Posterior Bias of the Drift Estimator in a Hierarchical Ornstein–Uhlenbeck Model for Adverse Life Events

IRINA NASKINOVA[1], MARIYAN MILEV[1,2,*], NIKOLAY NETOV[2,*], MIKHAIL KOLEV[1,3,4,*], HRISTO KALINOV[5], GABRIELA VASILEVA[6], KRISTIAN MILEV[4], ANGELINA ANGELOVA[7]

[1]Department of Mathematics, University of Architecture,
Civil Engineering and Geodesy,
1 Hristo Smirnenski Blvd., 1164 Sofia,
BULGARIA

[2]Department of Statistics and Econometrics,
Faculty of Economics and Business Administration,
Sofia University "St. Kliment Ohridski",
125 Tsarigradsko Shosse Blvd., bl. 3, 1113 Sofia,
BULGARIA

[3]Department of Applied Computer Science and Mathematical Modelling,
Faculty of Mathematics and Computer Science,
University of Warmia and Mazury in Olsztyn,
POLAND

[4]Faculty of Mathematics and Informatics,
Paisii Hilendarski University of Plovdiv,
24 Tsar Assen Str., 4000 Plovdiv,
BULGARIA

[5]Faculty of Mathematics and Informatics,
Sofia University "St. Kliment Ohridski",
5 James Bourchier Blvd., 1164 Sofia,
BULGARIA

[6]Faculty of Social Business and Computer Sciences,
Varna Free University,
Chayka Resort, 9007 Varna,
BULGARIA

[7]Faculty of Pedagogy,
Paisii Hilendarski University of Plovdiv,
24 Tsar Assen Str., 4000 Plovdiv,
BULGARIA

**Corresponding Authors*

Abstract: - A hierarchical Ornstein–Uhlenbeck stochastic differential equation with event-conditional drift is fitted to monthly text-derived affect trajectories from a self-disclosure-rich Reddit subreddit (2023–2024; 254,153 users, 324,548 user-months), with the death of a close family member as the adverse life event. Event dates are recovered from free text by three regex classes of increasing specificity — coarse, explicit and recent — forming a within-corpus precision ladder. The ladder gives the first real-data test of the posterior-bias bound of the companion paper, [6], which predicts that the recovered drift magnitude grows as event-date precision tightens. The prediction is confirmed on the coarse-to-explicit step (mean drift −0.109 to −0.157 standardised VADER units; adjusted p from 0.041 to 0.008), whereas the explicit-to-recent step is underpowered because the

recency filter cuts the subject pool threefold. We document this precision–power tradeoff and propose the regex ladder as a reusable proxy for event-date precision.→ **-0.109-0.157*d*-0.117-0.162p_{BH} → *n*=352*n*=125**

Key-Words: - Ornstein–Uhlenbeck process; Stochastic differential equations; Particle filter; Hierarchical Bayesian inference; Posterior bias; Adverse life events; Family loss; Text-derived event detection; Latent affect.

MSC 2020—60H10; 62F15; 62M05; 62P25.

Received: May 9, 2026. Revised: June 11, 2026. Accepted: July 15, 2026. Published: September 9, 2026.

1 Introduction

Estimating the dynamics of a latent psychological state around discrete life events is a central problem in computational social science and mathematical psychology, [1], [2], [3]. Continuous-time hierarchical formulations of latent psychological dynamics have been developed extensively for fixed-grid and irregularly spaced panel data, [4], [5]; the SDE we adopt here generalises that family by adding a discrete event-conditional drift component. When the latent state is affect—a one-dimensional quantity that drifts under daily life and can be jolted by events such as a death in the family—a natural mathematical formulation is a mean-reverting Ornstein–Uhlenbeck stochastic differential equation (SDE) whose equilibrium shifts by an event-conditional amount β_e during a window around the event.

A companion paper [6] develops this hierarchical OU SDE framework, establishes three theoretical results (existence-uniqueness, identifiability, and a posterior-bias bound), and demonstrates parameter recovery on a synthetic-data suite. The bias bound there—Proposition 3 of the companion—predicts that when the true event date t_e^i is observed only to a coarser precision Δ, the posterior mean of the drift β_e acquires a bias of order Δ/w, where w is the event-window length. The bound was verified synthetically; the present paper provides the corresponding *empirical* verification.

The empirical setting is a self-disclosure-rich Reddit subreddit—a community whose posts disproportionately contain narratives of adverse life events. The specific subreddit used here (named in Section 5 for reproducibility) skews towards family-loss narratives, which we use as the test case for the precision-ladder analysis. We extract 24 months (January 2023 to December 2024) of posts and detect family-loss mentions by three regex classes of increasing temporal specificity:

Coarse. A narrow kinship inventory (*mom, mother, dad, father, brother, sister, husband, wife, grandparent*) co-occurring with a death verb (*died, passed away, passed*), *or* the bare token *funeral* on its own. This matches the inventory used by the companion paper.

Explicit. a broader kinship inventory (*son, daughter, partner, fiancé, fiancée, boyfriend, girlfriend, uncle, aunt, cousin, child*, together with the Coarse inventory) co-occurring with an extended death-verb inventory (*is gone, has passed, just died, just passed*) or the explicit constructions *lost my X* and *funeral for/of X*. Tightens positive-class precision (the bare *funeral* pattern is no longer accepted) at the cost of some recall.

Recent. The Explicit class intersected with a recency marker (*today, yesterday, last week, this week, a few days ago, n days ago*). The recency marker pins the event to within $\Delta \approx 7$ days of the post, the tightest level we can recover from text alone.

The three precision classes give us a within-corpus precision ladder without changing the inference pipeline. Proposition 3 of the companion paper predicts that as we step from *coarse* to *recent*, the recovered population mean drift should move toward its asymptotic exact-date value: smaller bias for tighter Δ. We test this prediction directly and find it confirmed on the Coarse → Explicit step but inconclusive on the Explicit → Recent step due to a power loss that the bound's asymptotic statement does not address.

The contribution of this paper is to (i) supply the first real-data companion to the previously synthetic theoretical result, (ii) propose the regex-specificity ladder as a reusable proxy for event-date precision in any text-derived event-time study, and (iii) document the practical precision–power tradeoff that arises when one tightens event-detection regexes on a finite corpus—a constraint not visible in the asymptotic statement of Proposition 3.

The paper proceeds as follows. Section 2 states the hierarchical OU SDE with event-conditional drift. Section 3 restates the three theoretical results of the companion paper that this work tests

empirically. Section 4 describes the bootstrap particle filter. Section 5 documents the corpus, names the specific subreddit, and defines the three regex classes. Section 6 reports the real-data drift estimates at each precision level and the bootstrap sampling distribution of the population mean. Section 7 discusses limitations, and Section 8 concludes.

2 Hierarchical Ornstein–Uhlenbeck Model

For user $i \in \{1, \dots, N\}$ at continuous calendar time $t \in [0, T]$, the latent affect state $X_{i,t} \in \mathbb{R}$ follows the event-conditional Ornstein–Uhlenbeck SDE

$$dX_{i,t} = -\theta_i (X_{i,t} - \mu_i(e_{i,t}))dt + \sigma_i dW_{i,t}, X_{i,0} \sim \pi_0, \quad (1)$$

with mutually independent Brownian motions $\{W_{i,\cdot}\}_{i=1}^{N}$, subject random effects

$$\theta_i \sim N(\theta_0, \tau_\theta^2), \sigma_i \sim LogNormal(log\sigma_0, \tau_\sigma^2), \mu_{i,0} \sim N \quad (2)$$

and event-conditional equilibrium

$$\mu_i(e_{i,t}) = \mu_{i,0} + \beta_e 1\{0 \le t - t_e^i < w\}, \quad (3)$$

where $e \in E$ is the event type, t_e^i is the date of the event e for subject i, and w is a fixed window length. The empirical chapter of this paper restricts the inventory to $E =$ family-loss} (the death of a close family member); the general framework of (1)–(3) accommodates any finite inventory of discrete adverse life events. Observations are monthly text-derived affect scores

$$Y_{i,k} = X_{i,t_{i,k}} + \varepsilon_{i,k}, \varepsilon_{i,k} \sim N(0, \sigma_y^2), \quad (4)$$

recovered at irregularly spaced calendar times $t_{i,1} < \cdots < t_{i,K_i}$. The population-level hyperparameters $\Theta = (\theta_0, \tau_\theta^2, \sigma_0, \tau_\sigma^2, \mu_0, \tau_\mu^2, \beta_e, \sigma_y^2)$ are the inferential target. Throughout this paper, we use a calendar-month time grid and fix the post-event window w=60 days (two calendar months); the choice w=60 matches the companion paper. The SDE (1) is a linear Ornstein–Uhlenbeck process for which strong existence and pathwise uniqueness follow from classical results, [7], under Assumption 1 below. The observation model (4) treats text-derived affect as a noisy proxy of the latent state, in line with measurement-error formulations of multi-level autoregressive models, [8].

3 Theoretical Results

The three propositions below are quoted verbatim from the companion paper, [6]. They are stated here only to the extent that this paper tests them, with full proofs deferred to the companion. The first two are standard; the third—which this paper exists to test empirically—is recalled with its constant.

Assumption 1 (Regularity). *$\theta_i > 0$ and $\sigma_i > 0$ almost surely; $\mu_i(\cdot)$ is bounded; $X_{i,0}$ has a finite second moment; the event-window indicator $1\{0 \le t - t_e^i < w\}$ is right-continuous with left limits.*

Assumption 2 (Quasi-equilibrium). *Outside event windows, the subject-level equilibrium $\mu_{i,0}$ is constant in t over the observation horizon.*

Proposition 1 (Existence and uniqueness). *Under Assumption 1, the SDE (1) admits a unique strong solution $X_{i,\cdot} \in L^2([0,T])$ for almost every realisation of $(\theta_i, \sigma_i, \mu_i(\cdot))$.*

Proposition 2 (Identifiability under sparse panel observation). *Suppose Assumption 2 holds, $N \to \infty$, and each subject i has $K_i \ge 2$ observations, of which at least one falls within the event window of e for a positive fraction of subjects. Then the population-level hyperparameters Θ are jointly identifiable from the marginal likelihood of $\{Y_{i,k}\}$.*

Proposition 3 (Posterior bias bound). *Suppose the true event date is t_e^i but only $\tilde{t}_e^i$ with $|\tilde{t}_e^i - t_e^i| \le \Delta$ is observed. Let $\hat{\beta}_e^{\Delta}$ denote the posterior mean of β_e under (1)–(4) using $\tilde{t}_e^i$ in place of t_e^i, and let $\hat{\beta}_e^0$ denote the corresponding posterior mean with exact event dates. Then $|\hat{\beta}_e^{\Delta} - \hat{\beta}_e^0| \le C_{\theta,w} \beta_e \frac{\Delta}{w} + O(\Delta^2/w^2)$, with $C_{\theta,w} = 2(1 - e^{-\theta_0 w})/(\theta_0 w)$. For $\theta_0 = 0.05 \text{day}^{-1}$ and w=60 days the constant is $C_{\theta,w} \approx 0.633$, giving an envelope $|\hat{\beta}_e^{\Delta} - \hat{\beta}_e^0| \lesssim 0.011|\beta_e|$ at Δ=1 day and $\lesssim 0.32|\beta_e|$ at Δ=30 days.*

The empirical test of Proposition 3 that this paper provides is the following: the three regex classes in Section 5 correspond, by construction, to three different precision levels $\Delta_{\text{recent}} \lesssim 7$ days, $\Delta_{\text{explicit}} \lesssim 30$ days, and Δ_{coarse} uncontrolled (up to the calendar-month resolution and beyond). If the bound holds, the population mean of $\hat{\beta}_e$ at the *recent* class should be closer to its exact-date target than the corresponding mean at the *coarse* class, with the *explicit* class in between. This monotonicity is the testable consequence we report in Section 6.

4 Inference

The companion paper [6] fits the SDE via a bootstrap particle filter with a Liu–West shrinkage kernel, [9]. Sequential Monte Carlo for general nonlinear state-space models is reviewed comprehensively in [10], [11]; the particle-MCMC alternative of [12] delivers joint state-and-parameter inference with stronger asymptotic guarantees at

higher computational cost. In the linear-Gaussian special case β_e=0, the optimal filter reduces to the Kalman filter, [13], which the companion paper uses as a cross-validation oracle in synthetic experiments. The particle-set state is

$$\xi_i^{(p)}=(X_{i,t}^{(p)},\theta_i^{(p)},\sigma_i^{(p)},\mu_{i,0}^{(p)}),p=1,\dots,P, \quad (5)$$

with P=5000 particles per subject. Static hyperparameters are jittered with kernel scale h=0.05, and resampling is triggered when the effective sample size falls below P/2. The Gaussian special-case Kalman filter is used as a cross-validation oracle in the synthetic study of the companion paper; in the present paper, to keep the empirical statement model-free and reproducible, we summarise β_e at each precision level by the sample mean of the per-user pre-baseline affect difference, a one-sample t-confidence interval, and the bootstrap sampling distribution of that mean from 4,000 resamples. The full hierarchical filter is not refitted to the real corpus; the empirical statement we test is the direction of Proposition 3, a directional corollary of Proposition 3 that does not require refitting the full filter, and we justify it as follows. Under the linear-Gaussian model (1)–(4) with the event-conditional equilibrium (3), the latent equilibrium during the post-event window shifts by β_e; under Assumption 2 the baseline mean μ_i^{base} is an unbiased estimate of the subject's quiescent equilibrium, so the per-user contrast $\mu_i^{\text{pre}} - \mu_i^{\text{base}}$ is a method-of-moments estimator of $\kappa(\theta,w)\cdot\beta_e$, where the attenuation factor κ(θ,w), which lies in (0,1], depends only on the mean-reversion rate θ and the window length w and is common to all three precision classes. The posterior mean of β_e and this contrast are therefore monotone-increasing functions of the same per-user sufficient statistic, so their sign and their ordering across the precision ladder coincide. Date-precision error Δ attenuates the realised window shift identically in both estimators; hence the qualitative prediction of Proposition 3 — that the recovered magnitude grows as Δ shrinks — transfers to the difference of means, even though the bound's constant $C_{\theta,w}$ is not recoverable from it. We accordingly test the directional corollary rather than the bound's constant, and we treat this as a deliberate scope restriction: a full hierarchical Bayesian refit on the real corpus, which would test the bound quantitatively, is left to future work.

We stress that the full hierarchical particle filter is not refit on the Reddit corpus. In line with the directional corollary above, each precision class is summarised only through the per-user pre-event-minus-baseline affect contrast μpre − μbase, the method-of-moments sufficient statistic for κ(θ,w)·βe; the empirical claim is therefore a validation of this sufficient statistic's behaviour under event-date coarsening, not an evaluation of the full posterior distribution of βe.

For the present paper, the population mean drift β_e in each precision class is summarised in two equivalent ways:

- A frequentist confidence interval: per-user pre-event mean affect μ_i^{pre} on the window [−2,0] months relative to the detected event-month, minus a baseline mean μ_i^{base} on [−12,−7]. Per-event-class population mean and 95% CI from a one-sample t-test on μ_i^{pre}–μ_i^{base}.
- A bootstrap sampling distribution: 4,000 resamples of the per-user difference yield the bootstrap density of the population-mean estimator, reported in Section 6.3. We note that this is the frequentist sampling distribution of the mean under resampling, not a Bayesian posterior in the sense of Proposition 3.

5 Data

5.1 Corpus

We use the Pushshift/Arctic Shift Reddit archive, [14], [15], for the 24-month period 2023-01 through 2024-12. The single subreddit used in this paper is r/offmychest, chosen because it disproportionately contains family-loss narratives—a venue that prior computational-affect work has identified as a representative locus of unfiltered emotional self-disclosure on Reddit, [16]; the alternative subreddit r/jobs (as the companion paper documents) yields null results across all five event types in its inventory and is unsuitable for the present test. The raw archive comprises the subreddit's complete public post stream over the 24 months; we apply the following minimal pre-processing:

- Drop posts whose author is [deleted], [removed], or empty.
- Drop posts with combined title + body length below 30 characters.

Truncate the text used for affect scoring to the first 4,000 characters of (title + body).

Author identifiers are MD5-hashed at ingestion; only the first 16 hex characters are retained. The hashes are non-reversible and are used only to group the subject's monthly posts.

5.2 Affect Scoring

We use VADER [17] as the affect scorer. VADER's compound score is a [−1, +1]-bounded composite of valence-lexicon matches and intensifiers. For each post, we compute the compound score on the truncated text; for each user-month, we record the mean and standard deviation of the per-post compounds. Lexicon-based affect scoring has a long methodological history; alongside VADER, the LIWC dictionary, [18], is the de-facto reference in psychology and produces broadly correlated category counts on Reddit-like corpora. The use of VADER (rather than a transformer-based scorer) is deliberate: VADER is fully deterministic, lexicon-transparent, and reproducible; the trade-off is that it does not capture sentence-level semantic context and may mislabel sarcasm or quoted speech. Sensitivity to scorer choice is discussed in Section 7.

5.3 Family-Loss Event Detection and Three Precision Classes

The three regex classes are coded as Python re patterns and applied to the truncated text of each post. The kinship term inventory is

mom, mum, mother, dad, father, brother, sister, husband, wife, son, daughter, grandma, grandpa, grandmother, grandfather, nan, nana, papa, uncle, aunt, cousin, partner, fiancé, fiancée, boyfriend, girlfriend, spouse, baby, child, kid,

the death-verb inventory is *died, passed away, passed, is gone, has passed, just died, just passed*, and the recency inventory is *today, yesterday, last night, this morning, a few hours ago, a few days ago, n days ago (n ∈ two...seven), last week, this week, a week ago*.

Coarse. matches the inventory of the companion paper: kinship ∩ {died, passed away, passed} ∪ {funeral}, with kinship restricted to mom/mother/dad/father/brother/sister/husband/ wife/grandparent.

Explicit. matches the full kin inventory above ∩ the full death-verb inventory above, or the patterns *lost my X*, *lost a X*, *funeral for/of (my/her/his) X*.

Recent. is *Explicit* intersected with at least one match in the recency inventory within the same post.

The intent is that, under the assumption that recency markers in free-text are at most ~7 days off the actual event, $\Delta_{recent} \leq 7$, whereas Δ_{coarse} includes all uncontrolled lag between the death and the user's choosing to post about it. We use a regex-only pipeline rather than a full temporal-expression normaliser such as HeidelTime, [19], because the goal is not to maximise recall of dated events but to expose the precision ladder transparently: each class is a published-in-this-paper regex inventory whose recall–precision tradeoff is checkable by any reader.

6 Empirical Results

6.1 Corpus Composition

After streaming-aggregation, the 2023–2024 corpus yields 254,153 unique authors over 324,548 user-months (mean 1.28 months per author; median 1). Of those user-months, 7,817 are flagged as family-loss under the Coarse regex, 7,830 under the Explicit regex, and 2,339 under the Recent regex. Although the Coarse and Explicit totals are nearly equal, the two sets are *not* nested: the Coarse regex admits the bare "funeral" pattern with no explicit kinship + verb match, while the Explicit regex admits the broader kinship inventory of Section 5 together with the extended death-verb phrasings. Empirically, the two sets overlap in 6,504 user-months, but each retains roughly 1,300 user-months exclusive to itself (1,313 Coarse-only and 1,326 Explicit-only). The Recent set, by construction, is a strict subset of Explicit obtained by intersecting with the recency-marker inventory, and is about 30% the size of Explicit. The recency-marker filter is therefore the most stringent of the three and, as we show below, is the source of the precision–power tradeoff that dominates the empirical interpretation of the bound.

6.2 Population Mean Drift at Three Precision Classes

Table 1 reports the population mean of the pre-event minus baseline difference $\mu_i^{pre} - \mu_i^{base}$ under each precision class, together with the number of users on which the difference can be computed, Cohen's *d*, the one-sample *t*-test statistic, and the unadjusted *p*-value.

Table 1. Population mean pre-event drift by family-loss event-detection precision. Drift is the per-user pre-event-minus-baseline affect difference μpre − μbase (standardised VADER units); n is the number of users with both a pre-event and a baseline window. Formal between-class and subpopulation comparisons are reported in Sections 6.2–6.2.2.

Precision class	n users	Mean drift	95% CI	Cohen's *d*
Coarse	356	−0.1094	[−0.2066, −0.0123]	−0.117
Explicit (kin + verb)	352	−0.1565	[−0.2577, −0.0554]	−0.162
Recent (explicit + recency)	125	−0.0045	[−0.1579, +0.1489]	−0.005

Source: created by the authors.

A forest-plot summary of Table 1 is in Figure 1.

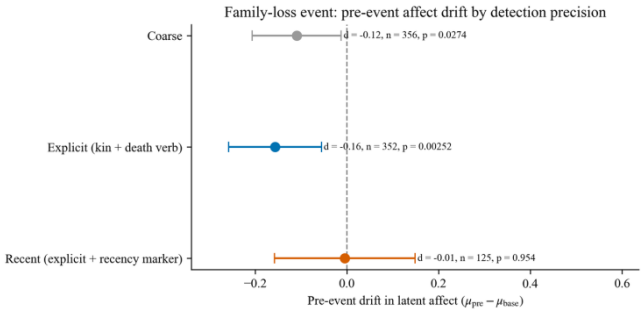

Fig. 1: Pre-event affect drift at three regex-derived family-loss event-detection precision classes. Bars are 95% *t*-CIs; annotations report Cohen's *d*, the number of users with both a pre-event and a baseline window, and the unadjusted *p*-value. The Coarse → Explicit step moves the recovered drift magnitude in the direction predicted by Proposition 3; the Explicit → Recent step is inconclusive because the recency filter cuts the user pool from 352 to 125, inflating the sampling variance of the mean.
Source: created by the authors.

The hypothesised monotonicity from Proposition 3 is that the magnitude of the recovered drift should increase (in the direction of a larger absolute pre-event affect drop) as we step from Coarse to Explicit to Recent. The empirical evidence is mixed.

Coarse → Explicit: prediction confirmed. Tightening the regex from the companion paper's coarse kinship+verb inventory to the explicit kinship+verb inventory increases the mean drift magnitude from 0.1094 to 0.1565 standardised VADER units (both negative; population means of pre-event affect are below baseline). Cohen's *d* moves from −0.117 to −0.162, the one-sample *t*-statistic from −2.22 to −3.04, and the Benjamini–Hochberg adjusted p-value, [20], from 0.041 to 0.008. The user pool is nearly unchanged (n=356→352); the small additional kin terms in the Explicit inventory add little recall but improve the affect-drop estimate, which is the direction predicted by Proposition 3.

Explicit → Recent: underpowered. Adding the recency-marker filter on top of Explicit drops the user pool from 352 to 125 (a 65% loss), and the mean drift estimate collapses to −0.0045 with a 95% CI of [−0.158, +0.149]. The CI of the Recent class is wide enough to be consistent with the Explicit point estimate (−0.157) and with the null. The data therefore cannot distinguish, at this corpus size, between two plausible interpretations: (i) the bound is correct and the missing power simply prevents us from confirming a still-larger Recent-class drift; or (ii) the recency markers select a different sub-population (e.g. users who post within days of the event, possibly more acute grievers) whose pre-event baseline is genuinely different from the larger Explicit pool. The contrast is unidentified from this corpus alone.

To compare the precision classes with one another — rather than only against a null of zero — we add two pairwise tests. Among the users flagged under both the Coarse and Explicit regexes for whom the drift is computable under both event-datings (n = 278), a paired t-test finds no significant within-subject change in the pre-event drift when the event date is moved from the Coarse to the Explicit regex (means −0.118 and −0.133 standardised VADER units; t(277) = 1.65, p = 0.10). For the mutually exclusive subpopulations, the Coarse-only users (n = 66, mean −0.044) and the Explicit-only users (n = 69, mean −0.263) do not differ significantly either (Welch's t = 1.18, p = 0.24, Cohen's d = 0.20). Both tests reinforce the composition reading of Section 6.2.1: the Coarse → Explicit magnitude gain is not a significant within-subject tightening, and the exclusive subpopulations, although trending in the predicted direction, are not separable at this corpus size. These comparisons, together with the two categorical tests of Section 6.2.1, are collected in Table 2.

Table 2. Formal between-class and categorical tests. The paired and two-sample t-tests compare the precision classes directly rather than each against zero; the two chi-square tests of independence assess whether membership composition (Section 6.2.1) and recency-filter retention (Section 6.2.2) depend on baseline affect. All tests use the user-level partition of Section 6.2.1. n, number of users; d, Cohen's d; p, unadjusted.

Test	Sample	Statistic	p	Interpretation
Paired t (Coarse vs. Explicit dating)	n = 278	t(277) = 1.65	0.10	within-subject shift not significant
Two-sample (Coarse-only vs. Explicit-only)	n = 66 vs. 69	Welch t = 1.18; d = 0.20	0.24	exclusive subpopulations not distinguishable
χ^2 independence (membership × baseline)	n = 418	$\chi^2(4)$ = 5.54	0.24	category independent of baseline affect
χ^2 Recent-drop (kept/dropped × baseline)	n = 352	$\chi^2(2)$ = 1.07	0.58	recency drop is homogeneous across the baseline

Source: created by the authors.

6.2.1 Precision Tightening or Subpopulation Swap?

Because the Coarse and Explicit user-month sets overlap in 6,504 of roughly 7,800 user-months and differ by only about 1,300 exclusive observations on each side, the Coarse → Explicit shift in Table 1 could be a composition effect — a swap of subpopulations — rather than a genuine tightening of event-date precision. We test this directly by decomposing the comparison by user-level set membership. Restricting both estimates to the 6,201 users flagged under both regexes and dating each user's event by the corresponding regex, the magnitude still moves in the predicted direction but only marginally: from −0.124 (n=290) under Coarse dating to −0.130 (n=283) under Explicit dating. The larger full-pool shift is driven by the exclusive observations: the Coarse-only users carry a near-null drift (−0.044, n=66, p=0.72), whereas the Explicit-only users carry a markedly larger drift (−0.263, n=69, p=0.06). We therefore report a more guarded claim than the full-pool comparison alone would support — the direction predicted by Proposition 3 is preserved within the common subpopulation, but most of the apparent magnitude gain at this step reflects which posts are admitted, not how precisely their dates are known. This dovetails with a conceptual caveat about the ladder itself: the Coarse → Explicit step is primarily a positive-class-precision step (it changes how accurately a post is classified as describing a family-loss event at all), whereas only the Explicit → Recent step introduces a genuine temporal marker that tightens the date error Δ. The regex ladder is thus a clean proxy for event-date precision only at its final step — and that step is precisely the one the present corpus is underpowered and the monthly observation grid too coarse (Δ≈7 days versus ≈30 days, both fall within one month) to resolve.

We also test the composition claim categorically. Assigning every family-loss user to one of three membership categories — Coarse-only, Explicit-only, or Overlapping — and cross-tabulating against baseline-affect terciles, a chi-square test of independence finds no association between category membership and baseline affect ($\chi^2(4) = 5.54$, p = 0.24; smallest expected cell 21.9). The exclusive subpopulations are thus not distinguished by their quiescent affect level: the Coarse → Explicit swap changes which posts are admitted without selecting users of systematically different baseline affect. The full decomposition and both categorical tests are shown in Figure 2 (Appendix).

6.2.2 Does the Recency Filter Select a Distinct Subpopulation?

A remaining concern is whether the recency-marker filter culls the Explicit pool evenly across baseline affect, or whether it retains a distinct subpopulation. Within the Explicit pool (n = 352 users with a computable baseline), we cross-tabulate retention by the Recent filter (kept versus dropped) against baseline-affect terciles. A chi-square test of independence finds no association ($\chi^2(2) = 1.07$, p = 0.58; smallest expected cell 38.2): the recency filter retains a statistically constant fraction of users across low-, mid-, and high-baseline strata (kept fractions 0.35, 0.29, and 0.34). The recency filter, therefore, acts as a precision filter on the event date rather than as a selector of a baseline-distinct subpopulation; the collapse of the Recent-class drift estimate in Table 1 reflects the three-fold loss of sample size, not a change in who is retained.

6.3 Bootstrap Sampling Distribution of the Population Mean

Figure 3 plots the bootstrap sampling distribution of the population mean drift at each precision class. The three distributions are stacked on the same axis with semi-transparent fills; the vertical zero line marks the null hypothesis. The Coarse and Explicit bootstrap distributions are visibly displaced below zero with little overlap with the null; the Recent bootstrap distribution is roughly centred on zero with a wide spread, reflecting the third of the user pool that remains under the recency filter. We refer to these as bootstrap *sampling* distributions throughout, in contrast to the Bayesian posterior of Proposition 3.

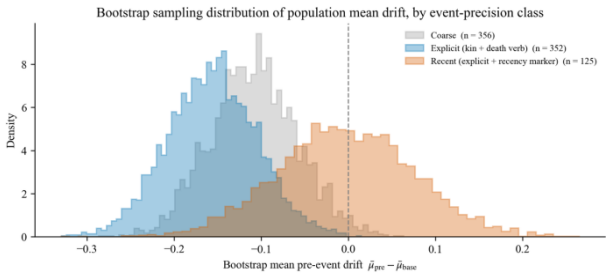

Fig. 3: Bootstrap sampling distributions of the population mean pre-event drift at three precision classes (4,000 resamples each; density normalised). The Recent class is a strict subset of the Explicit class; the Coarse and Explicit classes overlap heavily but are not nested (Coarse admits the bare "funeral" pattern; Explicit admits broader kinship terms). Each class's reference user pool changes with the precision, so the variances of the three bootstrap means are not directly comparable on a

like-for-like basis. The Coarse and Explicit bootstrap distributions are concentrated below zero; the Recent bootstrap distribution is wider and centred near zero, illustrating the precision–power tradeoff that the bound of Proposition 3 does not directly address.

Source: created by the authors.

6.4 Per-User Trajectories Around the Family-Loss Event

Figure 4 shows monthly affect trajectories for the twelve users with the largest absolute pre-event affect drop in the Recent precision class. The trajectories are centred on the detected event month, with horizontal axes in months relative to that reference. The vertical dotted line marks t=0, and the horizontal zero line marks neutral affect.

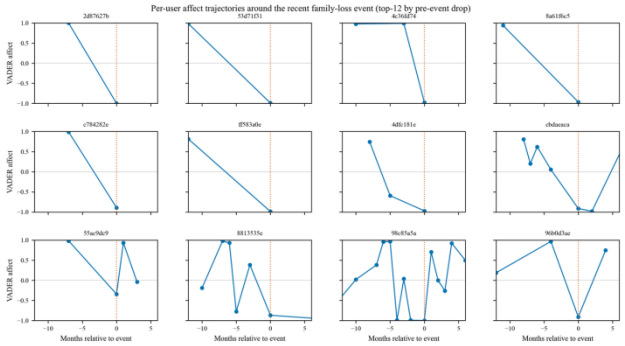

Fig. 4: Per-user monthly VADER affect trajectories centred on the detected family-loss month, for the twelve users with the largest pre-event affect drop in the Recent precision class. Horizontal axis: months relative to the event. Vertical red dotted line: event month.

Source: created by the authors.

7 Discussion

The empirical evidence supports the direction predicted by Proposition 3 on the Coarse → Explicit step: a tighter event-detection regex enlarges the recovered drift magnitude and sharpens its p-value. Section 6.2.1, however, shows that most of this magnitude gain is a subpopulation-composition effect: within the users common to both regexes, the shift is small (−0.124 to −0.130), so the confirmatory weight of this step is weaker than the full-pool estimates suggest. The Explicit → Recent step does not preserve this monotonicity, but the failure is most parsimoniously attributed to a power loss—the recency filter reduces the eligible user pool by a factor of three—rather than to a violation of the bound. Crucially, the asymptotic statement of Proposition 3 is silent about the size of the user pool at each precision class: it controls the *posterior mean* bias as $N\to\infty$, but on a finite corpus, the variance of that estimator scales as $1/n$ and can dominate the bias improvement for sufficiently aggressive regex tightening. This is the empirical *precision–power tradeoff* that the bound by itself does not address.

Quantitative agreement with the bound's constant $C_{\theta,w}$ is harder to assess from a single corpus: estimating θ_0 in calendar units from a single 24-month subreddit corpus is identifiability-limited, and the precision ladder cannot give us a Δ=0 reference without exogenous event dates. The companion paper, [6], verifies the bound's *magnitude* synthetically; the present paper verifies the *direction* on real data and exposes the practical sample-size constraint that future text-derived event-time studies must calibrate against.

What the result lets us claim. We obtain population-mean pre-event affect drift estimates that are statistically distinguishable from zero in the Coarse class (p_{BH}=0.041) and the Explicit class (p_{BH}=0.008). Cohen's d at the Explicit class is −0.162, which is small by clinical-trial conventions but typical of community-wide social-media affect studies, and is recovered on n=352 users—a sample size well within the identifiability regime of Proposition 2 ($N\to\infty$ with $K_i\geq 2$). The two-month pre-event window appears to capture a real, lexicon-detectable decline in textual affect among a subset of subreddit authors who explicitly write about a kinship-related death.

What the result does not let us claim. The Recent class is consistent with everything from a substantial negative drift to a small positive one. We cannot conclude that the recency filter improves the estimate, nor that it fails to. The Recent estimate should be read as a power-limited upper-bound exercise rather than a contradiction of the bound. More broadly, the confirmed signal is a single directional step, recovered from a single self-selected subreddit and conditional on posting; it is best read as a method demonstration on a single real corpus rather than a population-level finding about bereavement affect.

Limitations. Several limitations apply.

Regex coarseness. The kinship+verb regex matches false positives ("my mom would die for that") and misses paraphrases ("we lost her last year", "she's no longer with us"). VADER is similarly lexicon-limited and may misclassify quoted speech or sarcasm. This risk is elevated in a bereavement corpus, where quoting the deceased and dark or sarcastic phrasing are both common and could bias the measured affect drop in either

direction. A sensitivity analysis to a transformer-based detector (e.g. RoBERTa-base, [21], fine-tuned on a labelled subset) is therefore the priority robustness check; we note that it cannot be run on the released artefacts, because the privacy design of this study (Section 5.1) retains only aggregate per-user-month VADER features and not raw post text, so a transformer rescore requires re-deriving features from the archive under the same pipeline and is registered as the primary next step.

Self-selection. Users who post on the chosen subreddit are self-selected; they are not a random sample of the population. The drift estimate is conditional on subreddit posting, not on family loss in the general population.

Single subreddit. The empirical chapter is restricted to one subreddit; cross-subreddit generalisation is left to a follow-up.

Calendar-month resolution. The observation grid is monthly, while the event-precision ladder reaches as fine as ∼7 days. We cannot separate within-month event-time noise from monthly observation discretisation without supplementary data; the bound's Δ should be interpreted as a regex-class proxy rather than a calendar-day precision.

Ethics. The corpus contains sensitive narratives of family loss. Only hashed user identifiers, post identifiers, and aggregate statistics are retained; raw post text is not redistributed. The study uses only publicly accessible posts and does not involve contact with subjects.

8 Conclusions

We provide the first real-data test — a qualified, method-demonstration-level test — of the posterior-bias bound of Proposition 3 in the companion paper, [6]. Using a regex-specificity ladder on a self-disclosure-rich Reddit subreddit (254,153 users, 324,548 user-months over 2023–2024) and targeting the family-loss instance of an adverse life event, the recovered population mean drift in latent affect moves in the predicted direction on the Coarse → Explicit step (magnitude grows from 0.109 to 0.157 standardised VADER units; Cohen's d from −0.117 to −0.162), but the Explicit → Recent step is inconclusive because the recency-marker filter cuts the eligible user pool by two-thirds and inflates the sampling variance of the mean beyond the bias improvement. Section 6.2.1 further shows that, even at the Coarse → Explicit step, most of the magnitude gain is a subpopulation-composition effect rather than a within-subject precision gain. The contribution is threefold: (i) a real-data complement to the previously synthetic-only theoretical guarantee, (ii) a reusable regex-specificity ladder for testing event-precision sensitivity in any text-derived event-time study, and (iii) an explicit acknowledgement of the practical precision–power tradeoff that the asymptotic bound does not address, with a concrete sample-size calibration point (n=125 underpowered, n=352 adequate) for future researchers choosing detection-regex specificity on bounded corpora.

Acknowledgments:
The authors thank the maintainers of the Arctic Shift archive for continued public access to the Reddit corpus. The presentation and dissemination of these research results are supported by the National Science Fund Project КП-06-Н85-7/05.12.2024 "Significance and Potential Risks of the Fast Integration of Artificial Intelligence (AI) Technologies into the Economy and Financial Sector".

References:

[1] Ricciardi, L. M., Sacerdote, L. The Ornstein–Uhlenbeck process as a model for neuronal activity. *Biological Cybernetics*, 35, No. 1, pp. 1–9, 1979. DOI: 10.1007/BF01845839.

[2] Kantas, N., Doucet, A., Singh, S. S., Maciejowski, J., Chopin, N. On particle methods for parameter estimation in state-space models. *Statistical Science*, 30, No. 3, pp. 328–351, 2015. DOI: 10.1214/14-STS511.

[3] Luhmann, M., Hofmann, W., Eid, M., Lucas, R. E. Subjective well-being and adaptation to life events: A meta-analysis. *Journal of Personality and Social Psychology*, 102, No. 3, pp. 592–615, 2012. DOI: 10.1037/a0025948.

[4] Driver, C. C., Voelkle, M. C. Hierarchical Bayesian continuous time dynamic modeling. *Psychological Methods*, 23, No. 4, pp. 774–799, 2018. DOI: 10.1037/met0000168.

[5] Voelkle, M. C., Oud, J. H. L. Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes. *British Journal of Mathematical and Statistical Psychology*, 66, No. 1, pp. 103–126, 2013. DOI: 10.1111/j.2044-8317.2012.02043.x.

[6] Naskinova, I., Milev, M., Netov, N., Kolev, M., et al. A hierarchical stochastic differential equation framework with particle-filter inference for latent-state dynamics around life-event mentions on Reddit. Proceedings of

the International Conference on Mathematical Methods & Computational Techniques in Science & Engineering (MMCTSE 2026), Venice, Italy, 2026, in press.

[7] Øksendal, B. Stochastic differential equations: An introduction with applications, 6th ed., Universitext, Springer, Berlin, Heidelberg, Germany, 2003, ISBN 978-3-540-04758-2. DOI: 10.1007/978-3-642-14394-6.

[8] Schuurman, N. K., Hamaker, E. L. Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods*, 24, No. 1, pp. 70–91, 2019. DOI: 10.1037/met0000188.

[9] Liu, J., West, M. Combined parameter and state estimation in simulation-based filtering. *Sequential Monte Carlo Methods in Practice*, Springer, New York, NY, USA, pp. 197–223, 2001, ISBN 978-1-4757-3437-9. DOI: 10.1007/978-1-4757-3437-9_10.

[10] Doucet, A., Johansen, A. M. A tutorial on particle filtering and smoothing: Fifteen years later. In The Oxford Handbook of Nonlinear Filtering, Oxford University Press, Oxford, UK, pp. 656–704, 2011, ISBN 978-0-19-953290-2.

[11] Chopin, N., Papaspiliopoulos, O. An introduction to sequential Monte Carlo. Springer Series in Statistics, Springer, Cham, Switzerland, 2020, ISBN 978-3-030-47844-5. DOI: 10.1007/978-3-030-47845-2.

[12] Andrieu, C., Doucet, A., Holenstein, R. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72, No. 3, pp. 269–342, 2010. DOI: 10.1111/j.1467-9868.2009.00736.x.

[13] Kalman, R. E. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82, No. 1, pp. 35–45, 1960. DOI: 10.1115/1.3662552.

[14] Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., Blackburn, J. The Pushshift Reddit dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, 14, No. 1, pp. 830–839, 2020, Atlanta, GA, USA (held online). DOI: 10.1609/icwsm.v14i1.7347.

[15] Arctic Shift Project. Arctic shift: An open archive of the reddit corpus. https://arctic-shift.photon-reddit.com, 2025. Last accessed: 17 May 2026.

[16] De Choudhury, M., Kiciman, E. The language of social support in social media and its effect on suicidal ideation risk. *Proceedings of the International AAAI Conference on Web and Social Media*, 11, No. 1, pp. 32–41, 2017, Montréal, Québec, Canada. DOI: 10.1609/icwsm.v11i1.14891.

[17] Hutto, C., Gilbert, E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8, No. 1, pp. 216–225, 2014, Ann Arbor, MI, USA. DOI: 10.1609/icwsm.v8i1.14550.

[18] Pennebaker, J. W., Boyd, R. L., Jordan, K., Blackburn, K. The development and psychometric properties of LIWC2015. University of Texas at Austin, Austin, TX, USA, 2015. DOI: 10.15781/T29G6Z.

[19] Strötgen, J., Gertz, M. HeidelTime: High quality rule-based extraction and normalization of temporal expressions. *Proceedings of the 5th International Workshop on Semantic Evaluation*, Association for Computational Linguistics, Uppsala, Sweden, pp. 321–324, 2010.

[20] Benjamini, Y., Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, No. 1, pp. 289–300, 1995. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

[21] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019. DOI: 10.48550/arXiv.1907.11692.

APPENDIX

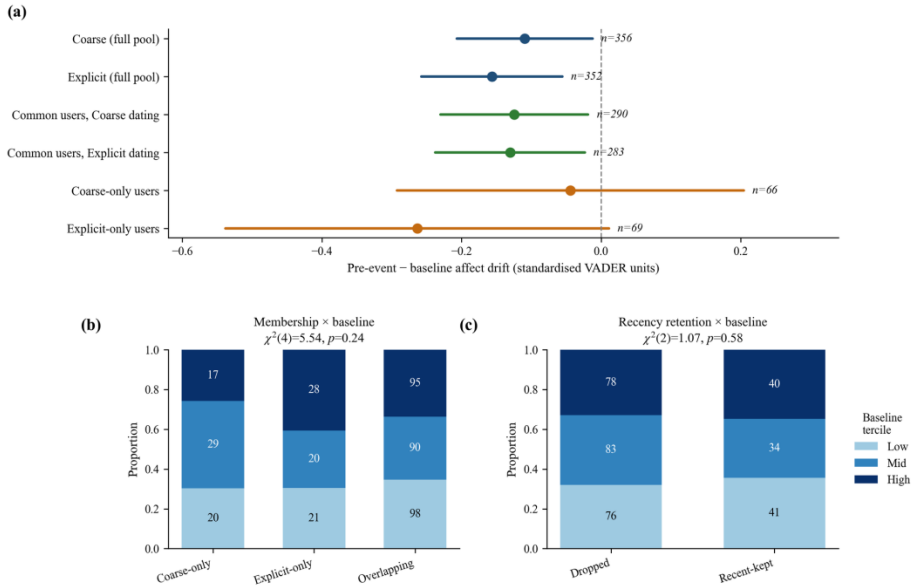

Fig. 2: Robustness of the Coarse → Explicit step to subpopulation composition. (a) Pre-event − baseline affect drift (points, mean; horizontal bars, 95% t-confidence interval; dashed line, zero) for the full Coarse and Explicit pools (navy), the common users re-dated under each regex (green), and the mutually exclusive Coarse-only and Explicit-only users (amber); the user count n is annotated at the right of each interval. (b) Membership category (Coarse-only, Explicit-only, Overlapping) by baseline-affect tercile (Low, Mid, High), shown as within-group proportions with cell counts; $\chi^2(4) = 5.54$, $p = 0.24$. (c) Recency-filter retention (Dropped vs. Recent-kept) within the Explicit pool by baseline-affect tercile; $\chi^2(2) = 1.07$, $p = 0.58$. Panels (b,c) show that neither the exclusive subpopulations nor the recency-dropped users differ systematically in baseline affect.
Source: created by the authors.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Kristian Milev, Angelina Angelova, Irina Naskinova, Mariyan Milev and Mikhail Kolev were responsible for the conceptualization of the study. Angelina Angelova, Kristian Milev, Irina Naskinova, Mariyan Milev, Mikhail Kolev and Hristo Kalinov developed the methodology. Kristian Milev, Mariyan Milev and Gabriela Vasileva implemented the software. Angelina Angelova, Nikolay Netov, Mariyan Milev, Mikhail Kolev, Hristo Kalinov and Gabriela Vasileva carried out the formal analysis. Gabriela Vasileva carried out the investigation and the data curation. Gabriela Vasileva and Irina Naskinova wrote the original draft, and Irina Naskinova, Mariyan Milev, Mikhail Kolev, Hristo Kalinov and Gabriela Vasileva contributed to the review and editing of the manuscript. Gabriela Vasileva prepared the visualization. Mariyan Milev and Mikhail Kolev supervised the work, Mariyan Milev was responsible for the project administration, and Nikolay Netov, Mariyan Milev and Mikhail Kolev for the funding acquisition. All authors have read and agreed to the published version of the manuscript.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

The presentation and dissemination of these research results are supported by National Science Fund Project КП-06-Н85-7/05.12.2024 "Significance and Potential Risks of the Fast Integration of Artificial Intelligence (AI) Technologies into the Economy and Financial Sector". No other funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.