Embedding-Based Measures of Life-Course Volatility and Disorganization in Multidimensional Trajectories

Extended Abstract

How should computational social scientists measure the complexity of modern life courses? Sociological theories of destandardization and individualization argue that contemporary biographies have become less linear, less institutionally scripted, and more difficult to summarize through conventional transition markers such as school completion, marriage, parenthood, or stable employment (Beck, 1992; Beck and Beck-Gernsheim, 2002; Brückner and Mayer, 2005; Elzinga and Liefbroer, 2007). Yet empirical research often treats life-course complexity as a single construct. Existing sequence-analytic measures such as turbulence and entropy are powerful descriptive tools (Abbott, 1995; Elzinga and Liefbroer, 2007), but they tend to compress multiple forms of complexity into one number, depend on researcher-defined state spaces and distance costs, and face practical limitations when life histories span several interdependent domains (Ritschard et al., 2023). This paper develops a machine-learning-based framework for measuring multidimensional life-course complexity and uses it to distinguish two analytically separate dimensions: *local volatility*, or short-run step-to-step movement between life states, and *global disorganization*, or longer-run inefficiency, backtracking, and pathway incoherence.

The empirical application uses the British Cohort Study 1970, a nationally important longitudinal cohort following individuals born in Great Britain in 1970. I reconstruct multidimensional life trajectories from ages 16 to 46 using quarterly life-state sequences across six domains: work, education, health, caregiving, partnership, and fertility. Each quarterly observation represents a composite state describing the individual's position across these domains. I then link the complexity measures of these trajectories to age-51 outcomes, including life satisfaction, relationship satisfaction, mental wellbeing, general health, social trust, work limitation due to health, weekly family income, and home ownership. This design allows the paper to ask not only whether life-course complexity can be measured in a more data-driven approach, but also whether different forms of complexity matter for later-life wellbeing and inequality.

Methodologically, the paper uses a Transformer autoencoder to learn dense vector representations of multidomain life states. The intuition is similar to representation learning in natural language processing: the meaning of a life state is inferred from the other states that precede, follow, and co-occur with it in observed trajectories. This builds on recent work using Transformer-based representations of life events and life histories (Vaswani et al., 2017; Savcīsēns et al., 2024). Rather than imposing substitution costs or assuming that all state changes are equivalent, the model learns a continuous "life-state space" from the data. Each individual trajectory is therefore represented as a path through this learned embedding space:

$$v_{i1}, v_{i2}, \ldots, v_{iT}$$

where v_{it} is the embedding of individual i's multidomain life state at quarter t.

From this embedded trajectory, I derive two measures, summarized in Table 1. The first is *local volatility*, measured as mean consecutive distance:

$$LV_i = \frac{1}{T-1} \sum_{t=1}^{T-1} d\left(v_{it}, v_{i,t+1}\right),$$

where $d(\cdot)$ is cosine distance. This captures the average magnitude of short-run movement between adjacent life states. A high value indicates frequent or large step-to-step changes. The second is *global disorganization*, measured as global optimality divergence:

$$GD_i = L_{\text{actual},i} - L_{\text{optimal},i}.$$

Here, $L_{\text{actual},i}$ is the length of the observed path through embedding space, while $L_{\text{optimal},i}$ is the shortest path needed to connect the unique states visited by that individual. A high value indicates that the observed trajectory is "inefficient" relative to the set of states visited, suggesting revisiting, oscillation, or backtracking. The two measures are adapted from embedding-based approaches to semantic trajectories (Nour et al., 2023) and are conceptually distinct: a life can involve frequent local movement while remaining globally coherent, or it can involve relatively small-step transitions arranged in a globally inefficient or disorganized way.

I validate the measures in three steps. First, I simulate synthetic life courses under low, medium, and high switching regimes. Both local volatility and global disorganization increase monotonically with switching probability, indicating that the measures recover known differences in trajectory instability. Second, I construct stylized life-course archetypes, such as stable marriage, persistent singlehood, work-only churn, care volatility, repeated partnership oscillation, and global churn. These archetypes map onto the expected locations in volatility–disorganization space: stable biographies score low on both dimensions, while global churn scores high on both, and domain-specific churn produces intermediate or off-diagonal profiles. Third, I compare the proposed measures with the established sequence-analysis turbulence index. As shown in Table 2, the Spearman rank correlations show moderate overlap rather than redundancy: local volatility and global disorganization are strongly related but distinct, and global disorganization is less strongly correlated with conventional turbulence than local volatility is.

The empirical results show that life-course complexity is multidimensional and consequential, with the main findings summarized in Table 3. In models predicting age-51 outcomes, higher ranks indicate lower underlying complexity, or more stable and organized trajectories. Lower global disorganization is associated with better mental wellbeing, better general health, higher weekly family income, greater home ownership, and a lower probability that health limits work, even after adjusting for sex, ethnicity, and father's occupational class. Global disorganization is especially important for outcomes linked to midlife wellbeing and economic stability, and it is the only index that remains significantly associated with weekly income in adjusted models. Local volatility is also associated with several outcomes, including life satisfaction, relationship satisfaction, general health, home ownership, and health-related work limitation, but its associations are generally weaker and less consistent. Traditional turbulence remains competitive for some outcomes, especially life satisfaction and home ownership. The results therefore do not imply that embedding-based measures simply replace traditional sequence analysis; rather, they show that global pathway organization captures an outcome-relevant dimension that conventional turbulence does not fully summarize.

The paper also examines whether volatility and disorganization are socially stratified. Women in the 1970 cohort display higher local volatility, higher global disorganization, and higher turbulence than men, consistent with gendered life-course patterns involving caregiving, part-time work, and family-related interruptions. Class differences are more pronounced for local volatility than for global disorganization. Respondents from professional-class origins show more local volatility, while long-run pathway organization is less clearly stratified by childhood class. This pattern suggests that short-run movement is not always a straightforward sign of disadvantage: for advantaged groups, some forms of volatility may reflect exploration, flexibility, or strategic reorientation rather than simple precarity. However, because higher volatility is still associated with worse later-life outcomes on several measures, the findings also caution against interpreting instability as uniformly liberating. The social meaning of mobility depends on whether movement is supported by resources and whether it remains globally organized over time.

Finally, domain-specific analyses show that different parts of the life course contribute differently to volatility and disorganization. Work and partnership turbulence are positively associated with both embedding-based indices, while sickness-related turbulence is negatively associated, suggesting that health disruption may constrain rather than diversify life paths. Care turbulence shows gendered patterns, operating differently for men and women, while fertility turbulence pulls in opposite directions across the two dimensions: it increases short-run volatility but is associated with lower global disorganization. These findings reinforce the central argument that life-course complexity cannot be reduced to the number of transitions alone. What matters is how transitions are sequenced, across which domains they occur, and whether they form a coherent or backtracking pathway.

This paper contributes to computational life course analysis in three ways. First, it introduces a representation-learning approach for measuring multidimensional life trajectories without relying on researcher-defined distance costs. Second, it proposes a conceptual and empirical distinction between short-run volatility and long-run disorganization, showing that these dimensions are non-identical. Third, it demonstrates that embedding-based measures can be sociologically interpretable: they predict later-life wellbeing, health, and economic outcomes; reveal gender and class patterning; and complement rather than replace established sequence-analysis tools. More broadly, the paper shows how machine learning can be used not merely for prediction, but for theoretically motivated measurement in social science.

Tables and Figures

Table 1: Embedding-based measures of life-course complexity

Measure	Definition	Interpretation
Local volatility	Mean cosine distance between consecutive embedded life states.	Captures short-run step-to-step switching; high values indicate frequent or large local moves.
Global disorganization	Actual path length minus shortest path connecting all unique visited states.	Captures inefficiency, revisiting, oscillation, and backtracking in the overall trajectory.
Average turbulence	Mean of domain-specific sequence turbulence scores.	Conventional benchmark summarizing sequence complexity using established sequence-analysis methods.

Table 2: Descriptive statistics and rank correlations of trajectory indices

Index	N	Mean	SD	Selected rank correlation
Average turbulence	13166	5.610	1.504	$\rho(\text{Turb.}, LV) = 0.609$
Local volatility	13564	0.012	0.007	$\rho(LV, GD) = 0.605$
Global disorganization	13564	0.192	0.223	$\rho(\text{Turb.}, GD) = 0.379$

Note: Correlations are Spearman rank correlations. LV denotes local volatility; GD denotes global disorganization. All reported correlations are statistically significant at $p < 0.001$.

Table 3: Summary of main empirical findings

Dimension	Main finding
Measurement validity	Simulations and stylized archetypes show that both indices recover expected differences between stable, domain-specific churn, and globally churning trajectories.
Outcome associations	Global disorganization is especially predictive of mental wellbeing, general health, and weekly family income at age 51; local volatility has weaker and less consistent associations.
Comparison with turbulence	Traditional turbulence remains competitive for some outcomes, especially life satisfaction and home ownership, suggesting complementarity rather than replacement.
Social stratification	Women have more volatile and disorganized trajectories. Class differences are stronger for local volatility than for global disorganization.
Domain-specific patterns	Work and partnership turbulence are associated with greater complexity; care turbulence is negatively associated with complexity among men but positively associated with complexity among women; fertility turbulence increases local volatility but is associated with lower global disorganization.

References

Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology*, 21:93–113.

Beck, U. (1992). *Risk Society: Towards a New Modernity.* Sage.

Beck, U. and Beck-Gernsheim, E. (2002). *Individualization: Institutionalized Individualism and Its Social and Political Consequences.* Sage.

Brückner, H. and Mayer, K. U. (2005). De-standardization of the life course: What it might mean? and if it means anything, whether it actually took place? *Advances in Life Course Research*, 9:27–53.

Elzinga, C. H. and Liefbroer, A. C. (2007). De-standardization of family-life trajectories of young adults: A cross-national comparison using sequence analysis. *European Journal of Population*, 23(3–4):225–250.

Nour, M. M., McNamee, D. C., Liu, Y., and Dolan, R. J. (2023). Trajectories through semantic spaces in schizophrenia and the relationship to ripple bursts. *Proceedings of the National Academy of Sciences of the United States of America*, 120(42):e2305290120.

Ritschard, G., Liao, T. F., and Struffolino, E. (2023). Strategies for multidomain sequence analysis in social research. *Sociological Methodology*, 53(2):288–322.

Savcīsēns, G., Eliassi-Rad, T., Hansen, L. K., Mortensen, L. H., Lilleholt, L., Rogers, A., Zettler, I., and Lehmann, S. (2024). Using sequences of life-events to predict human lives. *Nature Computational Science*, 4(1):43–56.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.