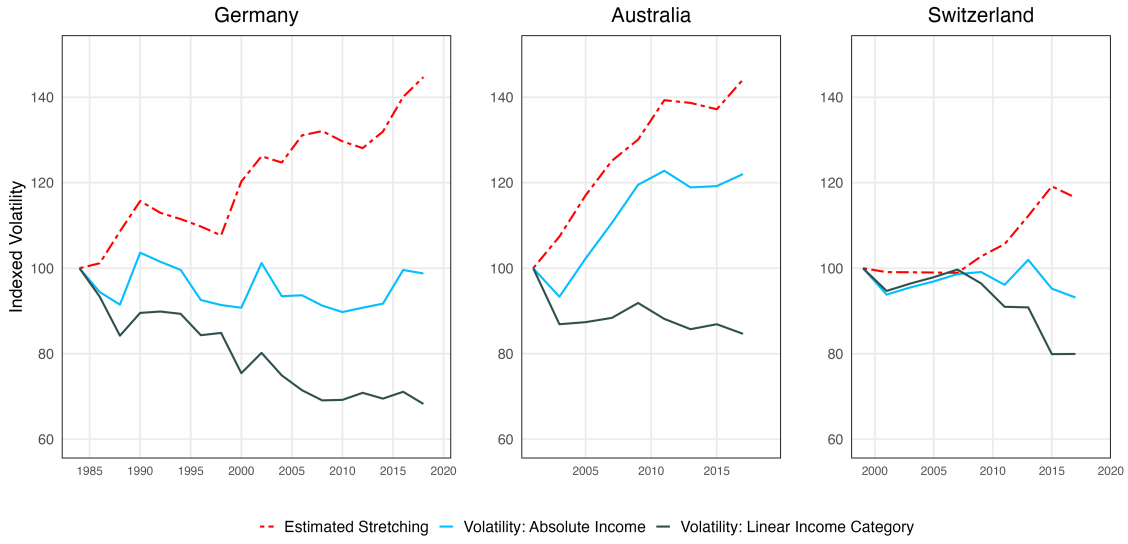
[13] We use equal-width income bands rather than deciles. Deciles are equal in population share, not in income range: the poorest deciles span far narrower income ranges than the richest. Decile assignment is therefore strongly non-linear in income, violating Assumption 1.

[14] Standard deviation is not an ideal income inequality measure, but we adopt it in this validation exercise for comparability with our main applications to subjective ratings.

[15] Figure 4 plots the top 10%/bottom 50% income and wealth ratios for each country. These rise in Germany and Australia, and fall slightly in Switzerland, where our sample nonetheless shows a rising standard deviation.



(a) Nominal and actual dispersion diverge



(b) Response volatility and the stretching parameter

Figure 1: Income validation. Panel (a): standard deviation of the income category (blue) and of household income (green). Panel (b): mean absolute within-person change in the income category and in household income, and the inverse stretching parameter $1/\widehat{d}_P$ (red dotted). All series indexed to 100 in the first period; two-year periods.

4.2 Applying the V-LINE

We now ask whether the V-LINE estimator recovers this known trend. Recall equation (13). Setting the numéraire $d_{P=0} \equiv 1$, this gives the stretching parameter d_P as the ratio of indexed reported volatility, $\overline{|\Delta R_P|}/\overline{|\Delta R_{P=0}|}$, to indexed latent volatility, $\overline{|\Delta q_P|}/\overline{|\Delta q_{P=0}|}$.

While for life satisfaction and political ideology, identification requires an assumption about latent volatility (i.e., Assumption 4), here it is not needed, because we can observe income and estimate its volatility directly.

Then, we apply Proposition 3 exactly as in Section 2: the standard deviation of the latent (here, the empirical standard deviation of income $\widehat{SD(q_P)}$) equals the standard deviation of the ordinal category (i.e., $SD(R_P)$, the empirical standard deviation of income category), divided by the stretching parameter ($\widehat{d}_P$). That is:

$$\widehat{SD(q_P)} = SD(R_P) \cdot \frac{1}{\widehat{d}_P} = SD(R_P) \cdot \frac{\overline{|\Delta q_P|}/\overline{|\Delta q_{P=0}|}}{\overline{|\Delta R_P|}/\overline{|\Delta R_{P=0}|}} \tag{22}$$

Figure 1b shows two volatility series entering this ratio, alongside the resulting stretching parameter. Volatility of the income category (grey) falls in all three countries, while the volatility of income itself (light blue) is either flat or rising. The inverse stretching parameter (red) increases, indicating scale stretching. We multiply the nominal series by 'estimated stretching' to recover the adjusted estimate.

From this plot, there are two initial findings. First, the V-LINE produces a directionally correct answer: it correctly predicts that inequality estimated from ordinal categories needs adjusting upwards. Second, the core intuition for the V-LINE is verified. That is, a stretching in the income distribution results in a lower volatility of income categories.

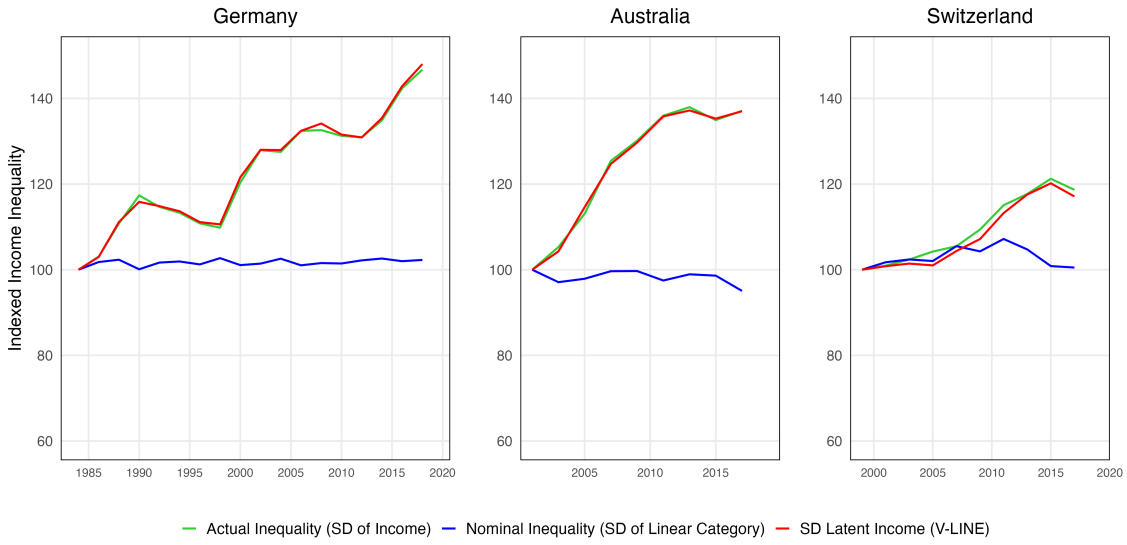


Figure 2: Actual, Nominal, and V-LINE-Adjusted Income Inequality

We can also check the final adjusted series – the estimated latent dispersion. Because household income is independently observable, we can compare this adjusted estimate directly against the observed trend in $SD(q_P)$. The result is shown as the third line in Figure 2: the V-LINE-adjusted series tracks the true rise in income inequality very closely. Visually, it is difficult to distinguish the red (estimated inequality) and the green (observed inequality) lines. The mean absolute error of the dispersion estimated via the V-LINE, relative to the one observed, is 0.53 percentage points for Germany, 0.56 for Australia, and 1.22 points for Switzerland. For Australia and Germany, this is around 40 times smaller than the mean absolute error when estimating the dispersion from income categories; for Switzerland the absolute error falls by a factor of 5. We believe that this is a remarkable level of precision.

What accounts for the residual error? We can largely rule out Assumption 1 (Linearity): our income categories are perfectly equidistant in their widths, and very close to linear in their means.[16] Assumption 4 was not required in this case (we adjusted for latent volatility). We can conclude, then, that the residual error is some combination of violations of Assumptions 2 (reporting heterogeneity) and 3 (continuous reporting drifts).[17]

So far, this exercise has relaxed Assumption 4. In Appendix G, we re-run the exercise assuming the absolute income volatility is constant over time – just as we do for happiness and political ideology. The accuracy of the V-LINE estimator decreases but remains far better than the nominal SD by a factor of 2.7 in Germany (20.93 to 7.69 indexed points), 1.7 in Australia (26.17 to 15.12), and 2.3 in Switzerland (6.75 to 2.91) (Table 13). The correction recovers the right direction of the inequality trend in all cases. Income volatility is close to constant in Germany and Switzerland. In Australia income volatility rises, so the V-LINE with Assumption 4 *under-estimates* the rise in income inequality (Figure 20).

This exercise demonstrates three things. First, a real rise in inequality can be almost invisible in the dispersion of the corresponding ordinal variable. Second, when latent volatility is observed – or when Assumption 4 holds – the V-LINE recovers the latent trend very accurately (mean absolute errors of 0.5–1.2 index points), though less well in years of sharp income shocks. Third, Assumption 4 appears to matter more for precision than for direction. Imposing it, as we must for happiness and ideology, raises the error, but the V-LINE remains 1.7 to 2.7 times more accurate than the nominal series and, importantly, carries the correct sign. We take this as reasonable grounds for confidence in the V-LINE method's application to variables where, unlike income, no independent check is available. These are the subject of the next section.

[16] We quantify any residual nonlinearity using the same cost function Kaiser and Lepinteur (2025) use for subjective scales, computed separately within each year and averaged across the sample. The result is close to zero (0.0001-0.0002 for all three countries). Our income categories are essentially linear.

[17] By design, we specified a common step-wise reporting rule: the same category boundaries, applied uniformly to everyone in a given year. This common rule is a close relative of Assumption 2 (though not exactly the same; income levels are not all equal within each category). Assumption 3 is, plausibly, more important. The income distribution stretches gradually, year by year, rather than jumping discretely from one two-year window to the next – so the reporting function drifts within, not just between, our estimation periods. Our method is less accurate in years with income shocks (e.g., 1990/1991 and 2008/2009 in Germany).

5 Results: Life satisfaction and Political Ideology

In this section, we proceed to our main applications of the V-LINE. We start by investigating happiness inequality, then move on to ideological polarisation. Estimation results for seven specifications are summarised in Tables 4 and 5. Our preferred specification (Model 5) controls for panel tenure and changes in demographics (cf. the empirical model in Section 3.4).

5.1 Life satisfaction inequality

5.1.1 The puzzle: material inequality has risen; happiness inequality has not

Happiness inequality seems to fall in all three countries. The standard deviation of life satisfaction fell from 2.14 to 1.65 in Germany (1984–2019), from 1.68 to 1.45 in Australia (2001–2019) and from 1.48 to 1.33 in Switzerland (2000–2019; Table 2). These declines are consistent with the wellbeing literature (see Clark et al. 2012 for Europe; Ambrey and Fleming 2014 for Australia) and have been documented also in the US (Stevenson and Wolfers, 2008; Dutta and Foster, 2013).

Material inequality, by contrast, has risen in these countries (see Figure 4, World Inequality Database (2026)).[18] For instance, the ratio of top-10 to bottom-50 per cent income rose by about 40 per cent in Germany between 1984 and 2019, and by about 12 per cent in Australia, with most of the rise concentrated before 2008. In Switzerland, the top 10/ bottom 50 ratio of personal wealth rose by about 30 per cent.

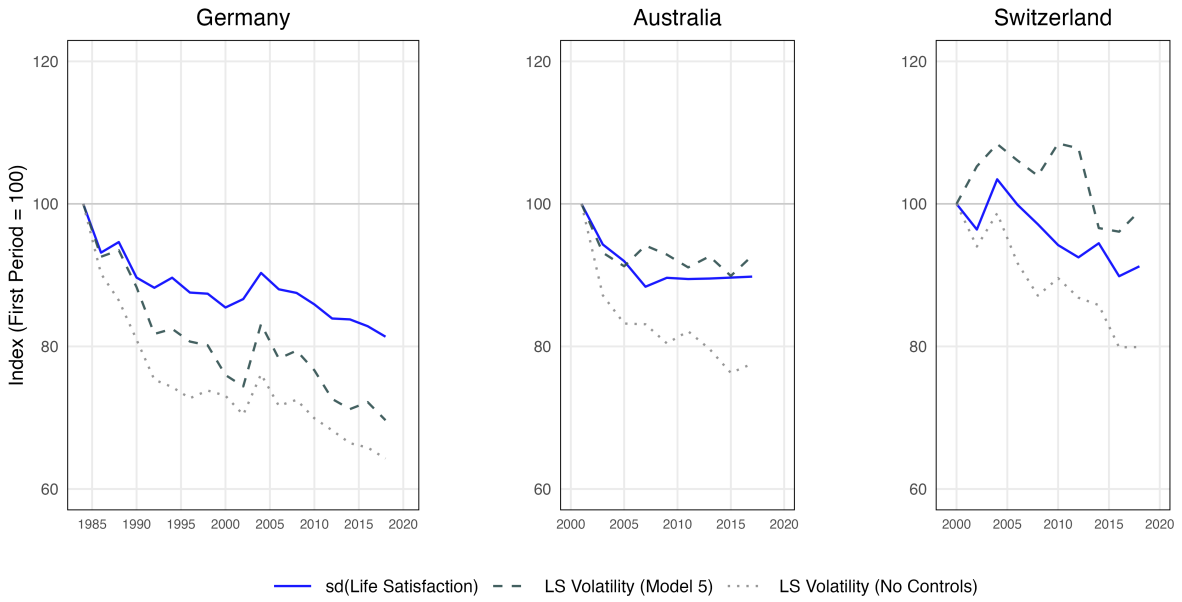
The decoupling of happiness inequality and material inequality is a puzzle because income and wealth are reliable predictors of life satisfaction within countries, with particularly solid evidence for Germany (D'Ambrosio et al., 2020; Jantsch et al., 2024). As material wellbeing diverges, we would expect subjective wellbeing to follow a similar pattern.

5.1.2 Adjusting happiness inequality for scale stretching

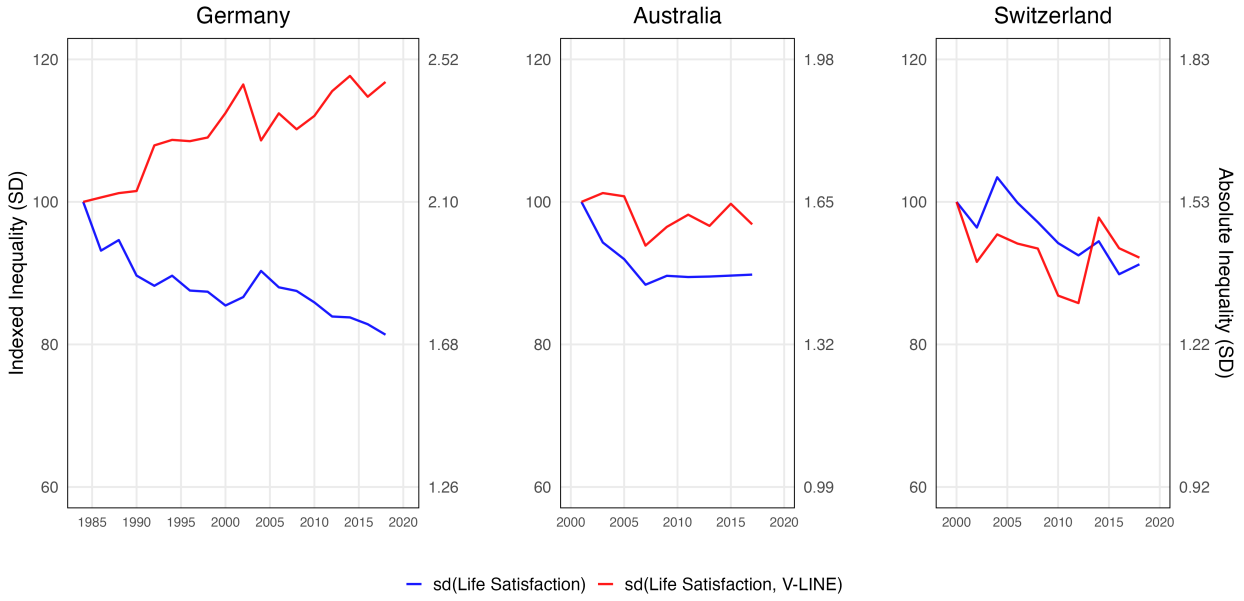
Within-person volatility of life satisfaction fell in Germany and Australia (Figure 3a). Part of the decline reflects panel conditioning: controlling for panel tenure shrinks the trend (Table 4). Under our preferred specification (Model 5), volatility fell by about 30 per cent in Germany and 7 per cent in Australia. Trends are flatter in Switzerland, where volatility rose in the 2000s and fell in the 2010s.

Correcting for scale stretching *reverses* the trend in happiness inequality in Germany. The paradox disappears: latent happiness inequality in Germany moves in the same direction as income and wealth inequality (Figure 4). Trends are not reversed everywhere, though. In Switzerland and Australia, latent inequality still falls slightly after correcting for rescaling. In these two countries, material inequalities have increased, but less markedly than in Germany over its longer observation period.

[18] Different inequality metrics and data sources might give different estimates, but the overall trends for these countries seem uncontroversial. For a detailed discussion, see Alvaredo et al. (2026).



(a) Nominal dispersion and response volatility



(b) Nominal and latent dispersion (V-LINE)

Figure 3: Life satisfaction. Panel (a): standard deviation of life satisfaction (solid), and mean absolute within-person change in life satisfaction under Model 5 (dashed; panel tenure and demographic controls) and without controls (dotted). Panel (b): standard deviation of life satisfaction (blue) and V-LINE-adjusted standard deviation using Model 5 volatility (red); right axis in first-period scale units. All series indexed to 100 in the first period; two-year periods.

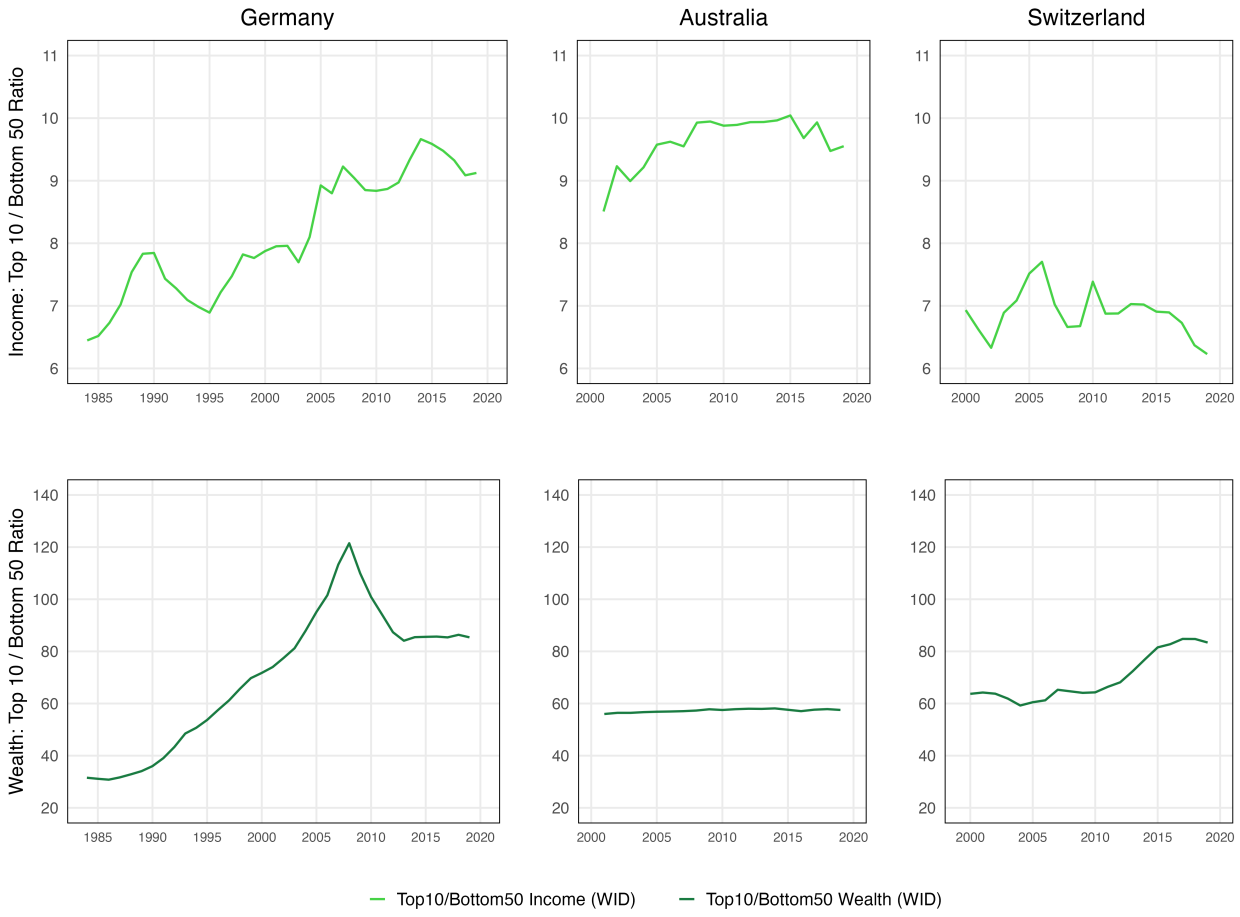


Figure 4: Income and wealth inequality. Ratio of the average pre-tax national income (net personal wealth) of the top 10 per cent to that of the bottom 50 per cent; adults 20+, equal-split, unindexed. Source: World Inequality Database (2026).

Importantly, the V-LINE does not always adjust nominal inequality upwards. When we estimate happiness inequality in the German SOEP during the COVID-19 pandemic, the V-LINE shows that the happiness inequality trend should be adjusted *downward*, consistent with a shrinking scale in a period of compressed aspirations. These analyses are reported in Appendix D, Figure 16.

5.2 Ideological polarisation

5.2.1 The puzzle: party polarisation has risen; voter polarisation has not

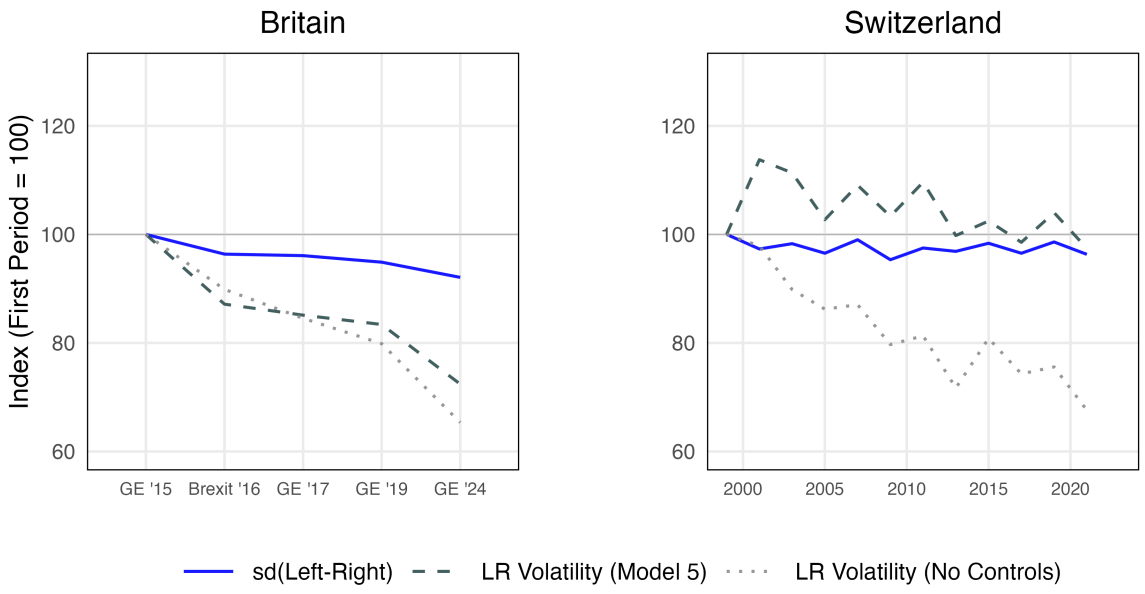
Party-level and voter-level ideological polarisation show opposite trends for the UK. The standard deviation of voters' left–right self-placement *fell* from 2.50 to 2.28 between the 2015 and 2024 general elections (Table 3). By contrast, elite left–right polarisation *rose* by about 40 per cent over the same period, with broadly similar trends on the economic, cultural and overall measures (Figure 6). This divergence implies a growing decoupling between the ideological positions of voters and parties. Theories of party cues would predict otherwise: voters follow their own party, or react against an opposing party as it becomes more extreme (Çakır, 2025). British voters themselves perceive the opposite trend: a large and growing majority say the country is more divided than ten years ago (Ipsos, 2025, p. 4).

For Switzerland, nominal dispersion changed little between 1999 and 2023 (Table 3); nor did elite polarisation on any dimension between 2010 and 2019 (Figure 6).

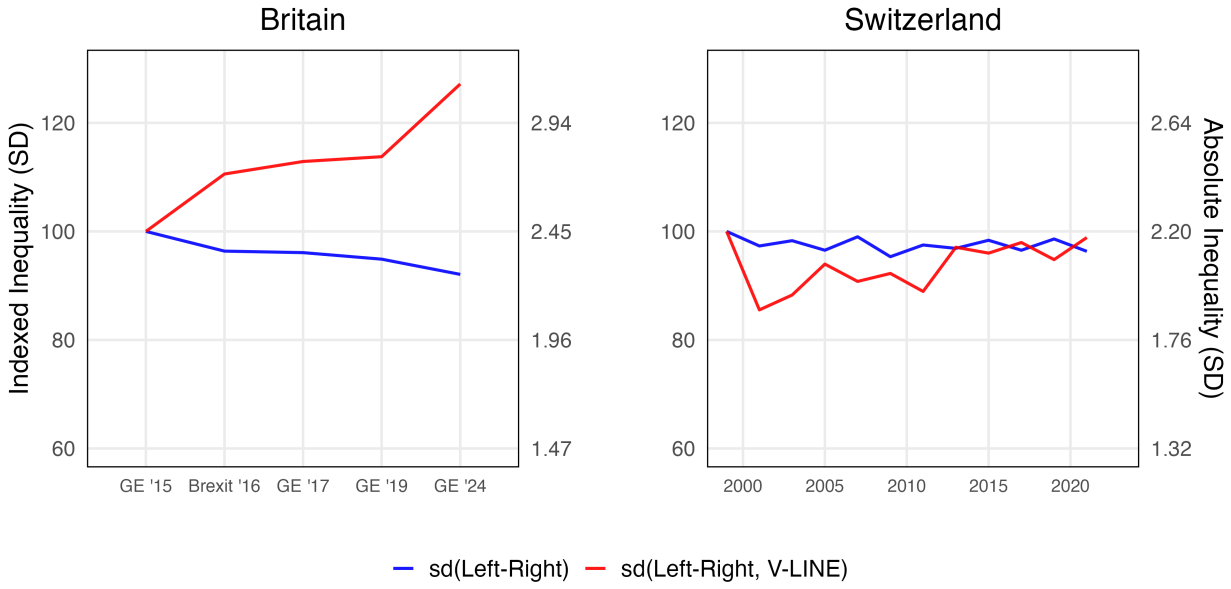
5.2.2 Adjusting polarisation trends for scale stretching

In the UK, political volatility fell substantially between the 2015 and 2024 general elections (Figure 5a). Under Model 5, the linear trend is -5.9 points per election cycle (Table 5), and it is robust to alternative volatility measures (Appendix C, Table 10). While nominal dispersion fell - consistent with declining polarisation - using the V-LINE reverses the trend: once corrected for the Left-Right endpoints drifting apart, latent polarisation rose by about 27 per cent (Figure 5b).

Hence, when using the V-LINE, another empirical paradox disappears: latent mass polarisation in the UK now rises alongside elite polarisation. Conversely, in Switzerland, neither mass nor elite polarisation shows a clear trend in the years in which the data overlap (2010–2019). In Switzerland, most of the decline in volatility without controls reflects panel conditioning. Under Model 5, volatility rose in the early 2000s and then drifted back down; the linear trend is -0.40 points per year (Table 5). Latent polarisation shows no rise over the full period, finishing close to its 1999 level.



(a) Nominal dispersion and response volatility



(b) Nominal and latent dispersion (V-LINE)

Figure 5: Left-right self-placement. Panel (a): standard deviation of left-right self-placement (solid), and mean absolute within-person change under Model 5 (dashed) and without controls (dotted). Model 5 controls for panel tenure and demographics; for the UK, also days until the next survey. Panel (b): standard deviation of left-right self-placement (blue) and V-LINE-adjusted standard deviation using Model 5 volatility (red); right axis in first-period scale units. UK: within-election volatility, by election cycle. Switzerland: yearly volatility, two-year periods. All series indexed to 100 in the first period.

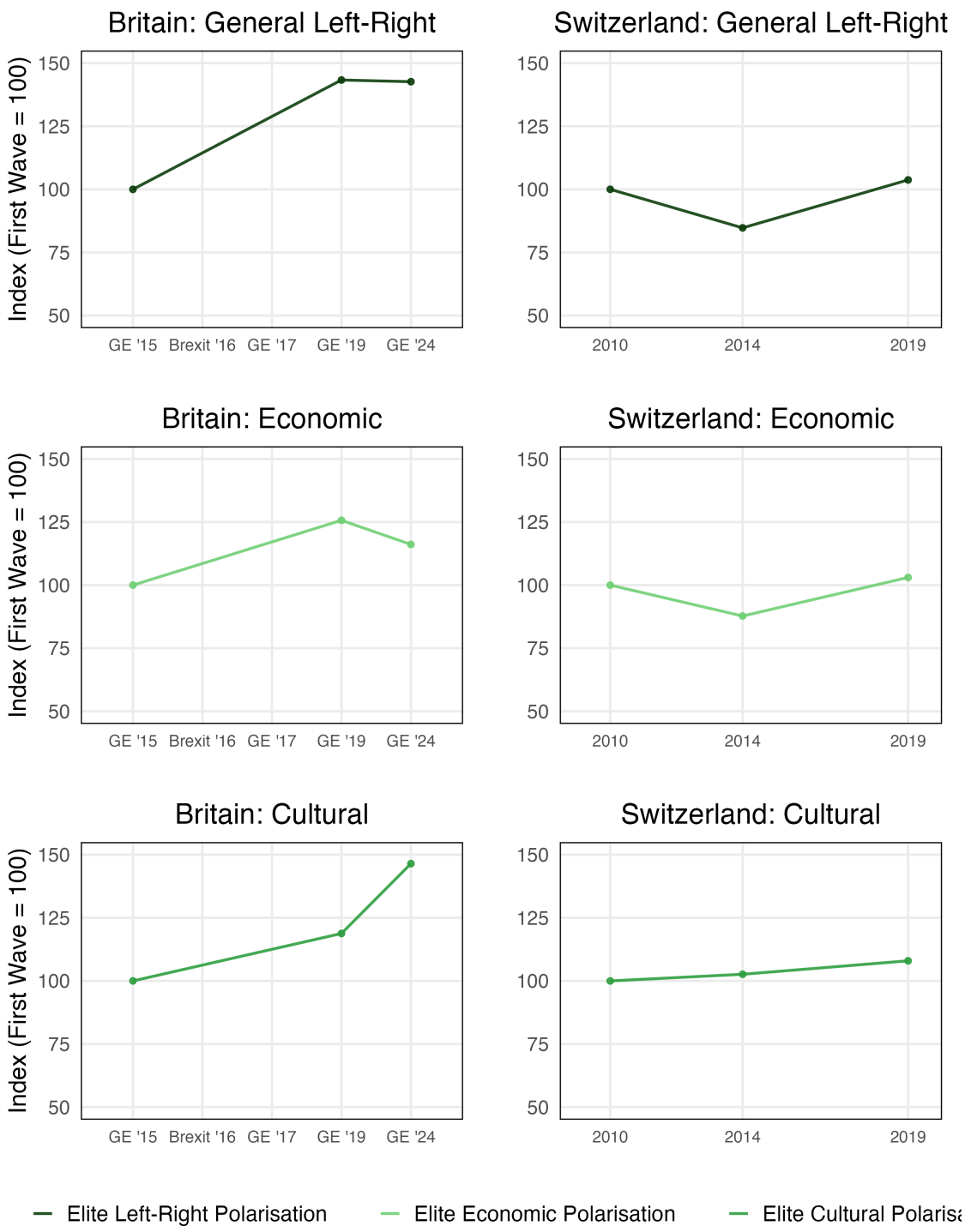


Figure 6: Elite ideological polarisation. Vote-share weighted standard deviation of party positions on the general left–right, economic left–right and cultural dimensions, as assessed by country experts in the Chapel Hill Expert Survey (Rovny et al., 2025). Weights are each party's vote share at the nearest election. UK: 2014, 2019 and 2024 waves, mapped to the nearest election. Switzerland: 2010, 2014 and 2019 waves. Indexed to 100 in the first wave. Expert coverage is reported in Table 7.

5.3 Robustness checks

Overall, we find evidence of hidden rising inequalities in two of our five cases: life satisfaction in Germany, and political self-placement in the UK. We now address two distinct sources of concern. First, is the decline in *observed* response volatility real, or an artefact of sample composition, panel conditioning or scale mechanics? Second, even granting that the observed decline is real, do our identifying assumptions correctly translate it into a claim about *latent* dispersion? We address the first in Sections 5.3.1–5.3.5, and the second in Section 5.3.6.

5.3.1 Sensitivity to model specification

Tables 4 and 5 report the linear time trend in indexed volatility under seven specifications. The trend is negative and statistically significant at the 5 per cent level in all five cases under every specification. Controlling for panel tenure (Models 3–5) shrinks the trend substantially in Australia and in both Swiss series, but barely in Germany. In the UK, controls reduce the trend by about a quarter.

Table 4: Indexed Time Trends in Within-Person Variation – Life Satisfaction

Model	Specification	Germany	Australia	Switzerland
1	No Controls	-0.673^{***}	-0.985^{***}	-0.976^{***}
		(0.064)	(0.141)	(0.103)
2	Demographic Controls	-0.671^{***}	-0.991^{***}	-1.012^{***}
		(0.061)	(0.135)	(0.102)
3	Linear Panel Tenure	-0.616^{***}	-0.383^{*}	-0.461^{***}
		(0.061)	(0.131)	(0.102)
4	Panel Tenure (Factor)	-0.658^{***}	-0.397^{***}	-0.425^{*}
		(0.052)	(0.091)	(0.151)
5	Panel Tenure (Factor) + Demographics	-0.657^{***}	-0.262^{**}	-0.371^{*}
		(0.051)	(0.084)	(0.155)
6	New Respondents Only	-0.844^{***}	-0.725^{***}	-0.863^{*}
		(0.097)	(0.162)	(0.306)
7	Balanced Panel	-0.523^{***}	-1.089^{***}	-0.733^{***}
		(0.116)	(0.164)	(0.156)

Note: Coefficient from an OLS regression of indexed within-person volatility on years since panel start, one observation per period. Dependent variable: $|\Delta\text{LS}|$ (absolute year-to-year change in life satisfaction), indexed to 100 in the first period. Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5: Indexed Time Trends in Within-Person Variation – Political Attitudes

Model	Specification	Switzerland (SHP)	UK (BES)
1	No Controls	-1.241^{***}	-7.935^{**}
		(0.112)	(0.945)
2	Demographic Controls	-1.285^{***}	-7.949^{**}
		(0.109)	(0.952)
3	Linear Panel Tenure	-0.459^{***}	-5.617^{**}
		(0.106)	(0.616)
4	Panel Tenure (Factor)	-0.450^{**}	-6.894^{**}
		(0.153)	(0.966)
5	Panel Tenure (Factor) + Demographics	-0.403^{*}	-5.893^{*}
		(0.155)	(1.177)
6	New Respondents Only	-0.904^{**}	-5.816^{*}
		(0.292)	(1.407)
7	Balanced Panel	-1.226^{***}	-9.866^{**}
		(0.161)	(1.668)

Note: Coefficient from an OLS regression of indexed within-person volatility on years (or election cycles for BES) since panel start, one observation per period. Dependent variable: $|\Delta \text{LR}|$ (absolute change in left-right self-placement), indexed to 100 in the first period. Switzerland: yearly trends. UK: election-cycle trends; for the UK, Model 5 also controls for the number of days until the next survey. UK coefficients are larger because they are per election cycle, and cycles are one to five years apart; Table 10 reports a quasi-yearly version. Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5.3.2 Sensitivity to volatility specification in the BES

In the BES, our main analysis measures volatility around each election rather than across years. Table 10 and Figure 12 report a yearly version that is comparable with the other panels. The yearly measure declines by about 23 per cent between 2014 and 2024. Volatility also falls at a similar rate whether measured between two pre-election waves or across the election (-5.5 and -6.5 points per election cycle respectively).

5.3.3 Attrition

Attrition could produce a spurious decline in volatility: for life satisfaction, Headey and Muffels (2017, p. 5) note that respondents with less stable lives drop out of panels disproportionately. The same may hold for individuals with unstable political views. We address this in two ways. First, we restrict attention to each individual's first two observations (Model 6), which exploits the regular refresher samples in each panel. Second, we repeat the analysis on balanced panels of individuals observed in every period (Model 7), although panel tenure cannot be controlled for because it is collinear with time.

Both approaches shrink the sample substantially (Table 6), but downward trends persist in all five cases and remain statistically significant (Tables 4 and 5; further plots in Appendix A). The declines are generally larger than under Model 5, the exceptions being the German balanced panel

(−0.52 against −0.66) and UK new respondents (−5.8 against −5.9).

5.3.4 A comparison item: self-rated health

A further way of addressing attrition and panel-conditioning concerns is to examine volatility trends in another variable. We look at self-rated health in SOEP and HILDA, a five-point question with labelled categories (poor, fair, good, very good, excellent); there is no equivalent item in SHP or the BES. The volatility of self-rated health rises by about 12 per cent in Germany between 1994 and 2019, and is roughly flat in Australia (Appendix, Figure 9). Falling volatility is therefore not a general feature of ordinal self-reports in these panels, implying that neither attrition nor panel conditioning alone drives our results for wellbeing and ideology.

5.3.5 Bunching effects

A further concern is that falling volatility is a mechanical consequence of the response distribution: if mean responses rise, respondents bunch closer to the scale's ceiling, with less room to move, and measured volatility shrinks (Goff et al., 2018). Floor effects could occur with falling means. The data do not support this interpretation. Nominal means are flat throughout our sample for both variables (Tables 2 and 3). The shares of respondents at 0/10 and 10/10 are generally *falling* (Figures 11a–11b, Appendix A.3). This holds for life satisfaction in all three countries, as Barrington-Leigh (2024) documents, and for ideology in Switzerland, though not in the UK.

5.3.6 Violations of Assumptions

Besides assuming the existence of a latent (Assumption 0) and of a common reporting function across respondents (Assumption 2), the V-LINE correction of latent dispersion depends on three other identifying assumptions (Section 2). Here, we consider possible violations of each assumption, and whether they would overturn our conclusions.

Assumption 1: Linear Reporting

In our exercise, we assume that equal steps on the reported scale correspond to equal steps in the latent quantity. Kaiser and Lepinteur (2025) estimate how far reporting departs from linearity in numerical response scales. We simulate non-linear reporting functions at magnitudes they consider plausible (Appendix E); even the most severe does not overturn our divergence results for Germany and the UK. This robustness has an intuitive explanation: non-linearity tends to move response volatility and nominal dispersion in the same direction, so the two effects partially cancel in the ratio that estimates latent dispersion. For example, if a popular category such as 7/10 covers a wider band of the latent scale, more respondents report it and fewer move off it between waves, reducing both dispersion and volatility.

Assumption 3: No Within-Period Rescaling

We assume the reporting function is fixed within each two-wave period. Appendix F derives an exact decomposition separating genuine latent change from the bias a continuous drift in reporting would introduce. For happiness, the bias is an order of magnitude smaller than our identifying signal

(a signal-to-noise ratio of roughly 10–20:1). A reasonable level of accuracy should be unsurprising because Assumption 3 corresponds to a step-wise approximation of a continuous drift.

Assumption 4: Conditional Homoskedasticity

Our most important assumption is that individual volatility in latent units is constant over time conditional on controls.

Yet, true latent volatility would need to have fallen by a significant fraction to reverse our results. Even a fall in true volatility of up to 14% in Germany or up to 21% in the UK leaves both hidden divergences intact (Appendix G). These numbers serve as lower bounds to the extent that response volatility contains measurement error which is constant across time.

While we cannot exclude that reporting precision has increased over time, we find no evidence of mechanical reasons for this decline. If falls in genuine state volatility drove the changes, there would need to be separate stories for ideology and wellbeing, with parsimony favouring the stretching interpretation which jointly explains both.

6 Surveys could also hide changing levels

On its own, the V-LINE tells us how much the scale has stretched, but not where it sits: whether a given response (e.g., 3/10) corresponds to a happier life or a more right-wing position than it once did. Recovering the latent *mean* therefore requires an additional assumption about which point on the scale stays fixed for each case. In this section, we briefly discuss what we can learn from the V-LINE once we assume a stable anchor. Results are reported in detail in Appendix D.

Wellbeing. A plausible anchor for wellbeing is that the bottom of the scale stays fixed: the latent wellbeing of a 0/10 life is unchanged, while that of a 10/10 life has risen. This 'upward stretching' was hypothesised by Kahneman and Krueger (2006), Deaton (2011) and Clark et al. (2012), and is the modal way people describe changes in their own scale use (Prati and Senik, 2026). Under this anchor,[19] latent life satisfaction in Germany rose by about 45% between 1984–1985 and 2018–2019, an average annual growth rate of about 1.1 per cent. In Australia, it rose by about 8 per cent in the early 2000s and then levels off. In Switzerland, it shows no clear trend. The Easterlin Paradox rests on flat series of reported life satisfaction during sustained income growth (Easterlin, 1974). Once rescaling is corrected for, the paradox does not hold in Germany, nor in Australia during the early 2000s, while in Switzerland, the evidence is consistent with it. However, it should be noted that if the latent wellbeing of a 0/10 life has itself fallen, these estimates would overstate the rise.

Our estimates are comparable to those of Prati and Senik (2026), who find that latent life satisfaction in the US grew by about 1.2 per cent a year between 1959 and 2008. Their method

[19] Interestingly, this anchor works very well for income. If we assume that the bottom income category did not shift, the V-LINE tracks the rise in mean income quite closely, because real incomes at the bottom of the distribution are indeed close to flat in our samples (Appendix Figure 13). Income categories also fail to capture the rise in mean household income.

differs in that it uses retrospective satisfaction reports from repeated cross-sections, and makes no assumption about respondents' latent volatility.

Importantly, latent wellbeing does not always rise after correcting for stretching. The COVID-19 pandemic is an instructive case: the interpretation of the "best possible life" plausibly shifted downward rather than upward. Between 2018–2019 and 2021–2022, response volatility *rose*, implying that the scale *contracted* — by about 2 per cent in Germany and 5 per cent in Switzerland. With nominal life satisfaction flat, the adjusted mean fell by about 2.5 per cent in Germany and 6 per cent in Switzerland (Figures 15 and 16): a welfare cost that raw reports miss. Prati and Senik (2026) find the same pattern using retrospective questions. They also find that in Germany, latent satisfaction fell in 2020 even as nominal satisfaction rose, with similar results for France.

Ideology. A fixed bottom anchor makes little sense for self-placement: it would require the ideology of the far Left to be unchanged. A fixed midpoint, the modal response in both surveys (Figure 11), is more reasonable. Because nominal means stay close to 5 (Table 3), stretching then barely affects average ideology: in the UK, it moves the adjusted mean by less than 0.1 of a point. Asymmetric shifts, with one side of the ideology scale moving further than the other, would go undetected. We trust this anchor less than the wellbeing one because it is not yet supported by respondents' own accounts of how they use the scale. Collecting such evidence is a natural route for future work.

7 Conclusion

In this paper, we show that, in contexts where self-anchoring scales may be subject to rescaling, the conjunction of puzzling inequality dynamics and puzzling volatility dynamics may be a measurement artefact. We develop a simple model to understand the impact of rescaling on estimates of inequality trends and propose a novel method to correct for it. When we apply this method to subjective ratings displaying puzzling inequality trends, the puzzles largely disappear.

There is no definitive way to establish which data series – the uncorrected one or the V-LINE – is closer to the ground truth. Both series recover a latent quantity by relying on assumptions that cannot be directly verified. If the reporting scale is assumed to be immutable despite changes in context, then self-reported ratings should not be corrected. If, instead, (conditional) latent volatility is assumed to remain constant, then the V-LINE provides a better representation of true inequality. Both assumptions are plausible, yet also likely to be violated to some extent. The ground truth is therefore probably somewhere in between. If both rescaling and a moderate decline in latent volatility occur, however, survey data will still tend to hide rising inequality (see Appendix G).

To be sure, rescaling is not the only explanation of the empirical puzzles that we document. For instance, happiness inequality might have genuinely declined because of lower inequality in non-material domains, or because of heterogeneity in tastes that is not captured by material inequality (Clark et al., 2012). In parallel, latent volatility could have genuinely fallen because, e.g., unemployment became less common (OECD, 2026) or because wider antidepressant use blunted emotional

responsiveness (Goodwin et al., 2017). Ideological polarisation in the UK might have genuinely shrunk as economic and cultural divisions have become less tightly aligned (Tasker, 2026) or as partisan identities weakened (Fieldhouse et al., 2019). And latent ideological volatility may have fallen as clearer party divergence after 2015 made it easier for British voters to see which positions belong to which side (Freeder et al., 2019). Besides, declining volatility is compatible with reduced measurement error if respondents report more precisely over time for reasons unrelated to scale stretching, for instance stronger numerical education (Barrington-Leigh, 2024). And these are just a few of many possible theories.

If the ground truth cannot be established, why should researchers use the V-LINE? Parsimony. Four puzzles need explaining: unexpected trends in inequality and in volatility for both happiness and ideology. Competing theories can accommodate these one at a time, while the V-LINE provides a single explanation for all four. By Occam's razor, that is reason to take rescaling – and the correction we propose – seriously.

Importantly, we do not claim that all inequality trends derived from self-reported ratings are unreliable. Although any self-anchoring scale may be subject to rescaling, small amounts of rescaling will have negligible effects on the estimands in many applications. Indeed, this paper documents cases in which the V-LINE leads to conclusions similar to those obtained from the uncorrected time series (e.g., the absence of increasing voter polarisation in Switzerland). In particular, rescaling is likely to be negligible when studying short time periods during which there is little reason to expect shocks to scale interpretation (but see Prati and Senik 2026 for rescaling during social shocks).

Instead, we recommend using the V-LINE when three conditions are met. First, there should be good theoretical reasons to suspect that rescaling may occur. We believe the most fruitful applications of the V-LINE are contexts in which rescaling is a serious theoretical concern but could not be tested for empirically using existing approaches. In this sense, we envisage the V-LINE primarily as a diagnostic tool for scales that are theoretically vulnerable to rescaling, rather than as a data-mining tool for discovering which scales suffer from it. Second, the time series should display a conjunction of two empirical puzzles: (i) a puzzling decline (or increase) in inequality and (ii) a puzzling decline (or increase) in respondents' response volatility. Third, suitable panel data must be available to estimate the V-LINE.

What is needed to calculate the V-LINE? The key technical requirement is panel data: participants must be interviewed repeatedly over time. Ideally, the panel should include frequent refreshment samples – as in the German SOEP – to help disentangle panel conditioning from rescaling. Importantly, the V-LINE requires no additional survey questions. Existing methods for correcting scale use typically rely on supplementary questions (Prati and Senik, 2026; Benjamin et al., 2026). Yet social survey questionnaires are subject to strict constraints, and adding questions can be costly or simply unfeasible. The V-LINE can therefore be applied to a wide range of existing self-reported panel data in which rescaling is a concern. We hope this paper will encourage researchers to reconsider the interpretation of inequality trends in self-reported ratings, and open up a new research avenue.

References

Aldrich, J. H. and McKelvey, R. D. (1977). A method of scaling with applications to the 1968 and 1972 presidential elections. *American Political Science Review*, 71(1):111–130.

Alvaredo, F., Bourguignon, F., Ferreira, F. H., Lakner, C., and Lustig, N. (2026). Is income inequality on the rise everywhere?

Ambrey, C. L. and Fleming, C. M. (2014). Life satisfaction in Australia: Evidence from ten years of the HILDA survey. *Social Indicators Research*, 115(2):691–714.

Bakker, R., de Vries, C., Edwards, E., Hooghe, L., Jolly, S., Marks, G., Polk, J., Rovny, J., Steenbergen, M., and Vachudova, M. A. (2015). Measuring party positions in Europe: The Chapel Hill expert survey trend file, 1999–2010. *Party Politics*, 21(1):143–152.

Bakker, R., Edwards, E., Jolly, S., Polk, J., Rovny, J., and Steenbergen, M. (2014). Anchoring the experts: Using vignettes to compare party ideology across countries. *Research & Politics*, 1(3):2053168014553502.

Barrington-Leigh, C. P. (2024). The econometrics of happiness: Are we underestimating the returns to education and income? *Journal of Public Economics*, 230:105027.

Benjamin, D. J., Cooper, K., Heffetz, O., and Kimball, M. (2024). From happiness data to economic conclusions. *Annual Review of Economics*, 16(1):359–391.

Benjamin, D. J., Cooper, K. B., Heffetz, O., Kimball, M. S., and Zhou, J. (2026). Correcting for heterogeneity in the use of subjective scales. Nber working paper, National Bureau of Economic Research.

Bernaerts, K., Blanckaert, B., and Caluwaerts, D. (2023). Institutional design and polarization: Do consensus democracies fare better in fighting polarization than majoritarian democracies? *Democratization*.

Çakır, S. (2025). How do masses react to party polarization? Limited effect of party polarization on mass polarization. *European Journal of Political Research*, 64(3):1533–1548.

Chinoy, S., Nunn, N., Sequeira, S., and Stantcheva, S. (2026). Zero-sum thinking and the roots of US political differences. *American Economic Review*, 116(3):1052–1096.

Clark, A. E., Flèche, S., and Senik, C. (2012). The great happiness moderation. IZA Discussion Paper 6761, Institute of Labor Economics (IZA).

Dalton, R. J. (2008). The quantity and the quality of party systems: Party system polarization, its measurement, and its consequences. *Comparative Political Studies*, 41(7):899–920.

Deaton, A. (2011). Comment on "Work disability, work, and justification bias in Europe and the U.S.". In Wise, D. A., editor, *Explorations in the Economics of Aging*, pages 312–314. University of Chicago Press, Chicago, IL.

Dow, J. K. (2011). Party-system extremism in majoritarian and proportional electoral systems. *British Journal of Political Science*, 41(2):341–361.

Dutta, I. and Foster, J. (2013). Inequality of happiness in the U.S.: 1972–2010. *Review of Income and Wealth*, 59(3):393–415.

D'Ambrosio, C., Jäntti, M., and Lepinteur, A. (2020). Money and happiness: Income, wealth and subjective well-being. *Social Indicators Research*, 148(1):47–66.

Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In David, P. A. and Reder, M. W., editors, *Nations and Households in Economic Growth*, pages 89–125. Academic Press, New York.

Enke, B., Rodríguez-Padilla, R., and Zimmermann, F. (2023). Moral universalism and the structure of ideology. *The Review of economic studies*, 90(4):1934–1962.

Ezrow, L. (2007). The variance matters: How party systems represent the preferences of voters. *Journal of Politics*, 69(1):182–192.

Fabian, M. (2022). Scale norming undermines the use of life satisfaction scale data for welfare analysis. *Journal of Happiness Studies*, 23(4):1509–1541.

Fieldhouse, E., Green, J., Evans, G., Mellon, J., Prosser, C., Schmitt, H., and van der Eijk, C. (2019). *Electoral Shocks: The Volatile Voter in a Turbulent World*. Oxford University Press, Oxford.

Fiorina, M. P. and Abrams, S. J. (2008). Political polarization in the american public. *Annual Review of Political Science*, 11:563–588.

Freeder, S., Lenz, G. S., and Turney, S. (2019). The importance of knowing "what goes with what": Reinterpreting the evidence on policy attitude stability. *The Journal of Politics*, 81(1):274–290.

Frijters, P. and Beatton, T. (2012). The mystery of the U-shaped relationship between happiness and age. *Journal of Economic Behavior & Organization*, 82(2-3):525–542.

Gethin, A., Martínez-Toledano, C., and Piketty, T. (2022). Brahmin left versus merchant right: Changing political cleavages in 21 western democracies, 1948–2020. *The Quarterly Journal of Economics*, 137(1):1–48.

Goff, L., Helliwell, J. F., and Mayraz, G. (2018). Inequality of subjective well-being as a comprehensive measure of inequality. *Economic Inquiry*, 56(4):2177–2194.

Goodwin, G. M., Price, J., De Bodinat, C., and Laredo, J. (2017). Emotional blunting with antidepressant treatments: A survey among depressed patients. *Journal of Affective Disorders*, 221:31–35.

Hare, C., Armstrong, D. A., Bakker, R., Carroll, R., and Poole, K. T. (2015). Using Bayesian Aldrich-McKelvey scaling to study citizens' ideological preferences and perceptions. *American Journal of Political Science*, 59(3):759–774.

Harrison, C. (2025). Scaling the Easterlin paradox: Measuring life events on stretching happiness measures. Working Paper 2502, Wellbeing Research Centre, University of Oxford.

Headey, B. and Muffels, R. (2017). Towards a theory of medium term life satisfaction: Similar results for Australia, Britain and Germany. *Social Indicators Research*, 134(1):359–384.

Helliwell, J. F., Huang, H., and Wang, S. (2016). The distribution of world happiness. In Helliwell, J. F., Layard, R., and Sachs, J., editors, *World Happiness Report Update 2016*, chapter 2. Sustainable Development Solutions Network.

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., and Wang, S. (2026). Why single-item measures of wellbeing are best. *Nature Human Behaviour*, 10:437–438.

Hill, S. J. and Tausanovitch, C. (2015). A disconnect in representation? comparison of trends in congressional and public polarization. *The Journal of Politics*, 77(4):1058–1075.

Hix, S., Noury, A., and Roland, G. (2006). Dimensions of politics in the european parliament. *American Journal of Political Science*, 50(2):494–520.

Ipsos (2025). Divisions: Charts. Technical report, Ipsos.

Jantsch, A., Le Blanc, J., and Schmidt, T. (2024). Beyond income: Exploring the role of household wealth for subjective well-being in germany. *Journal of Happiness Studies*, 25(7):101.

Jenkins, S. P. (2021). Inequality comparisons with ordinal data. *Review of Income and Wealth*, 67(3):547–563.

Johnston, C. D. and Ollerenshaw, T. (2020). How different are cultural and economic ideology? *Current Opinion in Behavioral Sciences*, 34:94–101.

Kahneman, D. and Krueger, A. B. (2006). Developments in the measurement of subjective well-

being. *Journal of Economic Perspectives*, 20(1):3–24.

Kaiser, C. (2022). Using memories to assess the intrapersonal comparability of wellbeing reports. *Journal of Economic Behavior & Organization*, 193:410–442.

Kaiser, C. and Lepinteur, A. (2025). Measuring the unmeasurable? Systematic evidence on scale transformations in subjective survey data. IZA Discussion Paper 18029, Institute of Labor Economics (IZA).

Kalmijn, W. and Veenhoven, R. (2005). Measuring inequality of happiness in nations: In search for proper statistics. *Journal of Happiness Studies*, 6(4):357–396.

King, G., Murray, C. J. L., Salomon, J. A., and Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1):191–207.

Kubiszewski, I., Zakariyya, N., Costanza, R., and Jarvis, D. (2020). Resilience of self-reported life satisfaction: A case study of who conforms to set-point theory in Australia. *PLoS One*, 15(8):e0237161.

Lehmann, P., Franzmann, S., Al-Gaddooa, D., Burst, T., Ivanusch, C., Regel, S., Riethmüller, F., Volkens, A., Weßels, B., and Zehnter, L. (2024). The manifesto data collection. manifesto project (mrg/cmp/marpor). version 2024a.

Lindberg, S. I., Düpont, N., Higashijima, M., et al. (2022). Codebook: Varieties of party identity and organization (V-Party) V2. Technical report, Varieties of Democracy (V-Dem) Project.

Moral, M. and Best, R. E. (2023). On the relationship between party polarization and citizen polarization. *Party Politics*, 29(2):229–247.

Ng, Y.-K. (2008). Happiness studies: Ways to improve comparability and some public policy implications. *Economic Record*, 84(265):253–266.

OECD (2026). Infra-annual labour statistics: Unemployment rate, total, 15 years or over. OECD Data Explorer, `https://data-explorer.oecd.org/`. Accessed 13 November 2026.

Oswald, A. J. (2008). On the curvature of the reporting function from objective reality to subjective feelings. *Economics Letters*, 100(3):369–372.

Perez-Truglia, R. (2020). The effects of income transparency on well-being: Evidence from a natural experiment. *American Economic Review*, 110(4):1019–1054.

Prati, A. and Senik, C. (2026). Is it possible to raise national happiness? *Journal of Public Economics*, 257.

Rovny, J., Polk, J., Bakker, R., Hooghe, L., Jolly, S., Marks, G., Steenbergen, M., and Vachudova, M. A. (2025). The 2024 Chapel Hill Expert Survey on political party positioning in Europe: Twenty-five years of party positional data. *Electoral Studies*, 97.

Seré, M. and Decancq, K. (2025). Room for happiness? cross-country variation in response scale. Mimeo.

Stevenson, B. and Wolfers, J. (2008). Happiness inequality in the United States. *The Journal of Legal Studies*, 37(S2):S33–S79.

Tasker, J. (2026). The changing meaning of left-right self-placement in the united kingdom. *Political Studies*. Published online September 16, 2026.

VanderWeele, T. J. and Johnson, B. R. (2025). Multidimensional versus unidimensional approaches to well-being. *Nature Human Behaviour*, 9:857–863.

Vivalt, E., Rhodes, E., Bartik, A., Broockman, D., Krause, P., and Miller, S. (2026). The employment effects of a guaranteed income: Experimental evidence from two US states. *The Quarterly Journal of Economics*, 141(4):2675–2734.

World Inequality Database (2026). WID – World Inequality Database. https://wid.world. Accessed 9 November 2026.

A Additional Figures and Tables

A.1 Panel Statistics

Table 6 summarises the structure and demographic composition of each estimation sample. Table 7 reports the number of CHES experts behind each elite polarisation measure.

Table 6: Estimation Samples: Panel Structure and Demographic Composition

	Germany	Australia	Switzerland		UK
Measure	**(SOEP)**	**(HILDA)**	**(SHP, LS)**	**(SHP, LR)**	**(BES)**
Years	1984–2019	2001–2019	2000–2019	1999–2023	2015–2024
Individuals	65,363	25,633	18,377	17,068	45,535
Mean Waves	9.7	10.4	8.2	9.4	6.3
Observations	573,229	234,880	133,380	130,524	142,487
New Respondent Obs.	118,363	46,586	31,775	28,125	24,648
Balanced Panel Obs.	28,489	96,386	37,243	30,549	2,490
Mean Age	40.2	38.5	41.0	41.4	52.3
Female (%)	51.9	51.6	46.7	48.4	48.7
Migrant (%)	19.1	22.4	15.7	14.5	—

Note: Columns describe the life satisfaction (LS) samples for SOEP, HILDA and SHP, and the left-right (LR) samples for SHP and BES. Years = first and last year with a valid response. Individuals, Mean Waves and Observations refer to the Model 1 sample (no controls). Mean Waves = valid responses per individual (BES: waves, out of 15, with a valid left-right answer). Observations = within-person changes ($|\Delta\text{LS}|$ or $|\Delta\text{LR}|$) used in estimation. New Respondent Obs. = observations in the Model 6 sample, from each individual's first or second wave. Balanced Panel Obs. = observations in the Model 7 sample: individuals observed in every period (BES: at least two valid $|\Delta\text{LR}|$ in each of the five elections). Demographics are measured at each individual's first observation in the Model 1 sample. Migrant background is not available for BES.

A.2 Volatilities: Other Models

Figures 7 and 8 show volatility trends for the two attrition checks in Section 5.3.3.

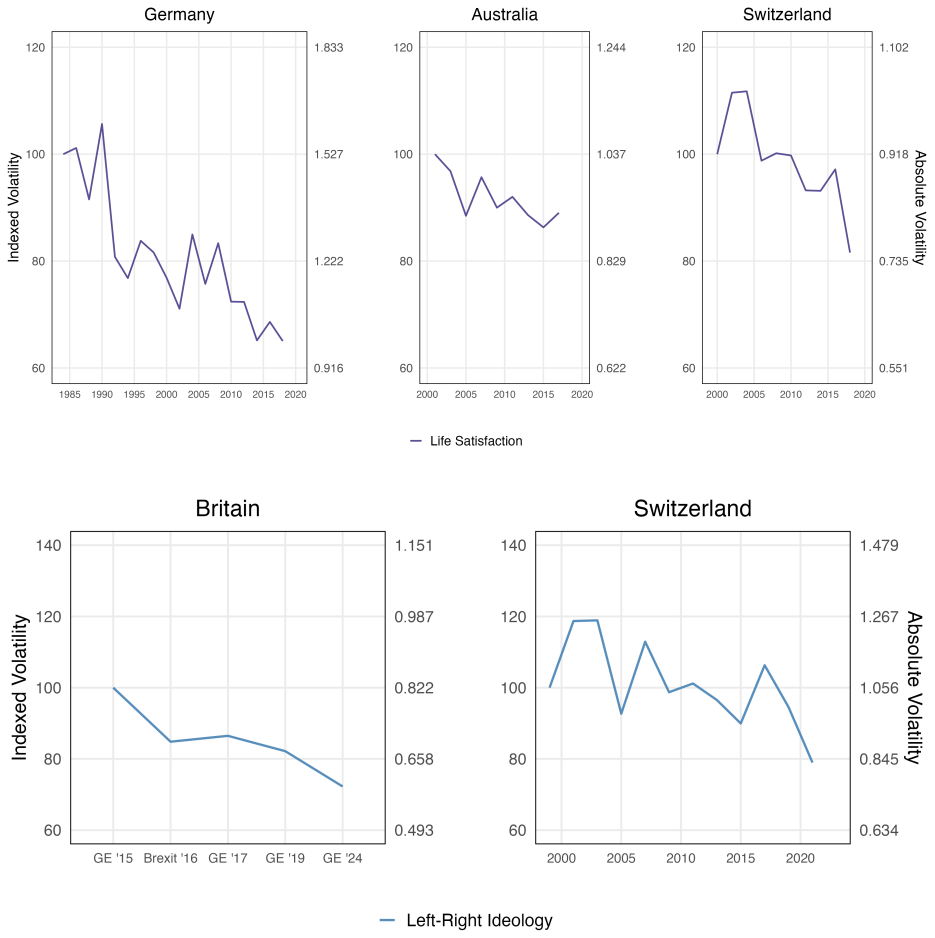


Figure 7: Within-person volatility for new respondents only (Model 6): each individual's first two observations. Top: life satisfaction. Bottom: left-right self-placement. Samples are small, particularly for Switzerland.

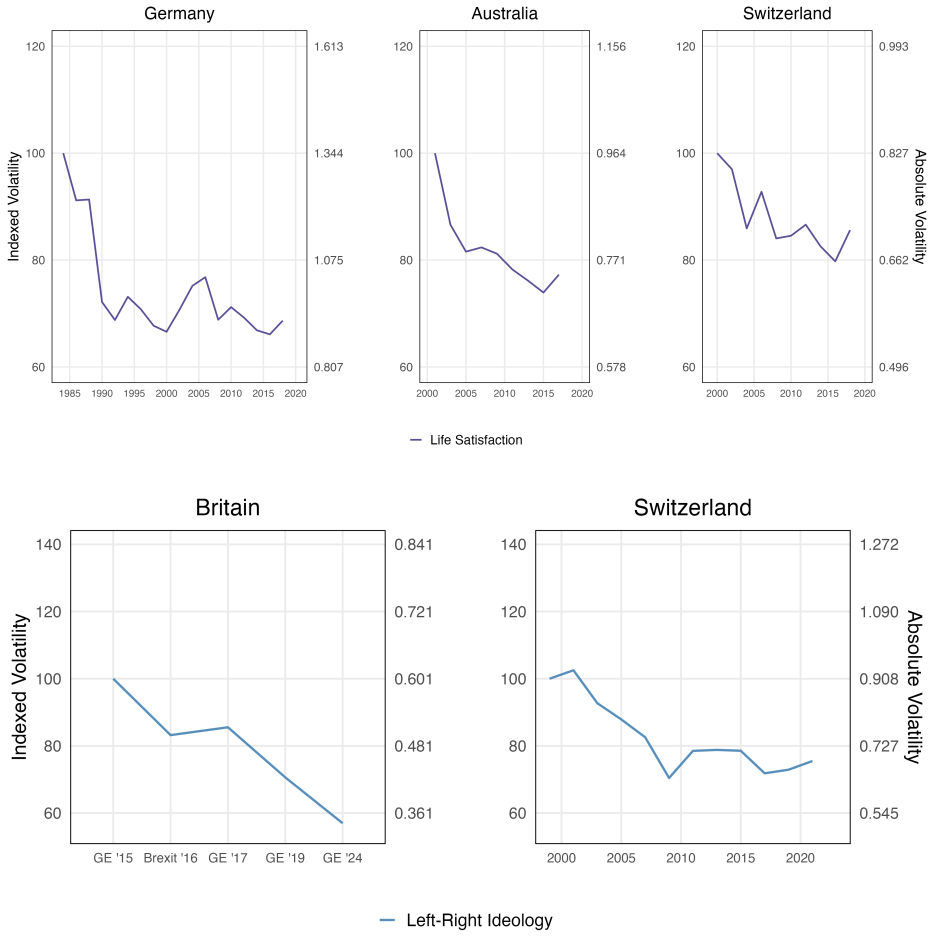


Figure 8: Within-person volatility for balanced panels (Model 7): individuals observed in every period. Top: life satisfaction. Bottom: left-right self-placement.

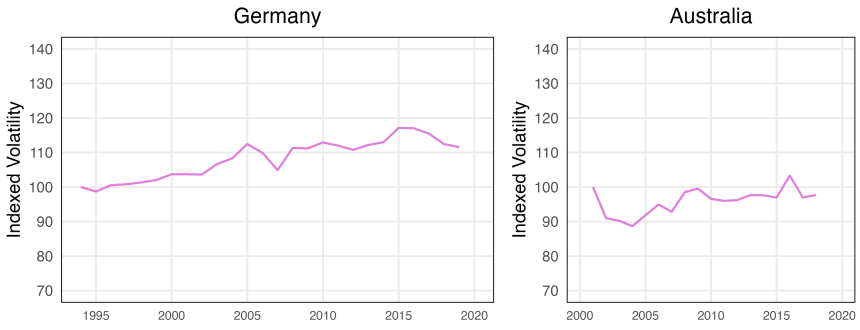


Figure 9: Volatility trend for self-rated health. Model 5 (panel tenure and demographic controls), indexed to 100 in the first period. Left: Germany. Right: Australia.

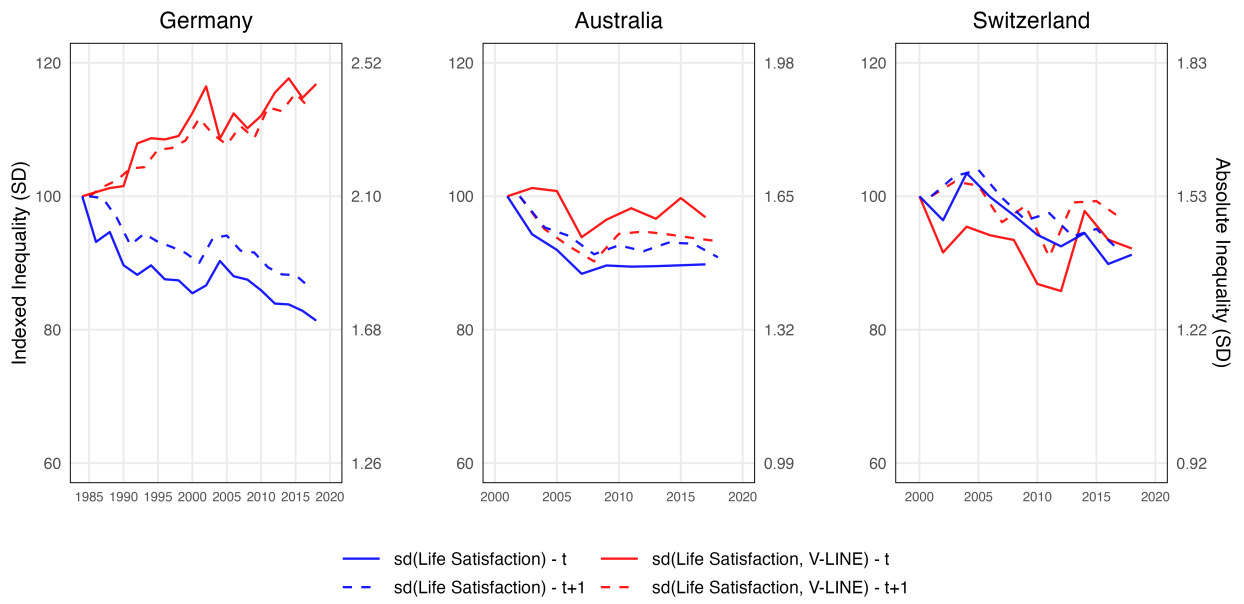
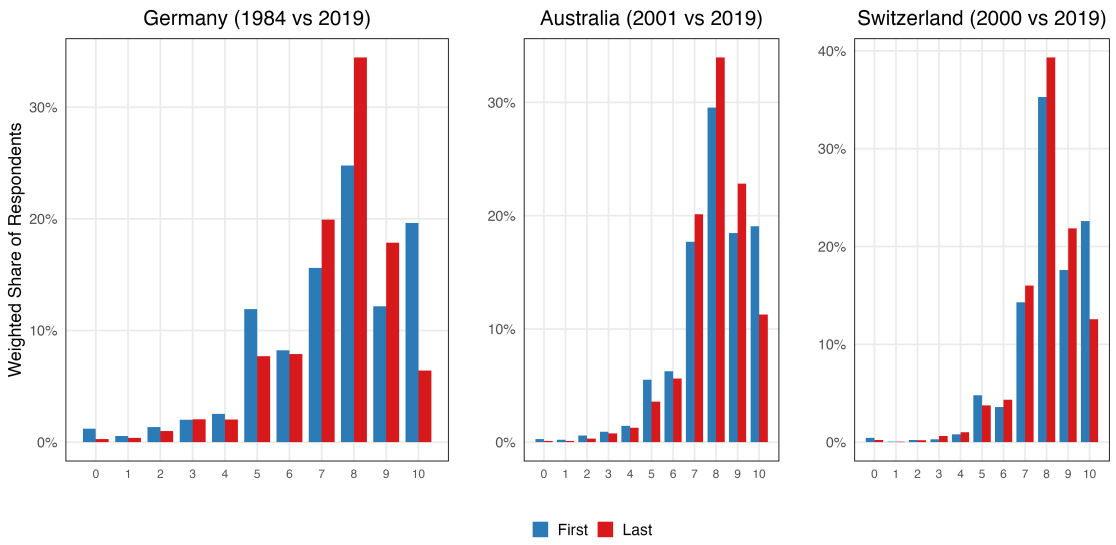


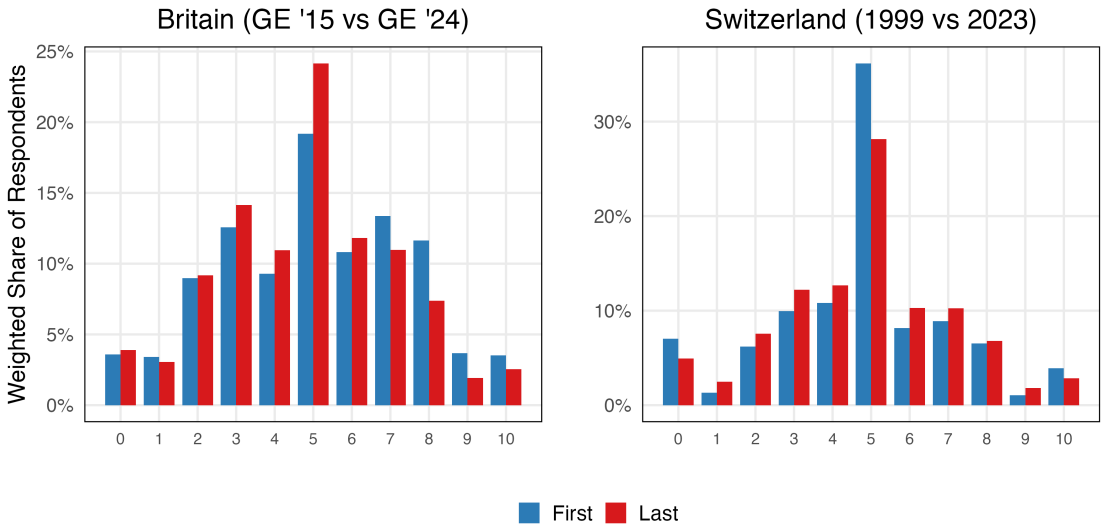
Figure 10: Sensitivity to the definition of two-year periods: life satisfaction inequality (SD). Solid lines: periods starting in the first survey year (t; e.g. 1984–1985, 1986–1987, ...). Dashed lines: periods starting one year later ($t + 1$; e.g. 1985–1986, 1987–1988, ...). Blue: nominal; red: V-LINE-adjusted, using Model 5 volatility. Each series is indexed to 100 in its own first period; right axis in first-period scale units.

A.3 Raw Response Distributions: First vs. Last Period

Figure 11 shows the full weighted distribution of raw responses (0–10) in the first and last period of each panel.



(a) Life satisfaction: Germany (1984 vs. 2019), Australia (2001 vs. 2019), Switzerland (2000 vs. 2019).



(b) Left-right self-placement: UK (GE '15 vs. GE '24), Switzerland (1999 vs. 2019).

Figure 11: Distribution of raw responses, first vs. last period.

B Elite Polarisation Data

B.1 Elite polarisation

Table 7: CHES Expert Coverage by Country and Wave

Country	Wave	Left-right	Economic	Cultural
Switzerland	2010	12	12	12
	2014	8	8	8
	2019	11	11	11
UK	2014	7	7	7
	2019	17	17	17
	2024	21	21	21

Note: Number of experts answering each item: overall left-right position (used in the main analysis), economic left-right, and the cultural dimension.

Wording of Elite Polarisation Questions

The wording in CHES has been essentially unchanged since 1999 (Bakker et al., 2015). All three items use a 0–10 scale with 5 as the centre. Experts are asked to place a party's leadership (e.g., the party chair/the parliamentary party, distinct from the party base) on:

General left–right. "Please tick the box that best describes each party's overall ideology on a scale ranging from 0 (extreme left) to 10 (extreme right)."

Economic left–right. "Parties can be classified in terms of their stance on economic issues. Parties on the economic left want government to play an active role in the economy. Parties on the economic right emphasize a reduced economic role for government: privatization, lower taxes, less regulation, less government spending, and a leaner welfare state." An 11-point scale ranges from 0 (extreme left) to 5 (center) to 10 (extreme right).

Cultural. "Parties can be classified in terms of their views on democratic freedoms and rights. 'Libertarian' or 'postmaterialist' parties favor expanded personal freedoms, for example, access to abortion, active euthanasia, same-sex marriage, or greater democratic participation. 'Traditional' or 'authoritarian' parties often reject these ideas; they value order, tradition, and stability, and believe that the government should be a firm moral authority on social and cultural issues." An 11-point scale ranges from 0 (libertarian/postmaterialist) to 5 (center) to 10 (traditional/authoritarian).

C BES analysis with alternative period definition

In the main analyses, our measure is not yearly, it is Election Volatility. In this section, we adopt a longer-run measure, *Quasi-Yearly Volatility*. Because BES fieldwork months vary from year to year, we select one post-election survey per year — drawn from local, national, or referendum contests — with most selections clustering around the middle of the year (Table 9).

Our main analysis relies on Election Volatility rather than Quasi-Yearly Volatility, for three reasons. First, each election contributes two volatility observations. Second, the time gaps between election-year waves are naturally regular, unlike the gaps in the quasi-yearly panel. Third, Election Volatility captures change over a short time-span, which is unique and interesting.

Table 8: Election Cycle Analysis: British Election Study

Election	Event	Event Date	Position	Wave	Field Dates
2015 GE	General Election	7 May 2015	Pre-election	W4	4 Mar – 30 Mar 2015
			Campaign	W5	31 Mar – 6 May 2015
			Post-election	W6	8 May – 26 May 2015
2016 Brexit	EU Referendum	23 Jun 2016	Pre-election	W7	14 Apr – 4 May 2016
			Campaign	W8	6 May – 22 Jun 2016
			Post-election	W9	24 Jun – 4 Jul 2016
2017 GE	General Election	8 Jun 2017	Pre-election	W11	24 Apr – 3 May 2017
			Campaign	W12	5 May – 7 Jun 2017
			Post-election	W13	9 Jun – 23 Jun 2017
2019 GE	General Election	12 Dec 2019	Pre-election	W17	1 Nov – 12 Nov 2019
			Campaign	W18	13 Nov – 11 Dec 2019
			Post-election	W19	13 Dec – 23 Dec 2019
2024 GE	General Election	4 Jul 2024	Pre-election	W27	24 May – 7 Jun 2024
			Campaign	W28	10 Jun – 3 Jul 2024
			Post-election	W29	5 Jul – 19 Jul 2024

Table 9: Yearly Analysis: British Election Study

Year	Wave	Field Dates	Event Context
2014	W3	19 Sep – 17 Oct 2014	Post-Scottish Independence Referendum (18 Sep)
2015	W6	8 May – 26 May 2015	Post-General Election (7 May)
2016	W9	24 Jun – 4 Jul 2016	Post-EU Referendum (23 Jun)
2017	W13	9 Jun – 23 Jun 2017	Post-General Election (8 Jun)
2018	W14	4 May – 21 May 2018	Post-Local Elections
2019	W19	13 Dec – 23 Dec 2019	Post-General Election (12 Dec)
2020	W20	3 Jun – 21 Jun 2020	COVID-19 Pandemic
2021	W21	7 May – 25 May 2021	Post-Local/Devolved Elections (6 May)
2022	W23	6 May – 26 May 2022	Post-Local Elections
2023	W25	5 May – 23 May 2023	Post-Local Elections
2024	W29	5 Jul – 19 Jul 2024	Post-General Election (4 Jul)
2025	W30	2 May – 22 May 2025	Post-Local Elections

Table 10: Indexed Time Trends in Within-Person Variation – UK Robustness Checks

Specification	**UK (BES)**
No controls (Model 1)	-7.935^{**}
	(0.945)
Election position only	-7.933^{**}
	(0.946)
Main specification (Model 5)	-5.893^{*}
	(1.177)
Without days control	-6.693^{**}
	(0.868)
Pre-election volatility only	-5.486^{*}
	(1.611)
Across-election volatility only	-6.455^{*}
	(1.287)
Quasi-yearly volatility	-2.291^{***}
	(0.359)

Note: No controls (Model 1) has no covariates; Election position only (R1) controls only for wave position. All other rows use the Model 5 controls (wave position, panel tenure as a factor, demographics, days until the next survey) unless stated. "Without days control" drops the days control. "Pre-election" uses only changes between the first two waves of each election cycle; "across-election" uses only changes spanning polling day; "Quasi-yearly" uses annual changes and its coefficient is per year, not per election cycle (it cannot use wave position as a control). Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

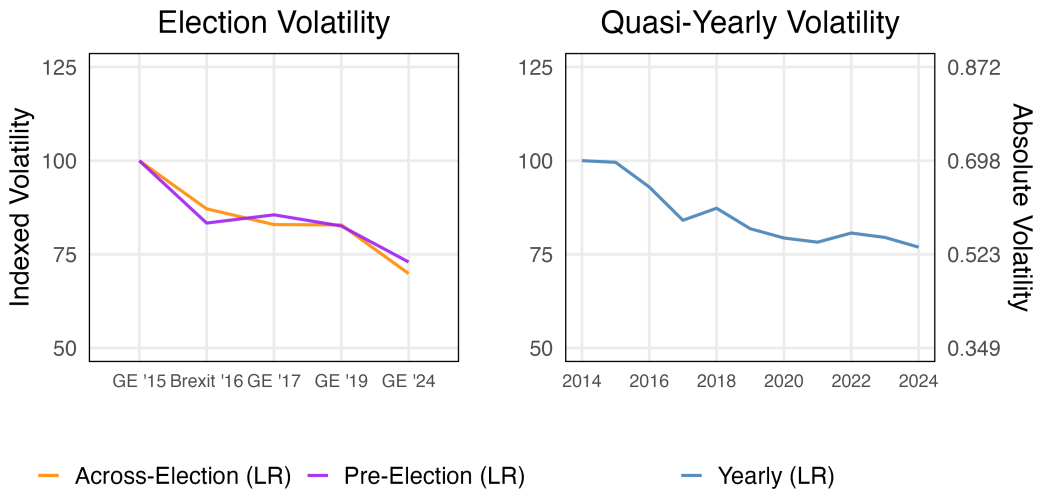


Figure 12: UK political volatility under alternative measurement windows. Left: pre-election and across-election volatility. Right: quasi-yearly volatility, with a linear control for days between surveys. All Model 5 controls; indexed to 100 in the first period.

Table 11: Within-election and quasi-yearly volatility in left–right self-placement (BES)

	Within-election		Quasi-yearly		
Cycle	$\overline{\lvert\Delta\text{LR}\rvert}$	Gap (days)	$\overline{\lvert\Delta\text{LR}\rvert}$	Gap (days)	Ratio
2015 GE	0.67	28	0.77	228	0.87
2016 EU ref.	0.60	33	0.70	352	0.86
2017 GE	0.59	24	0.67	337	0.88
2019 GE	0.56	20	0.59	336	0.95
2024 GE	0.47	19	0.56	364	0.83

Note: Mean absolute change in left–right self-placement (0–10) between consecutive surveys, for the 37,682 respondents observed in both samples. Gaps are medians. Quasi-yearly observations are assigned to the election cycle in which they fall; the first cycle covers only 2014–2015, so its quasi-yearly gap is shorter. Ratio is within-election divided by quasi-yearly volatility. Unweighted.

D Levels Analysis

Volatility identifies the stretching parameter d_P, but not the shifter k_P. Taking a period-averaged version of Equation 2:

$$\overline{R_P} = d_P \, \overline{q_P} + k_P \tag{23}$$

Since d_P and k_P are constants within P (Assumptions 2–3), adjusting each response and then averaging is equivalent to averaging and then adjusting. Recovering the latent mean requires taking a stance on k_P.

D.1 Average Life satisfaction

One way forward is to assume that rising dispersion in life satisfaction comes entirely from the top of the scale: $k_P \approx 0$. Rearranging Equation 23:

$$\widehat{\overline{q_P}} = \frac{\overline{R_P}}{\widehat{d}_P} \tag{24}$$

The latent mean is the nominal mean divided by the estimated stretching factor.

This "no shift" assumption is close to true in our income validation exercise. The V-LINE-adjusted mean tracks observed income closely (Figure 13).

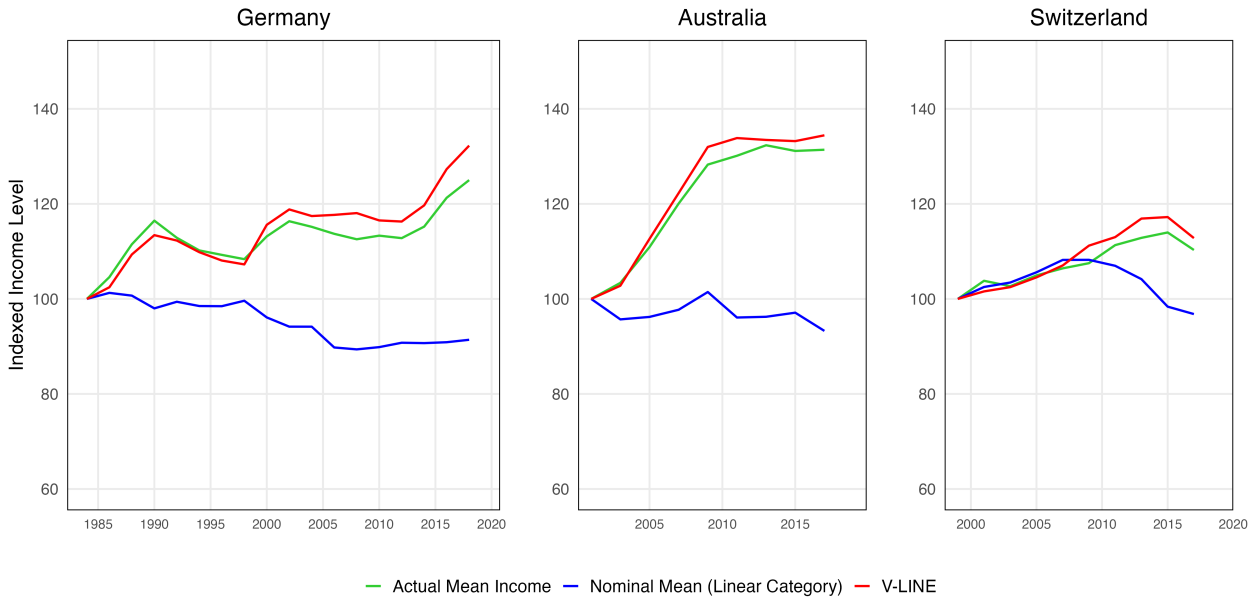


Figure 13: Validating the levels adjustment using household income: nominal mean of the linearised income category, V-LINE-adjusted mean, and observed mean household income. Germany, Australia and Switzerland.

Figure 14 applies the mean adjustment for our wellbeing data across each sample.

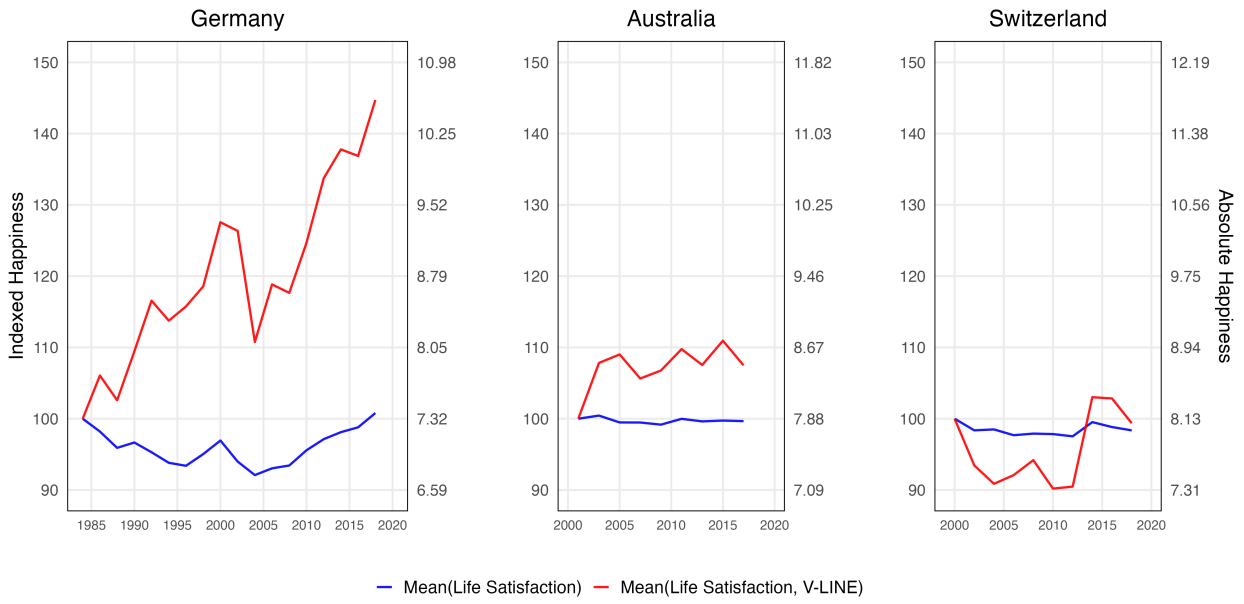


Figure 14: Mean life satisfaction: nominal versus V-LINE adjusted. Nominal means are flat; latent means are not.

We also treat COVID-19 as a natural experiment, comparing 2018–2019 with 2021–2022 in Germany and Switzerland (2020 is excluded due to survey timing). Nominal life satisfaction is flat in both countries. The adjusted mean falls by about 2.5 per cent in Germany and 6 per cent in Switzerland (Figure 15). Latent dispersion falls by a similar proportion (Figure 16).

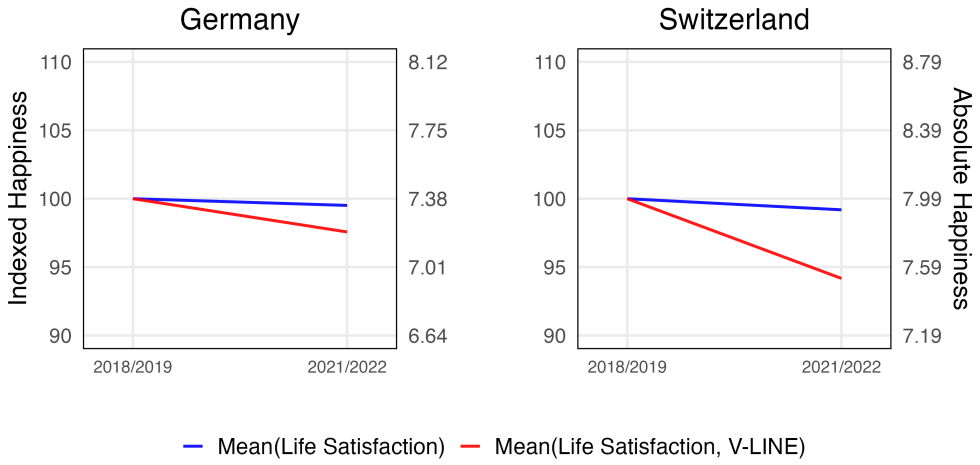


Figure 15: Life satisfaction during COVID-19: nominal versus V-LINE-adjusted means, 2018–2019 compared with 2021–2022. Nominal means are stable; latent means fall.

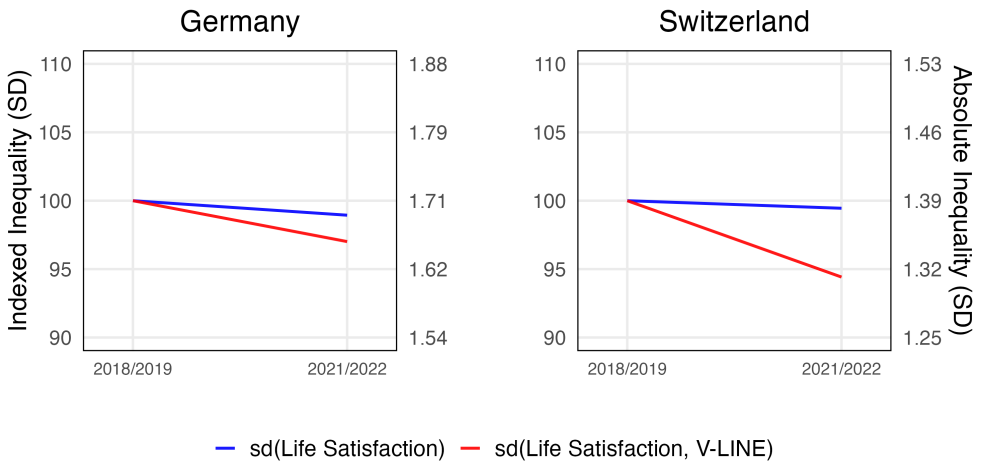


Figure 16: Life satisfaction during COVID-19: nominal versus V-LINE-adjusted dispersion, 2018–2019 compared with 2021–2022. Latent dispersion falls faster than nominal dispersion.

D.2 Average Political ideology

We anchor the scale at its midpoint. If a latent 5 is reported as 5 in every period, then $k_P = 5(1-d_P)$, and Equation 23 gives

$$\widehat{\overline{q_P}} = 5 + \frac{\overline{R_P} - 5}{\widehat{d}_P}. \tag{25}$$

Stretching therefore rescales only the distance between the nominal mean and 5, which is under 0.25 points in both the BES and the SHP (Table 3).

For the UK, response volatility in 2024 is about 73 per cent of its 2015 level, so by Proposition 2, $\widehat{d}_{2024} = |\Delta R_{2024}|/|\Delta R_{2015}| \approx 0.73$ (Figure 5a).

With a mean self-placement of 4.80 for the UK, Equation 25 gives

$$\widehat{\overline{q}}_{2024} = 5 + \frac{4.80 - 5}{0.73} \approx 4.73,$$

which is 0.07 points below (i.e., more left-wing) than the nominal mean. In Switzerland, where $\widehat{d}$ is close to 1, the adjustment is about 0.01 points. Under this anchor, stretching barely affects average ideology.

E Robustness to linearity violations

So far we have assumed a linear reporting function. Suppose instead that a monotonically increasing, but not necessarily linear, transformation $f(\cdot)$ of the reported value is affine in the underlying quantity:

$$f(R_{it}) = d_P\, q_{it} + k_P. \tag{26}$$

Our V-LINE estimator, which assumes linearity, effectively sets $f(R_{it}) = R_{it}$.

A non-linear f introduces bias through two separate channels: first, in the dispersion estimate itself, since R_{it} and $f(R_{it})$ differ; and second, in the estimated stretching parameter, since it is computed from $|\Delta R|$ rather than $|\Delta f(R)|$. These two channels need not point in the same direction, depending on the shape of f and the distribution of R_{it}. It is therefore not possible to sign the overall bias analytically without further assumptions on f. We instead assess its size and direction by simulation.

E.1 Simulating Non-Linearity

We simulate $f(\cdot)$ using three monotone shapes drawn from two parametric families. Each maps the 0–10 response scale onto a transformed 0–10 scale and fixes both endpoints. Since $d_P > 0$, gaps in $f(R)$ are proportional to gaps in the underlying quantity.

Power transformations.

$$f_{\text{power}}(R; p) = 10 \left(\frac{R}{10} \right)^p \tag{27}$$

1. Concave ($p < 1$): latent gaps between lower response categories are *larger* than those at the top.

2. Convex ($p > 1$): latent gaps between lower response categories are *smaller* than those at the top.

Inverse-S transformation. Latent gaps at the top *and* bottom of the scale are larger than those in the middle (Ng, 2008):

$$f_{\text{inv-S}}(R; \kappa) = \frac{R^2}{10\, s(R; \kappa)}, \tag{28}$$

where $s(R; \kappa)$ is a logistic S-curve centred at the scale midpoint and rescaled to the unit interval, with steepness κ. The transformation is mildly asymmetric, with larger gaps at the bottom of the scale than at the top.

We measure non-linearity with the cost function of Kaiser and Lepinteur (2025). Let $\tilde{l}_k = f(k)$ for $k = 0, \ldots, 10$, and let $\Delta\tilde{l}$ be the vector of gaps between adjacent transformed labels. Then

$$C = \sqrt{\frac{\text{Var}(\Delta\tilde{l})}{\max \text{Var}(\Delta\tilde{l})}}, \tag{29}$$

where $\max \mathrm{Var}(\tilde{\Delta l})$ is the variance when the entire range falls in a single gap. $C = 0$ under linearity and $C = 1$ for a single jump. For each of the three shapes, we numerically solve for the parameter value that achieves $C \in \{0.05, 0.10, 0.20\}$. Figure 17 plots the resulting transformations. We then apply each transformation to the reported values R_{it} and re-estimate our V-LINE adjustments using $f(R_{it})$ in place of R_{it}.

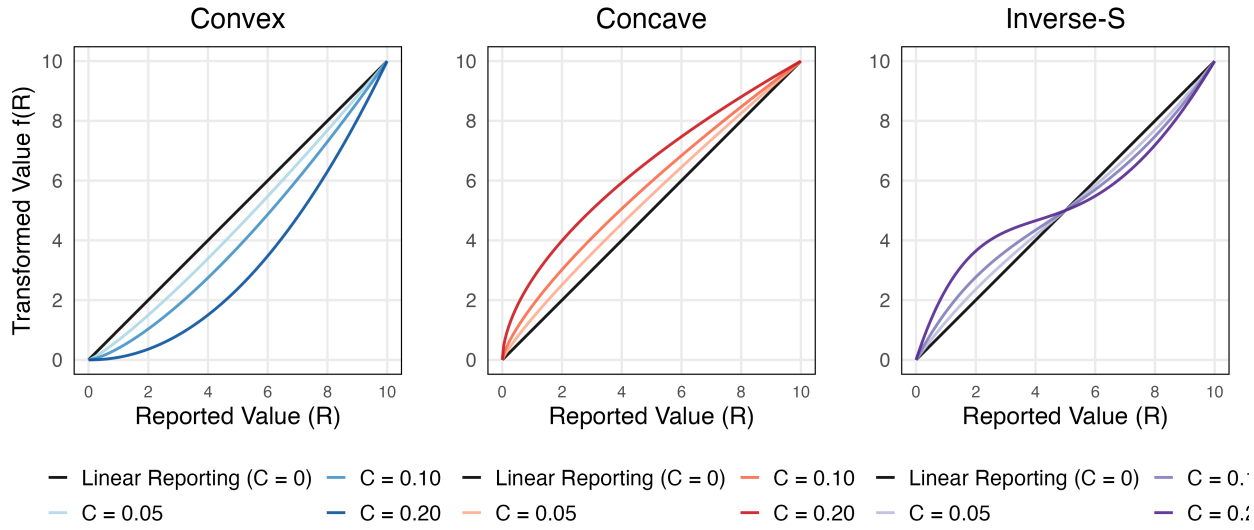
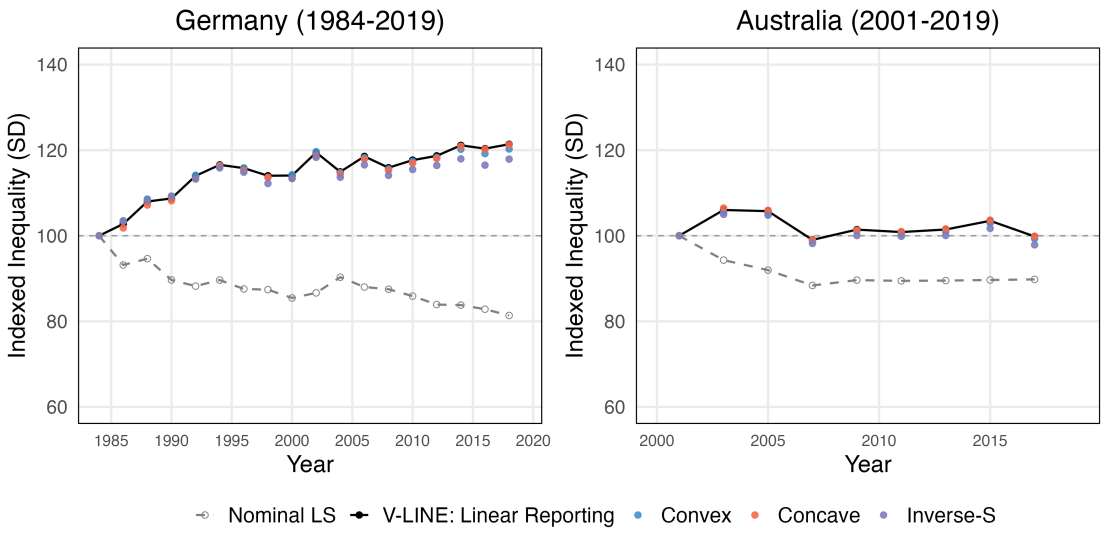
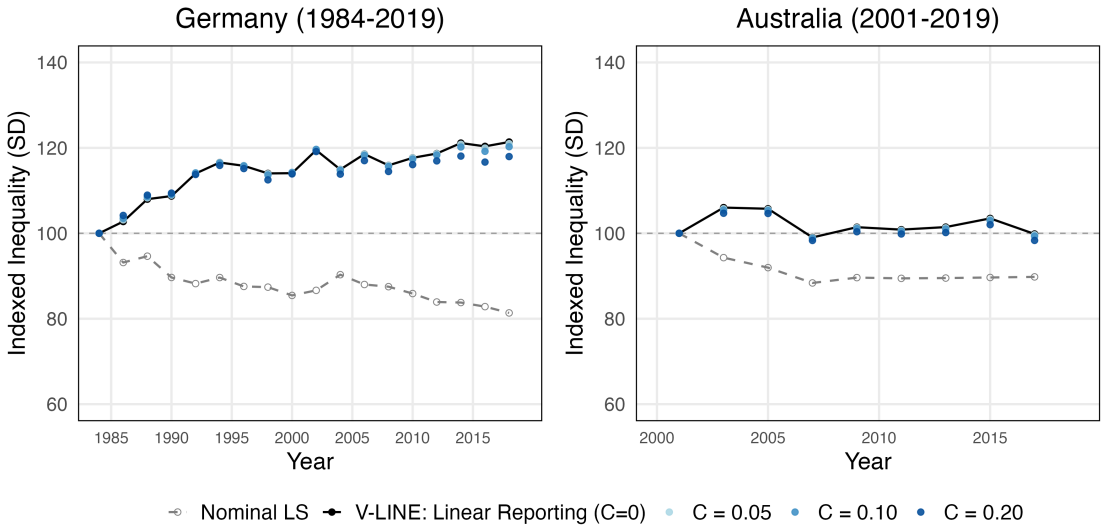


Figure 17: Simulated non-linear transformations of the 0–10 response scale. Each panel shows one transformation shape at the three target costs $C \in \{0.05, 0.10, 0.20\}$, with darker lines indicating greater non-linearity. The black line is the identity, corresponding to linear reporting ($C = 0$). A steeper segment indicates a larger latent gap between adjacent response categories.

Across all transformations with $C \leq 0.20$, a range that Kaiser and Lepinteur (2025) classify as plausible to conservatively plausible, the V-LINE continues to recover the rise in inequality in Germany (for wellbeing inequality) and in the UK (for ideological polarisation). For Germany and Australia, this exercise enters panel tenure linearly rather than as a set of dummies, to reduce computation across the many models estimated; the results therefore differ mildly from those in the main text.

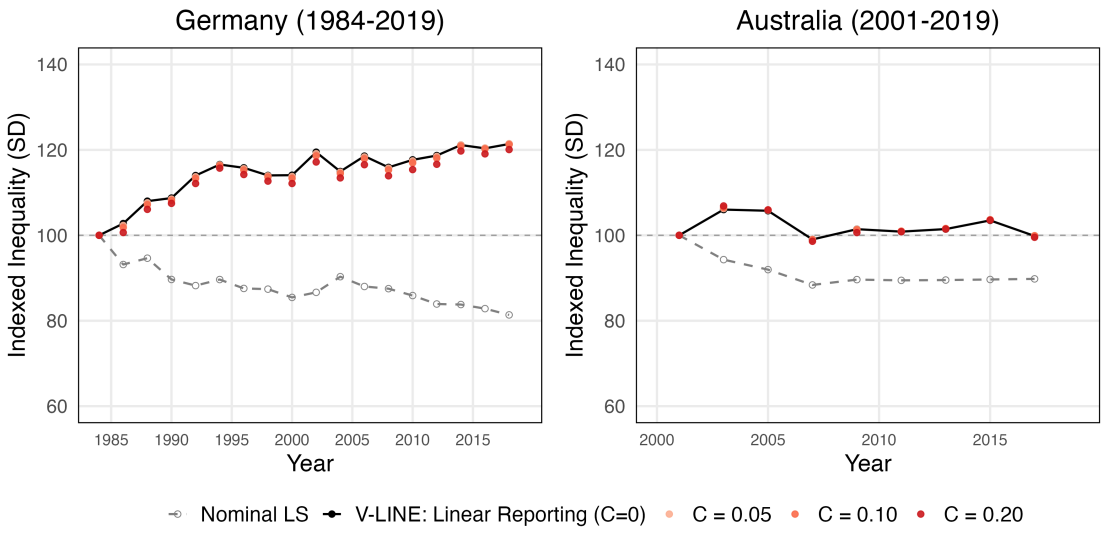


(a) All reporting functions, $C = 0.10$

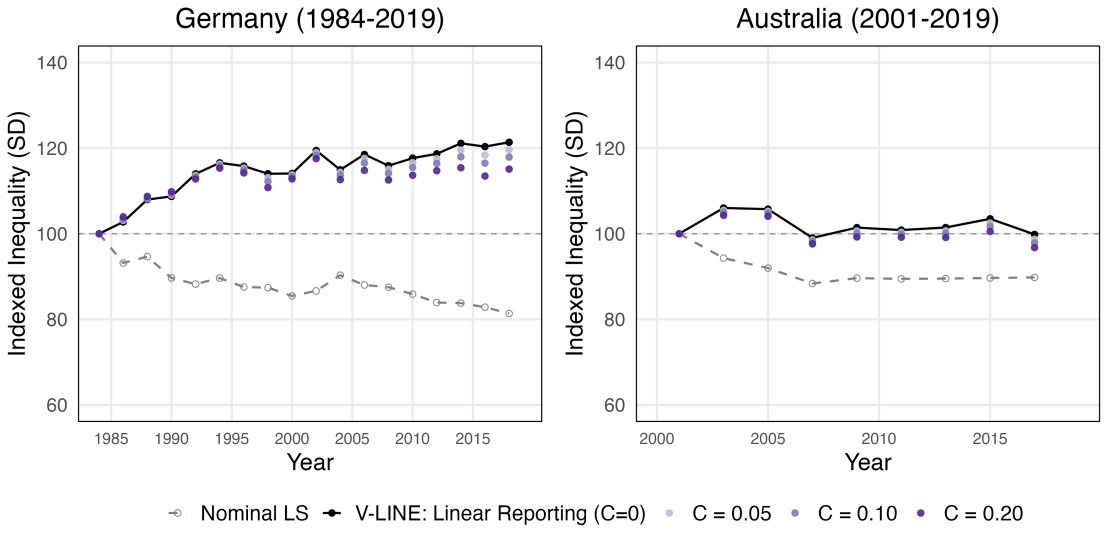


(b) Convex

Figure 18: V-LINE latent inequality under non-linear reporting functions, Germany (1984–2019) and Australia (2001–2019). The dashed grey line is the nominal SD of life satisfaction. The solid black line is the baseline V-LINE estimate, which assumes linear reporting ($C = 0$). In panel (a), coloured points re-estimate V-LINE under convex, concave and inverse-S reporting functions, each at $C = 0.10$. In panels (b)–(d), points re-estimate V-LINE under a single reporting function at $C \in \{0.05, 0.10, 0.20\}$, and darker shades denote more non-linearity. All series are indexed to the first period (= 100).

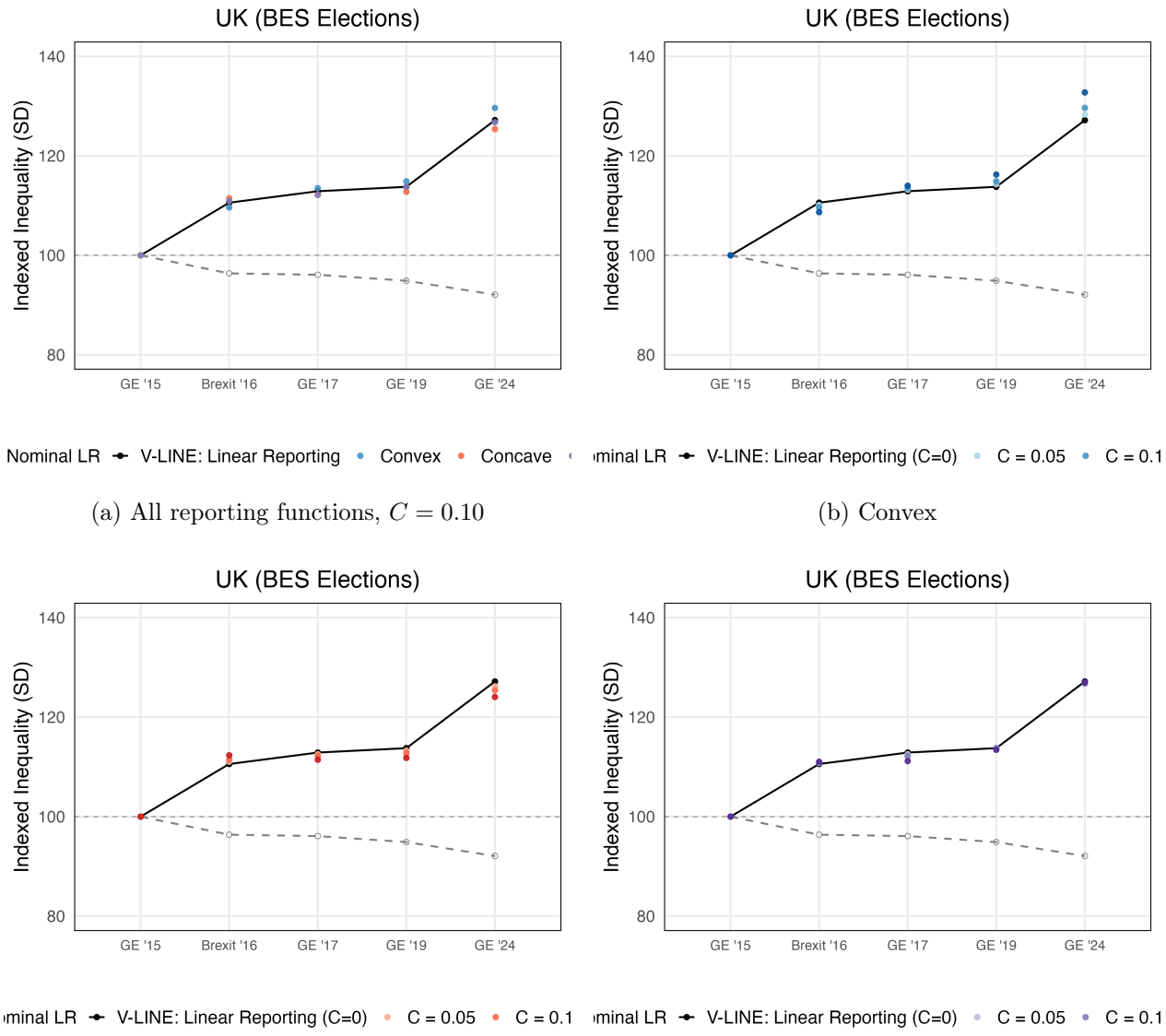


(c) Concave



(d) Inverse-S

Figure 18: V-LINE latent inequality under non-linear reporting functions, Germany and Australia (continued).



(a) All reporting functions, $C = 0.10$

(b) Convex

(c) Concave

(d) Inverse-S

Figure 19: V-LINE mass polarisation under non-linear reporting functions, Britain (BES). The dashed grey line is the nominal SD of left-right self-placement. The solid black line is the baseline V-LINE estimate, which assumes linear reporting ($C = 0$). In panel (a), coloured points re-estimate V-LINE under convex, concave and inverse-S reporting functions, each at $C = 0.10$. In panels (b)–(d), points re-estimate V-LINE under a single reporting function at $C \in \{0.05, 0.10, 0.20\}$, and darker shades denote more non-linearity. Periods are the five BES election cycles. All series are indexed to the first cycle (= 100).

F Robustness to smooth rescaling trends

Our identification strategy also relies on Assumption 3: that the reporting function remains constant within each 2-year period. This section derives an upper bound on the resulting bias, and shows that even in the worst case it is small ($< 10\%$).

F.1 Deriving a Term for Within-Period Rescaling Bias

Suppose within-period rescaling does occur, so that reporting parameters differ between years t and $t+1$ within a period. Our method identifies period-average parameters:

$$d_P = \frac{1}{2}(d_t + d_{t+1}), \qquad k_P = \frac{1}{2}(k_t + k_{t+1}), \qquad q_{iP} = \frac{1}{2}(q_{it} + q_{i,t+1}). \tag{30}$$

These imply $d_{t+1} = 2d_P - d_t$, and analogously for k_t and q_{it}.

Substituting a linear reporting function (Assumption 1) into the observed change $R_{i,t+1} - R_{it}$, then substituting reporting parameters using their period averages and collecting terms, yields our key decomposition:

$$\boxed{\underbrace{R_{i,t+1} - R_{it}}_{A} \;=\; \underbrace{d_P(q_{i,t+1} - q_{it})}_{B} \;+\; \underbrace{q_{iP}(d_{t+1} - d_t) + (k_{t+1} - k_t)}_{C}} \tag{31}$$

A is what we observe. B is the true, latent change, scaled by the period-average stretching parameter – what we ultimately want to identify. C is the bias term arising from within-period drift. Taking absolute values:

$$|A| = |B + C|. \tag{32}$$

If reporting parameters are constant within the period, $C = 0$, and Equation (32) gives $|A| = |B| = d_P|\Delta q_{iP}|$, recovering Proposition 2. If they are not constant, C is the part of an individual's observed change driven by instability in the reporting function rather than real movement in q: for someone whose latent position did not change within the period ($B = 0$), $|A| = |C|$ – their entire observed change is attributable to bias.

F.2 How substantial is the bias from within-period rescaling?

We bound how far $|A|$ can depart from $|B|$. Since $A = B + C$, the triangle inequality gives

$$\big|\,|A| - |B|\,\big| \;\le\; |C|. \tag{33}$$

Averaging across individuals – using $\overline{|B|} = d_P\overline{|\Delta q_{iP}|}$ (Assumption 2) – gives

$$\Big|\,\overline{|A|} - \overline{|B|}\,\Big| \;\le\; \overline{|C|}. \tag{34}$$

This holds regardless of how the bias term correlates with Δq_{iP} across individuals.

For happiness, Prati and Senik (2026) estimate annual linear rescaling of about 1 percentage point per year, giving $\overline{|C|} \approx 0.075$ from a baseline life satisfaction of 7.5. Mean within-person volatility (in the first period) is $\overline{|\Delta q_{iP}|} \approx 0.9$–$1.6$ (Table 2). Substituting into Equation (34), the distortion in $\overline{|\Delta R_{iP}|}$ is roughly 5–10% of the signal. (We are not aware of an equivalent estimate of annual within-period rescaling for political ideology.)

The true distortion is likely smaller still (as we found empirically in Section 4). This is because within-period drift shifts everyone's reports in the same direction, whereas changes in latent states run in both directions. This means that an inflated change for some respondents is offset by deflated changes for others.

G Robustness to violations of homoskedasticity

G.1 Sensitivity Analysis

We suggest hidden divergences have occurred for Germany (life satisfaction) and the UK (political ideology). How much would Assumption 4 have to fail before this conclusion reversed? Intuitively: if latent volatility has been falling, we will over-estimate latent inequality. However, we still get a hidden divergence (rising latent inequality) whilst accommodating fairly large falls in latent volatility – even intuitively large violations of Assumption 4 would not change our results.

Equation (13) gives the true stretching ratio without invoking Assumption 4:

$$\frac{d_P}{d_{P=0}} = \frac{\overline{|\Delta R_P|}}{\overline{|\Delta R_{P=0}|}} \div \frac{\overline{|\Delta q_P|}}{\overline{|\Delta q_{P=0}|}} \tag{35}$$

Assumption 4 sets the second ratio to 1, giving $\widehat{d}_P$ (Proposition 2). Suppose instead that true volatility genuinely fell by a fraction x, so that $\overline{|\Delta q_P|}/\overline{|\Delta q_{P=0}|} = 1-x$. The true stretching parameter is then $\widehat{d}_P/(1-x)$, and by Proposition 3 the true adjusted inequality index is:

$$\left(\frac{SD(q_P)}{SD(q_{P=0})}\right) = (1-x)\cdot\widehat{R}, \qquad \widehat{R} \equiv \frac{\widehat{SD(q_P)}}{\widehat{SD(q_{P=0})}} \tag{36}$$

where $\widehat{R}$ is our reported adjusted ratio. Setting this to 1 gives the threshold decline in true volatility, x^*, needed to eliminate the divergence entirely:

$$x^* = \frac{\widehat{R}-1}{\widehat{R}} \tag{37}$$

For Germany, $\widehat{R} = 1.168$ in the final period (2019), giving $x^* = 0.144$: even if latent happiness volatility had fallen by up to 14%, there would still be a hidden divergence (rising latent inequality). For the UK, $\widehat{R} = 1.272$ for the final election year (2024), giving $x^* = 0.214$ – a substantially larger margin. This 'allowable decline' is given on the far-right column in Table 12.

We can also express the allowable decline in natural units. For Germany: the allowable decline is roughly 3 times the size of the estimated sex gap in reported life-satisfaction volatility. Concretely, imagine the entire German sample was female, and then became entirely male. We estimate that the volatility index would fall by 4.6. This is calculated as the ratio between the raw coefficient (the relationship between sex and LS volatility) and the starting value of the volatility of LS. The allowable decline is 14.4 index points: just over three times that amount. So even if we had omitted a confounding variable as large as the whole sample switching sex, three times over, there would still be a hidden divergence in Germany. The same logic applies to migration background (2.7 times over). We found that older people have (surprisingly) higher happiness volatility, and that those who have been in the survey longer have lower volatility. The allowable range for a hidden divergence is a population being 110 years younger, or having completed an extra 34 waves of surveys.

These results are summarised in Table 12. The allowable range, when expressed in natural units – shown in the second column from the right – is similarly large for the UK and political ideology.

Table 12: Sensitivity of the Hidden Divergence to Omitted Confounders

Country	Variable	Raw Coefficient	Index-Point Effect	Allowable Multiples	Allowable Decline in Latent Volatility
Germany	Panel tenure	−0.0060	−0.42	34.16	14.40%
Germany	Age	+0.0019	+0.13	109.65	
Germany	Sex (Male→Female)	+0.0649	+4.55	3.16	
Germany	Migrant (Native→Migrant)	+0.0768	+5.39	2.67	
UK (BES)	Panel tenure (waves)	−0.0157	−2.57	8.30	21.36%
UK (BES)	Age	+0.0022	+0.36	59.61	
UK (BES)	Gender (Male→Female)	+0.0603	+9.88	2.16	

Note: Raw coefficient is in the outcome's own units (life-satisfaction or left–right points). Index-point effect expresses it as a percentage of initial volatility. Allowable decline (x^*) is the fractional fall in true volatility that would have to occur, undetected, to eliminate the hidden divergence. Allowable multiples is the allowable decline divided by the index-point effect. For panel tenure, it is the number of extra waves the whole sample would need to participate in. For age, it is the number of years by which the whole sample would need to be younger. For sex, gender and migrant status, it is the number of complete sample-wide category reversals required (e.g. every respondent switching from male to female).

G.2 Violations of Homoskedasticity in Income Exercise

We can also assess Assumption 4 empirically in the income exercise, where the latent quantity is directly observable. Table 13 and Figure 20 report results from Section 4 with and without Assumption 4.

Table 13: Mean Absolute Error: Validating the Linearised Income Category Adjustment

Country	Nominal	V-LINE (No Assumption 4)	V-LINE (With Assumption 4)
Germany	20.93	0.53	7.69
Australia	26.17	0.56	15.12
Switzerland	6.75	1.22	2.91

Note: Mean absolute error (indexed percentage points) between each estimate of income inequality (SD, indexed to 100 in the first period) and the observed SD of actual household income. Nominal: SD of the linear income category variable, unadjusted. V-LINE (No Assumption 4): full V-LINE adjustment, allowing latent income volatility to vary over time. V-LINE (With Assumption 4): V-LINE adjustment under the simplifying assumption that latent income volatility is constant (homoskedastic) over time.

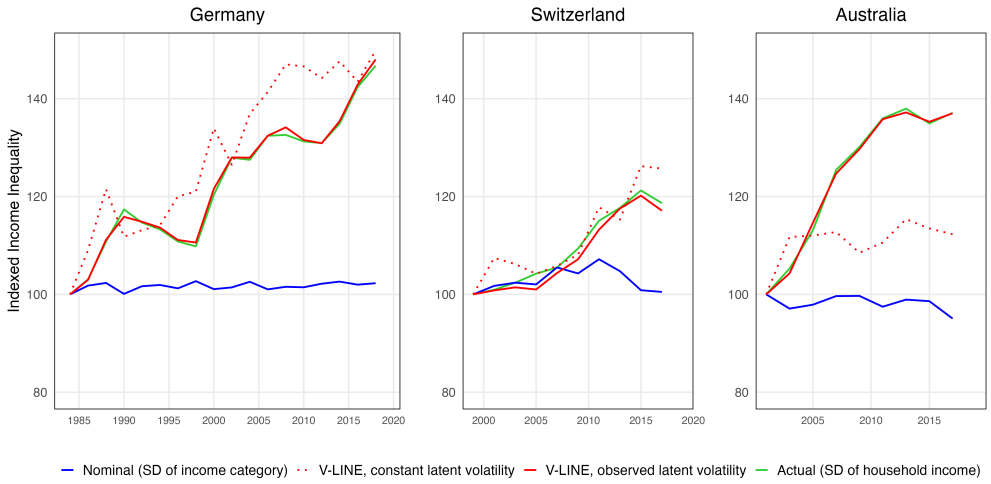


Figure 20: Income inequality estimates with and without the homoskedasticity assumption. Dotted red: V-LINE estimate assuming latent income volatility is constant over time (Assumption 4). Solid red: full V-LINE adjustment, allowing latent income volatility to vary. Solid green: observed SD of actual household income. See Table 13 for mean absolute error comparisons across all three series.