



# Stability and spread: A novel method for quantifying transitions within multivariate binary time series data

Katharine E. Daniel<sup>1</sup> · Robert G. Moulder<sup>1</sup> · Bethany A. Teachman<sup>1</sup> · Steven M. Boker<sup>1</sup>

Accepted: 22 July 2022 / Published online: 24 August 2022  
© The Psychonomic Society, Inc. 2022

## Abstract

We present a novel method for quantifying transitions within multivariate binary time series data, using a sliding series of transition matrices, to derive metrics of stability and spread. We define *stability* as the trace of a transition matrix divided by the sum of all observed elements within that matrix. We define *spread* as the number of all non-zero cells in a transition matrix divided by the number of all possible cells in that matrix. We developed this method to allow investigation into high-dimensional, sparse data matrices for which existing binary time series methods are not designed. Results from 1728 simulations varying six parameters suggest that unique information is captured by both metrics, and that stability and spread values have a moderate inverse association. Further, simulations suggest that this method can be reliably applied to time series with as few as nine observations per person, where at least five consecutive observations construct each overlapping transition matrix, and at least four time series variables compose each transition matrix. A pre-registered application of this method using 4 weeks of ecological momentary assessment data ( $N = 110$ ) showed that stability and spread in the use of 20 emotion regulation strategies predict next timepoint affect after accounting for affect and anxiety's auto-regressive and cross-lagged effects. Stability, but not spread, also predicted next timepoint anxiety. This method shows promise for meaningfully quantifying two unique aspects of switching behavior in multivariate binary time series data.

**Keywords** Multivariate binary time series · High dimensionality · Transition matrix · Switching · Emotion regulation

Patterns within multivariate binary time series occur everywhere. To understand individual differences in how people use various emotion regulation strategies throughout their lives, one could repeatedly ask them whether or not they were using each of the following common strategies: distraction, cognitive reappraisal, expressive suppression, experiential avoidance, acceptance, social support seeking, and rumination. Their answers would result in a multivariate binary time series space comprised of seven univariate time series (one time series per strategy).

To clarify how multivariate binary time series data can contain complex patterns, consider the choices of a chef. When cooking, a chef can choose to flavor their food with or without cumin. The sequence of a chef's choice to use or not to use this spice over time can be modeled as a binary time series. On its own, the information contained in this time series can provide some insight into what cuisine the chef likely specializes in. However, chefs generally combine many spices to create their desired flavor profiles. Supposing a chef stocks ten spices, then ten binary univariate time series—not one—comprise the choices that define a multivariate system. Modeling univariate binary time series in isolation of other relevant time series can obscure the true complexity of the system. Without considering the multivariate binary time series space of all ten spices, for example, it is not clear whether the chef's use of cumin at one time point is one of many spices used to create a rich curry or is the standalone spice in a burrito filling. Modeling the use of only one spice at a time also reduces the opportunity to observe the number of unique spice combinations a chef switches between each time they cook. This could obscure

---

✉ Katharine E. Daniel  
ked4fd@virginia.edu

Robert G. Moulder  
rgm4fd@virginia.edu

Bethany A. Teachman  
bat5x@virginia.edu

Steven M. Boker  
smb3u@virginia.edu

<sup>1</sup> Department of Psychology, University of Virginia, P.O. Box 400400, Charlottesville, VA 22904, USA

insight into how the chef varies the meals they cook over time. Novel and nuanced insights into a system are gained when interactions and transitions between multiple binary time series data are considered together. Although binary time series data are common, modeling complex patterns in a high-dimensional multivariate binary time series system can be challenging when using existing methods.

## Existing methods for analyzing binary time series data

Markov models are perhaps the most common class of models used to analyze binary time series outcomes. Anderson and Goodman (1957) first developed hypothesis testing and maximum likelihood estimation procedures to test transition probabilities of Markov chains. Traditionally, Markov models have been used to assess the likelihood that a transition will occur between two states of a variable of interest: e.g., identifying the point when a patient is most likely to transition from being alive to dead (Muenz & Rubinstein, 1985). More recently, Tian and Anderson (2000) generalized these procedures to study joint transition probabilities between more than one variable of interest at a time; however, these procedures are limited to a small number of variables (i.e., four) because they return nonidentifiable parameters when data are sparse. Sparse data matrices can occur when many time series are included because as more time series are added, the size of the state space and the number of potential combinations between time series grows combinatorically. Additionally, modeling every possible state combination as its own discrete state quickly becomes computationally intractable. To illustrate, consider that you are interested in 20 binary time series. If each possible state combination is modeled as its own unique state, then  $2^{20}$  combinations are possible, and the solution quickly becomes intractable.

Recurrence quantification analysis (RQA; Webber & Zbilut, 1994) has been used to understand switching patterns in univariate time series; visualizing and characterizing aspects of change in non-linear dynamical systems. Using a recurrence plot (Eckmann et al., 1987), researchers can derive metrics such as the probability that a specific state will recur (recurrence rate) and the predictability of the system (determinism), among others (see Webber & Marwan, 2015 for review). Since their debut in the early 1990s, bivariate extensions (Marwan & Kurths, 2002; Romano et al., 2004; Zbilut et al., 1998) allow researchers to study the correlation, coupling, or synchronization between two dynamical systems using cross recurrence plots or joint recurrence plots. Multidimensional cross-recurrence quantification analysis extended the bivariate case to study the relationship between multidimensional rather than binary time series (Wallot, 2019). However, the number of time

series that can be investigated jointly with multidimensional RQA becomes computationally intractable as the number of time series under study increases.

While Markov models and RQA have made exciting contributions to the study of complex multivariate time series, they were not designed to analyze high-dimensional multivariate binary data. When data are too sparse to return identifiable parameters with these methods, Tian and Anderson (2000) recommend that researchers collapse their data into fewer transition categories (e.g., by conceptually or empirically using factor-level information, by modeling only those states with conditional independence) or model the processes separately. While these dimension reduction approaches may be sufficient for some research questions, other questions depend on capturing transition information involving many time series. For example, a researcher might want to study how a person transitions between using 40 possible emotion regulation strategies; reducing strategy-level specificity to a few broad categories would prevent asking the research question. We propose the current method as one tool for researchers who are interested in studying transitions within large, complex, high-dimensional systems that are too sparse to be effectively analyzed using existing RQA or Markov chain methods.

## Use cases for high-dimensional binary time series systems

High-dimensional binary time series data are prevalent across many fields. For example, health insurance administrators track which service a patient receives each time they file a claim. Developmental psychologists code what classroom activity a kindergartener is engaging in every 5 minutes throughout the school day. Therapists note where a patient was located each time that they have a panic attack. Dieticians track what types of food their clients eat throughout the day. Linking response patterns across successive observations shows person-level patterns in time-ordered changes between measured states. As such, increasing modeling options for complex, high-dimensional systems have the potential to expand the range of testable research questions afforded by those data streams. For example, a health insurance payee with multiple comorbidities might alternate between using a wider range of services than a healthy payee. A student in a traditional public-school setting might explore fewer activities in the classroom compared to a student in a Montessori school. A patient with post-traumatic stress disorder might tend to exclusively have panic attacks in the location that is linked with a traumatic event whereas a patient with panic disorder might tend to have panic attacks across many different locations. A very picky eater might

report less variety in the food groups they eat over time compared to a more adventurous eater.

## Proposing a new method

To offer a way to analyze large, complex, sparse phase spaces, we present a new method to quantify switching behavior between binary variables over time using transition matrices. This method is specifically designed for research questions interested in how a multivariate binary system switches between *endorsed* states over time. We quantify binary switching according to two dimensions: stability and spread. We define stability as the proportion of transitions within the multivariate time series when the same binary time series is endorsed at consecutive timepoints (i.e., the trace of a transition matrix) relative to all consecutive between- and within-time series transitions observed within the multivariate time series (i.e., the sum of all elements within that transition matrix).<sup>1</sup> This metric is useful when the extent to which a system transitions from endorsing one binary variable to endorsing a different binary variable is of theoretical interest. We define spread as the proportion of unique transitions observed between all possible binary time series within a multivariate time series (i.e., the number of non-zero cells observed in a transition matrix) relative to the total number of possible transitions afforded by the time series (i.e., the number of cells within that transition matrix). This metric is useful when the diversity of the transition states that are observed within a multivariate binary time series affording many possible transition states is of theoretical interest.

However, the quantified pattern for how a system changes over time might not only vary between people but also within individual. For example, a health insurance payee might alternate between using fewer services when they are healthy compared to when they are actively seeking treatment for a health condition. A dietician's client with binge eating disorder might report eating a small range of "safe" foods during the day but report eating many food groups with frequent switching in consumed food during nightly binge episodes. Calculating all the transitions within these systems at once would obscure meaningful within person

changes over time. Thus, this method also incorporates the option to repeatedly calculate stability and spread on different parts of the full timeseries using a sliding series of transition matrices.

First, we mathematically define and describe characteristics of the method according to results from an initial simulation study. We also conduct an initial comparison between our method and RQA. Next, to illustrate its potential to advance theory, we apply the proposed method to a real data example with high socially anxious individuals who repeatedly reported their in-the-moment use of 20 emotion regulation strategies across 4 weeks.

## Methods

In this section, we define our method for measuring concepts of stability and spread within multivariate binary time series data. We calculate stability and spread by first constructing individual-level matrices that count all transitions that occur between successive time points within a multivariate binary time series. Then, we compute stability and spread from the resulting transition matrix. Instead of constructing only one transition matrix using data from the entire time series, which would result in only one stability and one spread value per person, we take a repeated measures approach that is similar to that which was used by Marwan and colleagues (2002). By using small windows that slide over the time series to segment it into a set of subseries, multiple transition matrices are constructed per person. This allows for the detection of within-person variation in stability and spread over time.

### Defining a transition matrix

We define a transition matrix as

$$\mathbf{X}_{ij} : i \in \{1, 2, \dots, N\}, j \in \{1, 2, \dots, J_i\} \quad (1)$$

Where  $\mathbf{X}_{ij}$  is a  $k \times k$  transition matrix for person  $i$  within time window  $j$ ;  $k$  is the number of binary variables to be included in the analysis;<sup>2</sup>  $N$  is the number of participants;

<sup>1</sup> Given our focus on quantifying patterns of switching amongst *endorsed* binary states over time, we define transitions as those that occur between *endorsed* variables in two successive observations (i.e., observed-to-observed). We do not treat unobserved-to-observed, unobserved-to-unobserved, or observed-to-unobserved as transitions because, in the high-dimensional and sparse data systems for which this method was designed, they would far outnumber the observed-to-observed transitions and potentially make it more challenging to detect these less frequent transitions that are of greater interest.

<sup>2</sup> We designed this framework to be flexible. If the researcher is interested in modeling  $n$  states where at least one of the states is always observed (e.g., weather—weather can be described even on a still, clear day), then  $k = n$ . If the researcher is interested in modeling states where it is possible that, at some time points, none of the relevant states would be observed (e.g., medication use—a person is not always taking a pill), then the researcher has two choices. First, they could pre-process the data to create a new time series that codes a 1 every time that none of the measured states were selected and 0 every time that at least one of the measured states was selected. This choice would be appropriate if the researcher is interested in also including transitions into and out of states of "none of the above" and thus  $k = n + 1$ . Alternatively, they could retain  $k = n$  such that if none of

$J_i$  is the number of transition matrices that are constructed for each individual after sub-setting all of the individual's observations into a series of smaller windows of observations. A hyperparameter  $W$  can be defined to set the number of observations that contributes to a given matrix  $\mathbf{X}_{ij}$ , if different from the total number of observations.  $W$  must be a positive integer  $\geq 2$  and cannot exceed the number of observations. The value of  $J_i$  is determined by  $W$  relative to the length of the individual's overall time series and the lag that is set between initial observations of the subseries that construct two successive transition matrices (i.e., the windowing lag). Assuming a windowing lag of one, then

$$J_i = L_i - W + 1 \quad (2)$$

where  $L_i$  is the number of observations in person  $i$ 's overall time series. As Eq. (2) shows, if  $W$  equals the total number of time points observed for person  $i$ , then only one transition matrix will be constructed for that person ( $\mathbf{X}_{i1}$ ). If  $W$  is less than  $L_i$ , then  $J_i > 1$ .

### Building $\mathbf{X}_{ij}$ to depict switches over a multivariate subseries

We first create  $\mathbf{X}_{ij}$  with  $k \times k$  dimensions and initialize all elements to zero. To build  $\mathbf{X}_{i1}$ , we iterate through person  $i$ 's subseries of length  $W$  and increment elements of  $\mathbf{X}_{i1}$  by 1 for each observed transition between  $k$  options for all observations within the first subseries. If the same binary variable (e.g., variable A) was selected at consecutive time points, we increment the diagonal of  $\mathbf{X}_{i1}$  in the element (a,a) of the matrix. If two different binary time series variables were selected at consecutive time points (e.g., variable A then variable B), we increment the off-diagonal of  $\mathbf{X}_{i1}$  in element (b,a). We continue this process until the last transition within individual  $i$ 's first subseries is accounted for, stopping with the  $W^{\text{th}}$  observation.

To build  $\mathbf{X}_{i2}$ , we iterate through person  $i$ 's second subseries of length  $W$ , starting with their second overall observation and stopping with time series observation  $W+1$ . We continue building transition matrices, sliding the subseries window down the length of person  $i$ 's overall time series by one each time until observation  $L_i$  is captured in  $\mathbf{X}_{i\max(ji)}$ . Unlike traditional Markov models, our method can account for multiple states being endorsed simultaneously. Additionally, the windowed approach to the data allows for this method to account for non-stationarity inherent in many time

series derived from human behavioral data (Boker et al., 2002; Molenaar et al., 2003).

### Visual demonstration

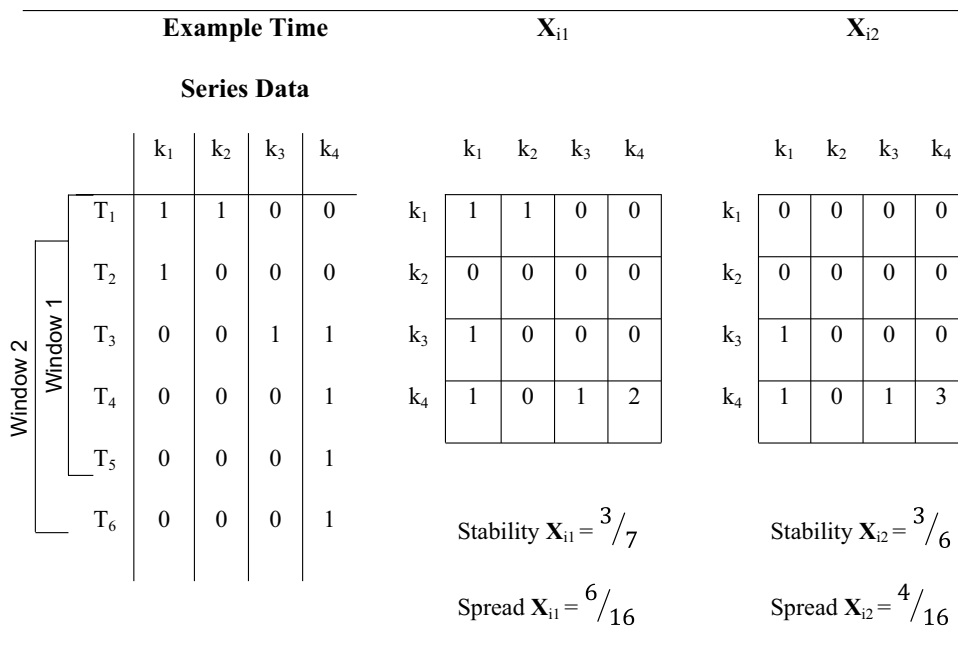
To demonstrate, we provide a verbal description of this process using a simple case that is accompanied by a visual representation in Fig. 1. Suppose a given individual  $i$  rated whether or not each of four different outcomes ( $k_1, k_2, k_3, k_4$ ) had occurred at six time points ( $T_1, T_2, T_3, T_4, T_5, T_6$ ). With these data, suppose we want to construct two transition matrices ( $\mathbf{X}_{i1}, \mathbf{X}_{i2}$ ), where each transition matrix contains data from five observations ( $W = 5$ ) within these multivariate binary time series data and the windowing lag is set to one.

To construct  $\mathbf{X}_{i1}$ , we would start by creating a  $4 \times 4$  matrix for which all elements are initialized to zero. Suppose the data show that  $k_1$  and  $k_2$  occurred at the first observation ( $T_1$ ) and  $k_1$  occurred again at the second observation ( $T_2$ ). This would suggest that a transition from  $k_1$  to  $k_1$  and a transition from  $k_2$  to  $k_1$  occurred between the first two time points. Given this pattern, we would increment the (1,1) element of  $\mathbf{X}_{i1}$  by one (to reflect the transition from  $k_1$  to  $k_1$ ) and we would increment the (1,2) element of  $\mathbf{X}_{i1}$  by one (to reflect the transition from  $k_2$  to  $k_1$ ). All other elements would remain at 0. Next, suppose  $k_3$  and  $k_4$  were both observed at  $T_3$ , indicating that a transition from  $k_1$  to  $k_3$  and a transition from  $k_1$  to  $k_4$  occurred between  $T_2$  and  $T_3$ . To account for these two transitions, we would increment the (3,1) element of  $\mathbf{X}_{i1}$  by one (to reflect the transition from  $k_1$  to  $k_3$ ) and we would increment the (4,1) element of  $\mathbf{X}_{i1}$  by one (to reflect the transition from  $k_1$  to  $k_4$ ). Next, suppose  $k_4$  was the only variable reported at  $T_4$ . This would indicate that a transition from  $k_3$  to  $k_4$  and a transition from  $k_4$  to  $k_4$  had occurred between  $T_3$  and  $T_4$ . In response, we would increment the (4,3) element of  $\mathbf{X}_{i1}$  by one (to reflect the transition from  $k_3$  to  $k_4$ ) and the (4,4) element of  $\mathbf{X}_{i1}$  by one (to reflect the transition from  $k_4$  to  $k_4$ ). Next, suppose  $k_4$  was the only variable reported at  $T_5$ , thereby indicating that a transition from  $k_4$  to  $k_4$  had occurred between  $T_4$  and  $T_5$ . In response, we would once again increment the (4,4) element of  $\mathbf{X}_{i1}$  by one, such that the (4,4) element now equals two. At this point, all transitions between the four binary variables across the first five time points are reflected in  $\mathbf{X}_{i1}$  (see Fig. 1).

To construct  $\mathbf{X}_{i2}$  we would start with a second  $4 \times 4$  matrix, also initialized to zero. The window of observations being read into  $\mathbf{X}_{i2}$  would be shifted down the time series by one compared to what was read into  $\mathbf{X}_{i1}$ , such that the transitions between  $T_1$  and  $T_2$  described above would not be captured by the new matrix. The transitions between  $T_2$  and  $T_3$ ,  $T_3$  and  $T_4$ , and  $T_4$  and  $T_5$ , however, would be incremented into the new matrix like in  $\mathbf{X}_{i1}$ . Finally, because the window of observations was shifted

Footnote 2 (continued)

the measured states were observed at successive timepoints within a given transition matrix, the stability calculation would return a "noUse" solution and spread would be 0.



*Note.* Two transition matrices constructed from example time series data with six observations ( $T_1$  through  $T_6$ ), window size of 5 observations per transition matrix, four binary time series ( $k_1$  through  $k_4$ ), and a windowing lag of one. We chose not to reduce the stability and spread fractions, when appropriate, to avoid obscuring the relationship between the transition matrices and the resulting stability and spread values.

Fig. 1 Visual demonstration of method

down one, there would be one new transition to add to  $\mathbf{X}_{i2}$  (i.e., the transition between  $T_5$  and  $T_6$ ). Suppose  $k_4$  was the only time series variable reported at  $T_6$ , thereby indicating that a transition from  $k_4$  to  $k_4$  had occurred between  $T_5$  and  $T_6$ . In response, we would once again increment the (4,4) element of  $\mathbf{X}_{i2}$  by one, such that the (4,4) element now equals three. At this point, all transitions between the four binary variables across the next five time points are reflected in  $\mathbf{X}_{i2}$  (see Fig. 1).

### Calculating stability

Stability is a proportion bounded between 0 and 1. It is defined as the trace of a transition matrix divided by the sum of all elements within that matrix, and thus is the proportion of transitions that are stable.

$$St_{ij} = \frac{tr(\mathbf{X}_{ij})}{\sum \sum \mathbf{X}_{ij}} \tag{2}$$

Here,  $tr(\mathbf{X}_{ij})$  is the sum of the elements along the diagonal of  $\mathbf{X}_{ij}$ ;  $\sum \sum \mathbf{X}_{ij}$  is the sum of all elements of  $\mathbf{X}_{ij}$ ; Stability is calculated for each  $\mathbf{X}_{ij}$  and is stored as a vector. An example of how two stability values are calculated from two example

transition matrices, each with  $4 \times 4$  dimensions and comprised of five time points, is provided in Fig. 1.

### Calculating spread

Spread is a proportion bounded between 0 and 1. It is defined as the number of all non-zero cells in a transition matrix divided by the number of all possible cells in that matrix.

$$Sp_{ij} = \frac{nz(\mathbf{X}_{ij})}{k^2} \tag{3}$$

$nz(\cdot)$  is a count of the number of non-zero elements in  $\cdot$ ;  $k^2$  is number of elements in  $\mathbf{X}_{ij}$ ; Spread is calculated for each  $\mathbf{X}_{ij}$  and is stored as a vector. An example of how two spread values is calculated from two example transition matrices, each with  $4 \times 4$  dimensions and comprised of five time points, is provided in Fig. 1.

### R package

We provide an R package on GitHub (<https://www.github.com/KatharineDaniel/transitionMetrics>) that includes

**Table 1** Selected simulation results when stability and spread indicators are randomly generated

Number of Timeseries	Number of Observations									
	25				100					
	Window Size				Window Size					
	0.02	0.05	0.1	0.2		0.02	0.05	0.1	0.2	
10	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	10	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	
	Stability	.110 (.08)	.110 (.07)	.111 (.07)		.110 (.05)	Stability	.112 (.08)	.112 (.05)	.111 (.04)
	Spread	.256 (.21)	.298 (.22)	.288 (.23)	.590 (.30)	Spread	.280 (.24)	.581 (.3)	.726 (.29)	.828 (.26)
30	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	30	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	
	Stability	.036 (.02)	.036 (.02)	.036 (.02)		.036 (.02)	Stability	.036 (.02)	.036 (.01)	.036 (.01)
	Spread	.290 (.20)	.272 (.20)	.273 (.20)	.579 (.30)	Spread	.290 (.22)	.647 (.29)	.634 (.30)	.872 (.20)

$N = 75$  for all cases

functions that transform binary time series data into transition matrices and then calculate stability and spread values per transition matrix per person. These functions allow researchers to specify their chosen  $W$  value and can operate on any number of time series variables or length of data.

### Simulation study

#### Method

To gain insight into the relationship between stability and spread and their reliability, we simulated multivariate binary time series data that varied according to set values along the following dimensions: Number of participants ( $N = \{20, 50, 75, \text{ or } 100\}$ ); number of variables included in the transition matrix ( $k = \{2, 10, 20, \text{ or } 30\}$ ); length of each person’s overall time series or the number of total observations per person ( $L = \{10, 25, 50, \text{ or } 100\}$ ); and the number of consecutive observations within a set of time series that contributes to a given matrix or window size ( $W = \{.02, .05, .1, .2 \text{ of } L\}$ ). Given that  $W$  is defined as a proportion of  $L$ , but by definition  $W$  must be a positive integer that is greater than or equal to 2, we constrained  $W$  to 2 if the percentage of  $L$  would have been below that lower bound. We set the windowing lag to one for all simulations.

We conducted 1000 runs for each possible combination of the above dimensions. Here we focus on results from simulation runs with randomly generated stability and spread values. However, we ran additional simulations with specific expected values of stability and spread ( $Stability = \{.01, .10, .25, .50, .75, .90\}$ ;  $Spread = \{.10, .25, .50, .75, .90, .99\}$ ) that are included in supplemental materials. Including those shown in the supplement, we ran 1728 different simulations taking approximately 3000 CPU hours on a high-performance computing cluster. For each set of simulated data, we calculated the mean and standard deviation of the resultant stability and spread values and calculated the correlation between the two stability and spread values.

### Results

Table 1 depicts how mean and standard deviation stability and spread values vary across differing  $W$  when:  $N = 75$ ,  $k = \{10, 30\}$ , and  $L = \{25, 100\}$ . Additional tables depicting how mean and standard deviation values vary across differing  $W$ ,  $k$ , and  $L$  when data were generated with different expected values for stability and spread (rather than having been randomly generated) are included in the supplement. We discuss general patterns observed within these simulations here, but provide all results as a  $4 \times 4 \times 4 \times 4 \times 5 \times 5 \times 3 \times 5$ -dimensional array in an R.data file on our OSF page (<https://www.osf.io/xqdk5/>).

#### Effect of $W$

Across all simulations, the number of observations that contribute to a given transition matrix,  $W$ , exerts a positive influence on the average spread value obtained across all  $X_{ij}$  while exerting little noticeable influence on the average stability value obtained across all  $X_{ij}$ . With increasing  $W$ , more observations are able to contribute to a given transition matrix. Greater observations afford greater opportunities to enter into new cells within the transition matrix, which necessarily increases spread values. The standard errors of the spread estimates do not appear to monotonically decrease as more observations are included until  $W = 5$ , which suggests that spread values calculated with fewer than five observations may not be trustworthy. Unlike spread, average stability values remain relatively unchanged due to the effect of taking an average across a sliding window. While the average stability values appear near-perfectly consistent in the large simulations run for the current study, within person variability in stability does occur across the sliding transition matrices (see Table 2 for a simplified example).

#### Effect of $k$

The number of variables that contribute to a given transition matrix,  $k$ , is functionally related to spread. Definitionally,

**Table 2** Depicting changes in stability and spread values calculated from the same timeseries data but with different window size

Timeseries where $k = 3$ and $L = 5$			Stability and Spread Values Derived from Transition Matrices Built on Same Timeseries but Differing $W$										
k1	k2	k3	k1		k2		k3						
1	1	0	W = 3										
0	0	1	k1	0	0	k3	1	k1	2	2	3		
1	1	1	k2	0	0	1	1	k2	1	1	2		
1	1	1	k3	1	1	1	1	k3	1	1	2		
1	0	0	stability	0.2		stability	0.3333		stability	0.3333		Average stability	0.2889
			spread	0.5556		spread	1		spread	1		Average spread	0.8519
			W = 4										
			k1	1	1	k3	2	k1	2	2	3		
			k2	1	1	2	2	k2	1	1	2		
			k3	2	2	2	2	k3	1	1	2		
			stability	0.2857		stability	0.3333					Average stability	0.3095
			spread	1		spread	1					Average spread	1
			W = 5										
			k1	2	2	k3	3						
			k2	1	1	2	2						
			k3	2	2	2	2						
			stability	0.2941								Average stability	0.2941
			spread	1								Average spread	1

$W$  window size,  $k$  number of timeseries,  $L$  number of observation in timeseries

spread is calculated with reference to the number of possible transition states (i.e.,  $k^2$  is the denominator). As such, variation in spread values is constrained by  $k$  such that, assuming sufficiently large  $W$ , the number of possible spread values for a given transition matrix is  $k^2+1$ . For example, the only possible spread values when  $k = 2$  are 0, .25, .50, .75, 1. Thus, as  $k$  increases, greater precision in spread between people and across transition matrices is possible. Whereas variance in spread is constrained by  $k$ , variance in stability is constrained by the number of observed transitions irrespective of  $k$ . That said, if a time series is randomly generated, maintaining a high stability value is less probable when there are more options available that would increment a transition matrix along its off diagonal.

**Effect of L**

The number of total observations within a timeseries,  $L$ , appears to be less influential on the mean stability and spread values calculated from these simulations than the  $W$  and  $k$  parameters. Indeed, in tables where  $L$  appears to increase along with spread values, it is important to note that it is increased  $W$ , rather than increased  $L$ , which explains these increases to spread ( $W$  is defined as a proportion of  $L$  and as such, when  $L$  increases, raw  $W$  increases in turn).

Key considerations for setting  $W$ ,  $k$ , and  $L$  are outlined in Table 3.

**Inverse relationship between stability and spread**

Plotting values of stability and spread against each other shows that while there is some overlap in these two metrics, stability and spread capture unique information about short-term switching behavior in multivariate time series data (see Fig. 2). The shapes of these plots show that stability and spread values have a moderate inverse association, such that

as a transition matrix is characterized by increasing levels of spread (i.e., more overall cells are populated within the transition matrix), stability values tend to decrease (i.e., more cells along the off-diagonal are populated). However, the curved banana-like shape suggests there is unique information captured by each metric. Further, these plots also show us that, assuming a random process, as stability approaches 1, spread necessarily converges to  $\frac{1}{k^2}$ . However, as spread approaches 1, stability converges to  $\frac{1}{k}$ .

**Interim discussion**

We varied the number of binary time series, number of observations in the sliding window, and length of the time series to explore the relationship between stability and spread and the effect of different parameters on stability and spread metrics. Simulation results found that stability and spread are moderately inversely correlated but capture unique information. Results also indicated that: (1) the number of observations that contribute to a transition matrix ( $W$ ) has a positive influence on average spread but little influence on average stability, (2) that the number of time series variables that contribute to a transition matrix ( $k$ ) has a probabilistically negative influence on average stability and mathematically constrains the number of possible spread values, and (3) the length of the overall time series ( $L$ ) has little effect on either average stability or spread. Notably, these metrics are based off the observed transition matrix, which implies that their statistical consistency is entirely dependent on the consistency of the observed data (i.e., the transition matrices and related metrics will accurately depict the transition behavior of the system if and only if the time series data that are fed into the transition matrices accurately capture the transitions within the system). As such, these metrics should be treated as sample statistics rather than parameter estimates.

**Table 3** Selecting parameter values

Parameter	Definition	Recommended Minimum	Effect on Average Stability	Effect on Average Spread	Consideration
W	Number of observations that contribute to a given transition matrix	5	Little influence	Positive influence	Avoid setting <i>W</i> to be significantly higher than <i>k</i> , as this would restrict variance in spread by increasing the likelihood to observe each state at least once within the transition matrix
k	Number of time series variables that contribute to a given transition matrix	4	Negative influence (assuming a random generating process)	Constrains the number of possible spread values for a given transition matrix ( $k^2+1$ )	To avoid overly sparse matrices, set larger <i>W</i> values if larger <i>k</i> or allow >1 time series to be endorsed simultaneously
L	Number of total observations within a timeseries	9*	Little influence	Little influence	The total number of transition matrices for individual <i>i</i> is given by $L - W + 1$ . Collect sufficiently long time series relative to chosen <i>W</i> to observe within-person change across transition matrices

\* = assuming a windowing lag of 1 and researcher interest in within-person differences in stability or spread

### Relationship between stability and spread

The current method calculates two inversely related measures that capture unique information about transitions in multivariate binary time series data. To elucidate the difference between stability and spread, consider these examples: a person who alternates between using cognitive reappraisal and suppression to regulate their emotions would receive the same stability score as a person who switches from cognitive reappraisal to distraction to acceptance, but the latter would receive a larger spread score than the former based on greater diversity in the specific strategies they used over time. Conversely, although two people who both used cognitive reappraisal and distraction as their only emotion regulation strategies would earn the same spread score, they could still earn different stability scores based on the order in which they reported using those two strategies (i.e., cognitive reappraisal to distraction to cognitive reappraisal to distraction is more unstable than distraction to distraction to distraction to cognitive reappraisal).

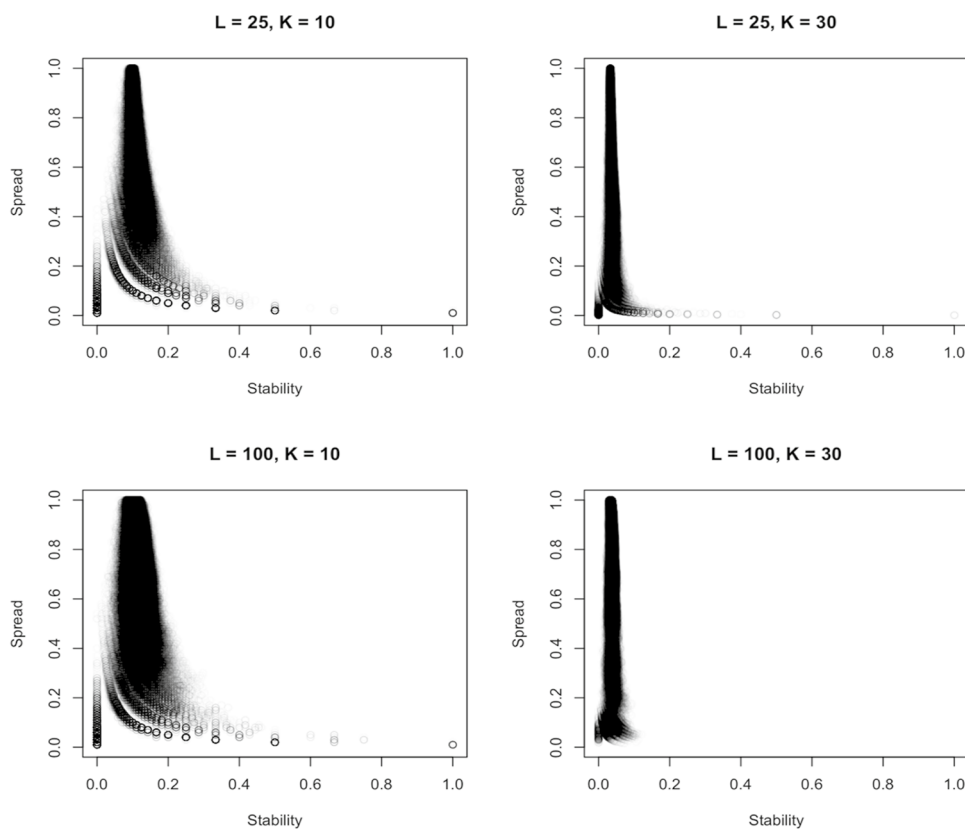
The differences in these metrics are not only mathematically distinct; they also capture theoretically interesting information. The degree of stability in one's emotion regulation strategy selections, for example, speaks to whether or not a person tends to rigidly employ the same strategy from one moment to the next (i.e., higher stability) or to vary their strategy use across time (i.e., lower stability). Given the presumed adaptiveness of flexible emotion regulation (Aldao

et al., 2015), some degree of instability (i.e., some shifting between strategies over time) is likely to be associated with positive emotional outcomes. However, complete instability may also indicate that a person is undiscerning and erratic in their attempts to regulate their emotions (Moulder et al., 2021). Separately, the greater number of unique emotion regulation strategy transitions that a person uses, the more "spread out" their observations will be across their transition matrix. This suggests that the relative spread of one's emotion regulation strategy selections speaks to the breadth of their strategy repertoire, which has been positively associated with psychological well-being (Rusch et al., 2012).

### Considerations for selecting parameter values

The hyperparameter *W* affects the stability and spread metrics. Simulation results suggest that researchers seeking to apply this method to their own data should refrain from setting a particularly small window size, given this would depress possible variance in spread values. For example, if  $W = 2$ , there is only one transition opportunity per matrix, making it challenging to observe between-person significant differences in spread. Window sizes smaller than 5 also do not evidence the expected relationship between increased observations and reduced standard errors, which further supports the importance of including at least five observations per transition matrix. However, researchers should also refrain from setting a very large window size relative to the number





*Note.* Stability and spread values, assuming a random process, generated when  $N = 75$  and number of observations ( $L$ ) is 25 or 100 and number of timeseries ( $k$ ) is either 10 or 30. Each dot is the result of one simulation.

**Fig. 2** Relationship between stability and spread

of variables contributing to their transition matrices, given this would inflate spread values such that there would also be a restriction in variance preventing meaningful statistical inference. For example, if  $k = 4$  and  $W = 100$ , most transition matrices would evidence a spread score of 1 simply because there are so many opportunities to observe each transition state at least once within the transition matrix. Researchers should also consider theoretical aspects of the process under investigation when setting  $W$ . For example, if a chef changes jobs from an Indian to a Mexican restaurant, thereby changing their pattern of typical spice use, a large  $W$  would obscure this change whereas a small-to-medium  $W$  may not. Therefore, we recommend that researchers set their  $W$  according to the theory within their substantive field and the number and length of their time series of interest. To allow for sufficient within-spread variance, however, we recommend that researchers set  $W$  to be greater than or equal to 5.

Stability and spread metrics will also be influenced by the number of binary variables that a researcher includes in their transition matrices. The number of elements in a given matrix increases by the square of the number of binary time series included. Two binary time series yields a  $2 \times 2$  matrix with four elements, three binary time series yield a  $3 \times 3$

matrix with nine elements, and so on. As a result, including a greater number of binary time series while holding  $W$  constant yields sparser matrices because there are more elements in the resultant matrix to fill despite there being no additional transitions reflected in the matrix. Specifically, spread values will be systematically lower in larger transition matrices (given that  $k^2$  is the denominator for spread) and stability values will be probabilistically lower assuming a random process, but the effect of  $k$  on stability is not mathematically constrained (given that  $k$  is not directly included in stability's equation). Although we recommend that researchers set their specific  $k$  according to the theory within their substantive field and the window they use per transition matrix, we recommend that researchers use four or more timeseries ( $k > 4$ ), thereby allowing sufficient variability in spread values. Four is the suggested minimum because when  $k = 4$ , there are 17 different possible spread values, which means that the possible variance in spread behaves more like a continuous variable.

Notably, because  $W$  and  $k$  can each influence stability and spread values, raw stability and spread values should not be compared across samples that use different  $W$  values and/or different numbers of binary time series. For this reason,

researchers should always report the parameter values they select.

Although  $L$  has little effect on either average stability or spread, there is a functional relationship between  $L$  and  $W$ : The total number of possible transition matrices that can be calculated for individual  $i$  is  $L - W + 1$ . Thus, researchers should collect sufficiently long time series relative to their chosen  $W$  to be able to observe within-person change across these transition matrices. Given  $W = 5$  and a windowing lag of one, we recommend a minimum of  $L = 9$  observations to allow for five different stability and spread values per person over time.

### Assumptions and boundary conditions of stability and spread

Although we recommend minimum values for  $W$ ,  $k$ , and  $L$ , the value of this method comes from its ability to function with high-dimensional, sparse, multivariate binary time series that have been prohibitively difficult to analyze with existing methods. We show that this method works when analyzing as many as 30 time series, the highest value for  $k$  included in this initial simulation study. To work with these high-dimensional time series, stability and spread values function as summary statistics. A trade-off when capturing information in high-dimensional data is that stability and spread are agnostic to the specific variables included in each row and column of a given transition matrix. Thus, while spread and stability summarize the diversity and order in which a person selects between a set of binary variables over time, respectively, they do not differentiate between a person who occupies one off-diagonal element from a person who occupies a different off-diagonal element. As such, multiple transition patterns could occur that result in the same stability or spread value. Therefore, stability and spread can be thought of as summary statistics that capture dynamic change patterns within a system over a given window of time and should be interpreted accordingly. These metrics are designed to be used when interest is at the level of (in) stability or spread within a system without strong interest in differentiating between a given level of stability due to one state's frequent endorsement over the same level of stability due to a different state's frequent endorsement. Future work should seek to identify ways to compare specific patterns of elements observed within transition matrices if which binary options are endorsed over time is of theoretical interest. Additional measures may also be used alongside stability and spread to enhance understanding of the system from different perspectives (e.g., recurrence rate on dimension-reduced data along with stability: Wallot, 2019; Shannon entropy along with spread: Rajaram et al., 2017).

Similarly, the metrics that we present are not exhaustive of all that could be taken from these transition matrices.

For example, our operationalization of spread does not capture the degree or weight of certain transitions over others. Rather, its values are affected by whether each possible transition occurred, not by the extent to which each transition type occurred relative to the others. Future method development work may seek to extend the current spread metric to a continuous spread metric. Eigenvalues-based methods, matrix rotations, or decompositions, among others, may offer useful additional approaches towards leveraging the range of information that can be learned from these transition matrices. Further, our operationalizations of stability and spread only consider observed-to-observed transitions, irrespective of other potentially interesting transitions afforded by the multivariate binary time series (i.e., observed-to-unobserved, unobserved-to-observed, unobserved-to-unobserved). Should researchers wish to contextualize observed-to-observed transitions relative to all classes of transitions, it would be interesting to create block matrices that reflect all four types of transitions and subsequently construct equations that leverage the desired information contained within.

Finally, this method does not account for the amount of real time that passes between successive observations. Similarly, the statistical consistency of stability and spread are entirely dependent on the consistency of the sample transition matrix. As such, researchers should take care to sample at a rate that best captures the underlying process of interest because sampling frequency may influence the validity of the stability and spread values that are derived. Because this method does not account for elapsed time, and given sampling frequency may influence derived values, this method may be best suited to repeated-measures data that are collected with equal time intervals. However, given that we use overlapping sliding windows and time delay embedding has been shown to be robust to sampling interval misspecification (Boker et al., 2018), it is likely that equal interval measurement is not a necessary condition.

### Comparison against recurrence quantification analysis

As an initial comparison of stability and spread against common metrics from RQA, we simulated 1000 binary time series data sets with a random generating process where each data set was defined with  $N = 100$ ,  $L = 100$ , and  $k = 10$ . For each simulated participant in each data set, we calculated stability and spread with  $W = 20$  and a lag of 1, which is consistent with one of the simulation sets described above. We averaged the stability and spread scores from across the windows within a given simulation set to arrive at one stability and one spread score per simulated participant to reflect the system's transition behavior. On those same data we also calculated recurrence rate, determinism, and entropy

**Table 4** Correlations between average recurrence rate, determinism, entropy, stability, and spread values across 1000 simulated data sets of  $N = 100$ ,  $k = 10$ , and  $L = 100$ 

Variable	<i>M</i>	<i>SD</i>	1	2	3	4
1. Recurrence Rate	49.49	7.56				
2. Determinism	80.72	6.54	0.97 [0.97, 0.97]			
3. Entropy	1.67	0.30	0.98 [0.98, 0.98]	0.96 [0.96, 0.96]		
4. Stability	0.11	0.01	-0.08 [-0.09, -0.08]	-0.11 [-0.12, -0.11]	-0.07 [-0.08, -0.07]	
5. Spread	0.87	0.20	-0.49 [-0.49, -0.49]	-0.48 [-0.48, -0.48]	-0.48 [-0.48, -0.48]	-0.34 [-0.35, -0.34]

For each simulated data set,  $N$  number of participants,  $k$  number of binary timeseries, and  $L$  number of observations in each time series.  $M$  and  $SD$  represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. All correlations were significant at  $p < .001$

between each unique bivariate time series combination using the `crqa` package in R (Coco et al., 2022). We then averaged all pairwise recurrence rate, determinism, and entropy values for a given simulation set, respectively, to arrive at the average recurrence rate, determinism, and entropy scores for a given simulated participant for that multivariate system. Next, we used the `psych` package in R (Revelle, 2022) to calculate the correlations between each of these five metrics using all simulated data. After inspecting the correlation plots between stability and the RQA metrics, we decided to remove observations where stability was greater than 0.4 ( $n = 14$ ) because these behaved as outliers. See Table 4 for results of the correlation values after outlier removal. The simulation results suggest that stability and spread provide different information about the transition behavior of time series data than does RQA.

### Applied example: Emotion regulation

Although the study of emotion regulation (ER) initially focused on the effect of a single or a few strategies in isolation (see Ford et al., 2019, for review), there are myriad ways to influence emotions (e.g., Heiy & Cheavens, 2014) and people can use more than one strategy in response to the same emotional event (Ford et al., 2019). Importantly, using a limited number of ER strategies has been associated with impaired psychological well-being (e.g., Rusch et al., 2012) and greater levels of depression, trait anxiety, and social anxiety (Lougheed & Hollenstein, 2012). As such, researchers are increasingly recognizing the need to sample a wide range of ER strategies.

The risk to psychological well-being that is associated with a limited repertoire of ER strategies may be driven by context-insensitive, rigid application of a few strategies over time (Aldao et al., 2015), instead of selecting from numerous strategies to optimally match a given situation (Bonanno & Burton, 2013). Indeed, Cheng (2001) found that those who reported higher variability in problem-focused

and emotion-focused coping during a series of in-lab tasks endorsed lower levels of depression. More recently, Birk and Bonanno (2016) found that the ability to switch from a suboptimal strategy to an optimal strategy during an in-lab study was associated with higher life satisfaction. Similarly, variably choosing between different ER strategies within a given situation in a person's daily life was associated with reduced negative affect (Blanke et al., 2019) and people with major depressive disorder reported lower trait ER diversity across adaptive ER strategies than people without a history of depression (Wen et al., 2021). As such, ER variability, or the tendency to use different strategies across different contexts, is theorized to be adaptive insofar as it supports flexible ER (Aldao et al., 2015).

However, a lack of existing methods to dynamically quantify patterns in ER transitions over time, especially when a large range of strategies are assessed via binary response scales, has hindered robust empirical tests of theories that tout flexible ER as adaptive. As an initial step to address this gap, this study constructs transition matrices on 4 weeks of ecological momentary assessment (EMA) ER data to investigate the dynamic interplay between ER strategy transition patterns and affective experiences in the daily lives of socially anxious people.

### Overview and hypotheses

To better characterize the association between patterns in ER strategy selections over time with how affect changes over time, this study investigates the order of transitions that people high in trait social anxiety symptoms make, or fail to make, between 20 different ER strategies (i.e., their strategy switches). The sample is composed of individuals scoring relatively high on a measure of trait social anxiety symptom severity given previous research showing that socially anxious people have deficits in ER (Jazaieri et al., 2014) and have lower intensity and less lasting positive emotions than healthy, non-anxious people (Kashdan et al., 2011).

The current study aims to test whether including information about how people switch between ER strategies over time predicts subsequent in-the-moment, self-reported affect and anxiety beyond previous timepoint's affect and anxiety ratings. We approach this question by including stability and spread metrics into a large cross-lagged panel model and testing for path significance coming from stability and spread at one timepoint onto next-timepoint affect and anxiety. We hypothesize that an individual's stability and spread metrics will predict their next-timepoint affect and anxiety ratings, although we do not have specific hypotheses for the direction of the effects given the novel methodological approach. Thus, these analyses are primarily exploratory and are conducted as a means of furthering future hypothesis-driven work that attempts to employ similar methods to answer questions about affect and ER dynamics.

## Method

### Participants

One hundred and 14 individuals scoring relatively high on a measure of trait social anxiety symptom severity enrolled in the 5-week EMA study. Participants were eligible for the study if they scored at least a 29 on the Social Interaction Anxiety Scale (SIAS; Mattick & Clarke, 1998). Participants also had to own an Android or iPhone that was compatible with Metricwire (the EMA mobile phone sampling application used in the study). Given that we were interested in the effect of recent strategy switching behavior on in-the-moment affect and anxiety, we retained only those participants who contributed at least six survey responses, leaving a final sample of  $n = 110$  participants. Participants in the final sample were 73.64% female and 20.44 years old on average ( $SD = 2.96$  years old). The self-reported racial composition of the final sample was 69.09% white, 6.36% African American/Black, 15.45% Asian, 1.82% Middle Eastern, and 7.27% multiple races. The self-reported ethnicity of the final sample was 2.73% Hispanic and 97.27% not Hispanic.

### Study procedure

The University of Virginia's ethics review board approved all study procedures. Participants provided written, informed consent to participate in two, 1.5-h in-lab sessions separated by 5 weeks of EMA surveys on their personal smartphone. As part of a larger study, approximately half of the participants were randomized to receive an online cognitive bias training intervention designed to reduce anxious thinking half-way through the study period (i.e., during week 3; see Daniel et al.,

2020).<sup>3</sup> Within the EMA portion of the study, participants received up to six randomly timed surveys per day (although participants in the intervention group only received two surveys per day during week 3 to reduce participant burden), one end-of-day survey, and one end-of-week survey for 5 weeks. MetricWire delivered randomly timed surveys at a random time between each 2-h window from 9 am to 9 pm. Surveys were designed to take less than 2 min to complete and to remain active for no more than 45 min. Consistent with our pre-registered plans for analyses (<https://www.osf.io/xqdk5/>), the current study only uses data from the first 4 weeks of randomly timed surveys. A full list of the in-lab, randomly timed EMA, end-of-day EMA, and end-of-week EMA procedures and measures can be obtained by contacting the first author.

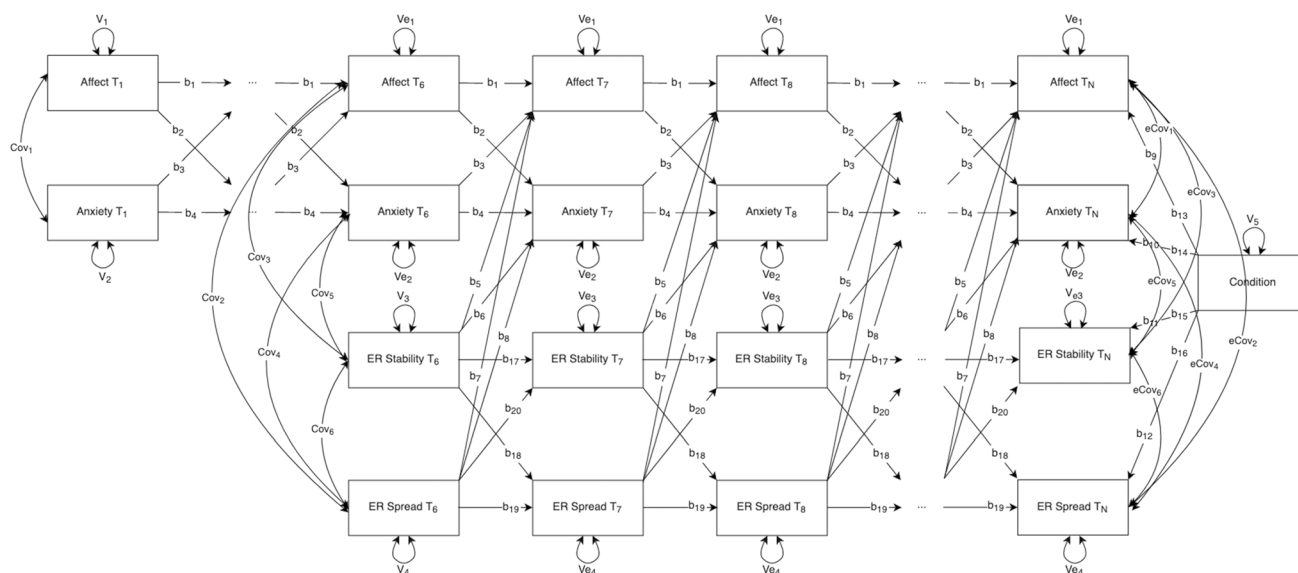
## Measures

**In-the-moment anxiety** At each randomly timed survey, participants rated their momentary anxiety using the single item, "Right now, I am feeling...", with anchors ranging from 1 (*very calm*) to 10 (*very anxious*).

**In-the-moment affect** At each randomly timed survey, participants rated their momentary affect using the single item, "Right now, I am feeling...", with anchors ranging from 1 (*very negative*) to 10 (*very positive*). We reverse-scored this item so that the interpretation of the direction of scores would be consistent with the in-the-moment anxiety item (i.e., higher scores indicate more distress on both items).

**In-the-moment emotion regulation** At each randomly timed survey, participants reported their momentary ER strategy attempts throughout the 30 min before the survey prompt. Participants could either report that they did not attempt to change their thoughts or feelings, or they could select from 19 unique strategies that were displayed using a check-all-that-apply list. Conceptual labels are provided for each strategy here, but participants saw lay-person descriptions of those strategies: rumination, problem solving, acceptance, self-criticism, cognitive reappraisal, thinking good thoughts, thought suppression, tackling the issue head on, alcohol, drugs, eating, exercising, TV/gaming, sleeping, advice-seeking, situational avoidance, expression suppression, doing something fun with others. See Daniel et al. (2020) for lay-person descriptions of strategies. Participants were not limited in the number of strategies they could select at each survey and each endorsed strategy was coded as a 1 (vs. 0). Participants could also report that they were not regulating their emotions at all.

<sup>3</sup> We discuss our approach to accounting for intervention condition in greater detail in the Analytic Approach section.



*Note.* ER = Emotion Regulation. Covariances between all variables at initial and final time points are included in the diagram to improve legibility, but these covariance paths were included at all time points in the model estimation. The paths from Condition to all other variables were constrained to zero within the first two weeks, freely estimated within the third week, and re-estimated within the fourth week. Parallel paths from the condition variable are constrained to be equal within the third week and are again constrained to be equal within the fourth week.

**Fig. 3** Complex model path diagram

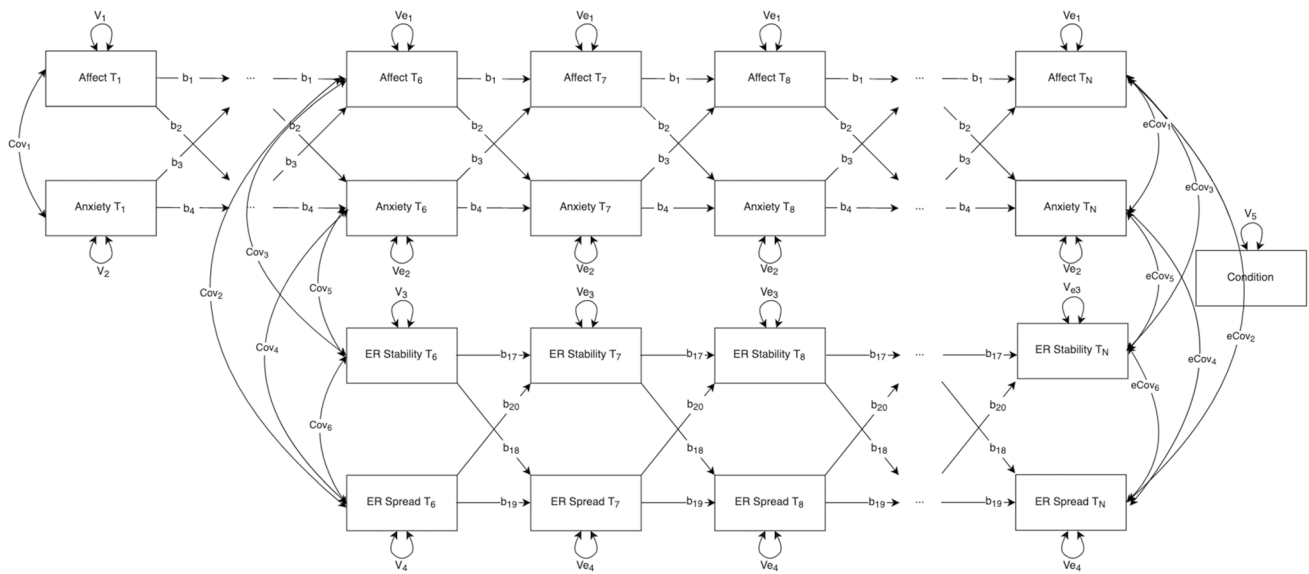
### Calculating stability and spread in ER transitions in daily life

We constructed transition matrices using the ‘buildTransArray’ function in the *TransitionMetrics* package on GitHub (Daniel & Moulder, 2020). The dimension of each transition matrix was 20-by-20. Six observations were included in each matrix and the windowing lag was set to one. We set  $W = 6$  because ER strategy use was sampled six times daily and we were interested in short-term ER dynamics. We pre-registered these decisions. We calculated stability and spread using the ‘transStats’ function in the same package. A stability value of 1 indicates rigid repetition of the same strategy over time and a value of 0 indicates no repetition of the same strategy across consecutive surveys. Spread values closer to 1 indicate more unique transitions between strategies and values closer to 0 indicate fewer unique transitions between strategies. To aid model convergence, we then rescaled the stability and spread by multiplying their values by 100 so their variances were closer in magnitude to those of the ten-point affect and anxiety variables.

### Analytic approach

Consistent with our pre-registered plans for analysis (<https://www.osf.io/xqdk5/>), we constructed a series of nested

cross-lagged panel models using OpenMX version 2.20.6 (Neale et al., 2016) in R version 4.1.3 (R Core Team, 2022) on the first 4 weeks of data from all participants. After transforming the data into wide format, we built a complex model for this data that included all paths to represent how stability and spread might be influencing affect and anxiety over time. The complex model auto-regressed affect, anxiety, stability, and spread, regressed affect and anxiety onto each other from  $T_t$  to  $T_{t+1}$ , regressed stability and spread onto each other from  $T_t$  to  $T_{t+1}$ , and regressed stability and spread onto the affect and anxiety variables from  $T_t$  to  $T_{t+1}$  (see Fig. 3). All autoregressive paths per variable, all cross-regressive paths between variables, and all intercepts per variable were constrained to be equal across time. Note that, although negative affect and anxiety are positively correlated in the present data ( $r = .52, p < .001$ ), they are not perfectly correlated and, as such, capture distinct yet related constructs. Further, note that while some authors call autoregressive paths “stability coefficients,” we reserve the word “stability” to mean the stability of the multivariate binary ER choices. We also modeled as a covariate the study condition to which each participant was randomly assigned. Specifically, this covariate was coded as “0” for participants who were not assigned to the intervention condition and “1” for participants who were assigned to the intervention condition. Paths going from this condition covariate to the affect, anxiety, stability, and spread variables were fixed to



*Note.* ER = Emotion Regulation. Covariances between all variables at initial and final time points are included in the diagram to improve legibility, but these covariance paths were included at all time points in the model estimation. Each set of covariances were held constant across all time points.

**Fig. 4** Constrained model path diagram

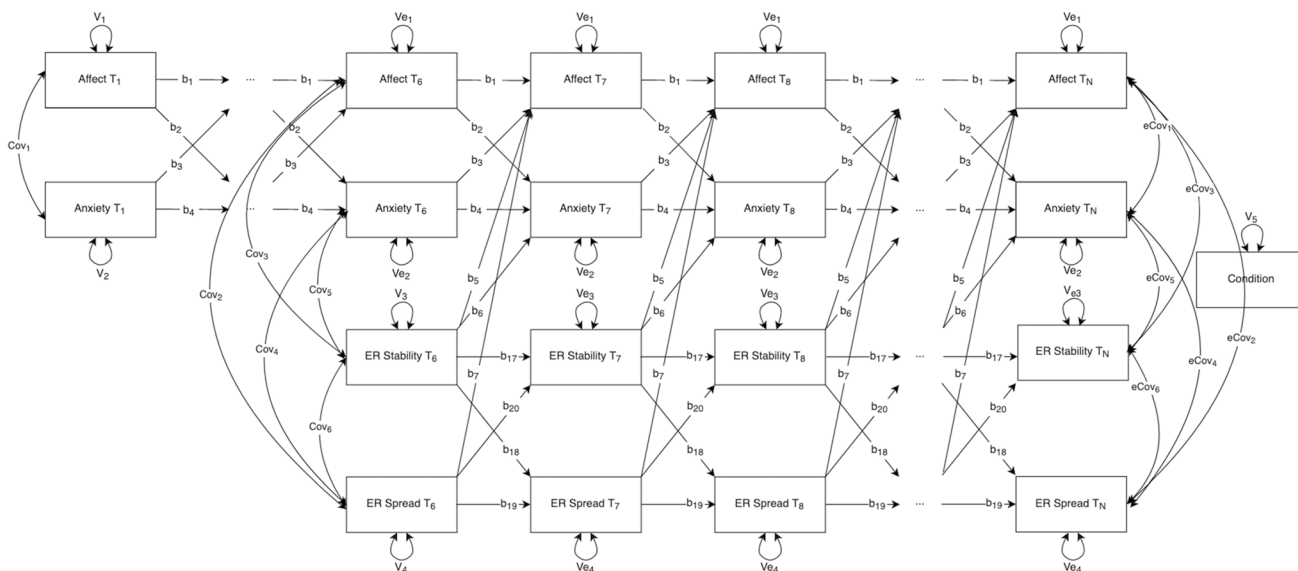
zero for all participants for all time points within the first 2 weeks of the study (i.e., prior to the intervention starting in week 3). These four paths were freely estimated for all time points within week 3 (i.e., estimating the effect of the treatment during the intervention period) and were re-estimated during the fourth week (i.e., estimating a lagged treatment effect). Paths from the condition covariate were constrained to be equal within each week.

Prior to testing our hypothesis that including paths from stability and spread at one timepoint to affect and anxiety at the next timepoint would improve model fit, we first reduced model complexity by testing if we could remove paths from the condition covariate. First, we constrained all paths from the condition variable to be zero and compared this condition-reduced model to the complex model depicted in Fig. 3. A likelihood ratio test found that all paths from the condition variable could be removed without worsening fit ( $\chi^2(8) = 5.41, p = .713$ ), indicating that the effect of condition could be entirely removed from the model. Thus, models that test the effect of stability and spread on next timepoint affect and anxiety did not estimate any regression paths coming from the condition covariate.

To test our hypothesis that including information about a person’s recent ER strategy switches improves prediction of affect and anxiety ratings, we constrained the paths from stability and spread at one timepoint to affect and anxiety at the next timepoint to be zero (see Fig. 4). We then again compared the constrained model to the complex model using a likelihood ratio test which resulted in significantly worse model fit ( $\chi^2(12) = 115.41, p < .001$ ), indicating

that the effect of stability and spread on next timepoint affect and anxiety should not be entirely removed from the model.

Given that we observed a significant difference in fit between the complex (Fig. 3) and constrained models (Fig. 4), we tested a series of intermediary models to better understand the effect of stability and spread within the system. One at a time, we constrained each path from stability or spread to either next timepoint affect or next timepoint anxiety to be zero and compared it to the complex model (Fig. 3). The order we took to constrain the paths to zero was selected based on the relative strength of each effect’s estimate found by the complex model (i.e., we first removed the path of the weakest effect, then the path of the second weakest effect, etc.). This iterative process concluded that the paths from one timepoint stability to next timepoint anxiety ( $\chi^2(9) = 51.55, p < .001$ ) and from one timepoint stability to next timepoint affect ( $\chi^2(9) = 81.24, p < .001$ ) were both meaningful and should be retained in the final model. However, the path from one timepoint spread to next timepoint anxiety ( $\chi^2(9) = 6.18, p = .722$ ) could be removed without harming model fit, but the path from spread to next timepoint affect should not also be removed ( $\chi^2(9) = 25.41, p = .005$ ). As a result, the final model depicted in Fig. 5 explains the data comparably to the complex reference model ( $\chi^2(9) = 6.18, p = .722$ ) that estimated all possible paths within the system. Interpretation of beta values from the final model are provided below. Raw data and the R scripts used for the current paper are openly available on the Open Science Framework (<https://www.osf.io/xqdk5/>).



Note. ER = Emotion Regulation. Covariances between all variables at initial and final time points are included in the diagram to improve legibility, but these covariance paths were included at all time points in the model estimation.

Fig. 5 Final model path diagram

Results

The direction of the estimates for the autoregressive paths indicates that, on average and across the full sample, previous timepoint scores for affect, anxiety, stability, and spread were positively associated with next timepoint scores for those same variables. The cross-lagged paths between affect

and anxiety were also positively associated, such that greater anxiety at one timepoint was associated with greater negativity at the next timepoint, and vice versa. The cross-lagged paths between stability and spread were negatively associated, such that greater stability at one timepoint was associated with less spread at the next timepoint, and vice versa. See Table 5. These regression paths were not tested for path significance.

Table 5 Unstandardized regression weights for single-headed arrows from the final model

Path	Estimate	Standard Error
Affect Autoregressive Effect (b1)	0.462	0.010
Affect to Anxiety Cross-Lagged Effect (b2)	0.064	0.011
Anxiety to Affect Cross-Lagged Effect (b3)	0.046	0.010
Anxiety Autoregressive Effect (b4)	0.412	0.010
Stability to Affect Cross-Lagged Effect (b5)	-0.005	0.001
Stability to Anxiety Cross-Lagged Effect (b6)	-0.005	0.001
Spread to Affect Cross-Lagged Effect (b7)	-0.036	0.008
Stability Autoregressive Effect (b17)	0.933	0.004
Stability to Spread Cross-Lagged Effect (b18)	-0.001	0.0002
Spread Autoregressive Effect (b19)	0.944	0.003
Spread to Stability Cross-Lagged Effect (b20)	-0.334	0.056

Likelihood confidence intervals (CIs) did not converge, indicating that 1.96\*standard error may not reflect an accurate 95% CI. Refer to Model 5 for a visual depiction of the final model

Including information about strategy switching (measured as stability and spread) improved model fit when predicting an individual’s subsequent in-the-moment affect and anxiety ratings. Specifically, instances of greater stability in ER reports were associated with less anxiety and less negative affect at the next timepoint, even after accounting for the autoregressive and cross-lagged effects of affect and anxiety. Further, instances of greater spread in ER reports were associated with less negative affect at the next time point, even after accounting for the autoregressive and cross-lagged effects of affect and anxiety. The effect of spread onto next time point anxiety was not significant, so it was removed from the final model (see Tables 5 and 6).

Interim discussion

The current study applies our approach to quantifying transitions within multivariate binary time series data to 4 weeks of ER EMA data collected from a sample of 110 socially anxious people. Results of this study demonstrate that ER stability and spread constructs calculated from transition matrices within high-dimensional time series data significantly predict next timepoint affect and/or anxiety, even after

**Table 6** Unstandardized estimates for variance, covariance, and mean score values from the final model

	Estimate	Standard Error
Variance of Affect (V1)	3.46	0.50*
Error Variance of Affect (Ve1)	3.83	0.05
Variance of Anxiety (V2)	4.23	0.64*
Error Variance of Anxiety (Ve2)	4.47	0.06
Variance of Stability (V3)	924.15	125.87*
Error Variance of Stability (Ve3)	149.83	2.12
Variance of Spread (V4)	4.96	0.67*
Error Variance of Spread (Ve4)	0.56	0.01
Variance of Condition (V5)	0.25	0.03
Covariance between Anxiety and Affect (Cov <sub>1</sub> )	1.77	0.47
Covariance between Anxiety and Stability (Cov <sub>5</sub> )	-1.43	3.18*
Covariance between Anxiety and Spread (Cov <sub>4</sub> )	-0.21	0.40*
Covariance between Affect and Stability (Cov <sub>3</sub> )	1.34	1.65*
Covariance between Affect and Spread (Cov <sub>2</sub> )	0.08	0.02
Covariance between Stability and Spread (Cov <sub>6</sub> )	-36.96	7.43*
Residual Variance between Stability and Spread (eCov <sub>6</sub> )	-3.08	0.01
Residual Variance between Anxiety and Stability (eCov <sub>5</sub> )	-1.41	0.26
Residual Variance between Affect and Stability (eCov <sub>3</sub> )	-2.12	0.25
Residual Variance between Affect and Anxiety (eCov <sub>1</sub> )	1.90	0.04
Residual Variance between Affect and Spread (eCov <sub>2</sub> )	0.08	0.02
Residual Variance between Anxiety and Spread (eCov <sub>4</sub> )	0.08	0.02
Mean of Condition	0.51	0.05
Mean of Anxiety	4.57	0.20
Mean of Affect	4.82	0.18
Mean of Stability	35.02	3.02*
Mean of Spread	2.22	0.22
Intercept of Anxiety	2.36	0.06
Intercept of Affect	2.55	0.06
Intercept of Stability	3.91	0.27
Intercept of Spread	0.14	0.02

\* = The gradient appears to be asymmetric so the standard error may not accurately reflect the variability of this estimate. Covariances, means, and intercepts were not drawn in the path diagrams for legibility, but their estimates are provided for full information. Refer to Model 5 for a visual depiction of the final model

accounting for the autoregressive and cross-lagged effects of previous timepoint affect and anxiety. Specifically, this exploratory cross-lagged panel model analysis showed that greater stability in ER behavior was associated with less anxiety and less negative affect at the next timepoint and greater spread in ER behavior was associated with less negative affect, but not less anxiety, at the next timepoint.

This work is exploratory and should be replicated to prevent overinterpretation. However, that stability and spread were each significant in the prediction of next timepoint affect and/or anxiety, even after accounting for the autoregressive and cross-lagged paths between one timepoint affect and anxiety and next timepoint affect and anxiety, demonstrates a promising new method for high-dimensional, complex datasets like the one used for this example. These findings suggest that the sequenced frequency and diversity in

switches that people make in their ER strategies are associated with affect and anxiety dynamics in daily life.

### Stability in ER

There is a robust literature to support the importance of flexible (i.e., not rigidly stable) ER strategy use over time (Aldao et al., 2015; Bonanno & Burton, 2013). However, flexibility does not necessitate constant changing over time. In fact, it is likely that there is some degree of optimal flexibility, such that both too much stability and too much instability in ER behaviors is maladaptive (Southward et al., 2018). In the current data, we observed a negative association between stability and affect and anxiety, such that greater stability in ER behavior was associated with less anxiety and less negative affect at the next time point. It could be that we



observed this pattern of results because, when there is a relative degree of stability in a person's strategy choices over a few consecutive survey responses, the person may have found a strategy or set of strategies that is effective for them during that period. Importantly, because stability is repeatedly calculated over time, this result does not imply that people who are more stable in their strategy selections across the entire study or throughout their entire life necessarily have better emotional outcomes. Rather, we found evidence to suggest that continuing with a strategy selection pattern within a segment of six surveys (i.e., in the short-term) is associated with better affect and anxiety outcomes at the next survey than is rapidly switching from one strategy to another strategy. Following Bonanno and Burton (2013) and Southward et al. (2018), it is possible that some amount of stability within active ER attempts is emotionally beneficial such that people can reap the emotional rewards of short-term continued use of a given strategy, especially if the situational demands do not change dramatically over the period that the surveys span.

### Spread in ER

There is a robust literature to support a positive association between size of ER strategy repertoire and long-term psychological well-being (e.g., Rusch et al., 2012). However, little attention has been paid to the importance of diversity in strategy transitions to affective dynamics (cf., Eldesouky & English, 2021). In the current data, we observed a negative association between one timepoint spread and next timepoint negative affect, such that greater spread in ER behavior was associated with less negative affect at the next timepoint. Greater spread might be associated with less negative affect at the next timepoint because a larger repertoire of strategies employed in the short term could be beneficial, following Rusch et al. (2012). Further, since participants could select multiple strategies at a time, participants who engage in emotion polyregulation earn greater spread scores and increased between-strategy variability has been associated with lower negative affect levels (Blanke et al., 2019). However, it is interesting to note that spread was not associated with next timepoint anxiety. This may be because some participants who identify as anxious might notice that they have used many ER strategies and therefore assume that must mean they are anxious. Other participants who do not identify as particularly anxious may not jump to the same conclusion. Thus, in the aggregate, the effect may be null.

### Limitations of applied ER example

The high dimensionality of these data and theoretical interest in ER switching patterns among a sample likely experiencing emotion dysregulation make this dataset a useful

candidate in which to apply this method. However, like many datasets collected through EMA, the amount of time between observations varied within and between people. There are open questions as to whether inconsistent sampling rates introduce bias into the outcomes of this method. However, given the sliding window approach we took, there is reason to expect bias from time misspecification may be minimal (Boker et al., 2018). Relatedly, although  $W$  was set to 6 because ER was measured six times daily, imperfect compliance resulted in transition matrices that conceptually did not always reflect within-day stability and spread in ER choices. Improved compliance through stronger incentive schemes would be helpful. Researchers could also choose to prevent the windowing procedure from combining data from multiple days, for example, if that would offer greater theoretical justification. However, we decided to allow multiple days' data to contribute to a given transition matrix to increase the number of observations and because our hypotheses did not depend on stability and spread being resolved each day.

Similarly, the validity of the stability and spread values solved for in this study depend on the statistical consistency of the data that were collected. Although this is ultimately unknown, given the data are irreducible and aperiodic and that the stability and spread sample statistics returned significant results that were consistent with theory, we view this application as useful.

Further, given the complexity of the cross-lagged panel models that we used in the current analyses, we constrained all auto-regressive and cross-regressive paths to be equivalent over time. Although this reduced the estimated parameters in the model by over 700, future models that allow for the association between stability and spread at one timepoint with next timepoint affect and anxiety to change over time would enable more nuanced relationships to be uncovered. Models that better leverage the repeated measures nature of stability and spread values per person to ask dynamic and developmental questions will be all the more exciting. That said, the analytic approach we took in this initial application still captures greater detail in the ways that people switch between ER strategies in their daily life than has been possible to date.

### Overall conclusions

We presented a novel method for quantifying transitions within high-dimensional multivariate binary time series by constructing transition matrices to derive metrics of stability and spread. We define stability as the trace of a transition matrix divided by the sum of all elements within that matrix. We define spread as the number of all non-zero cells in a transition matrix divided by the number of all possible

cells in that matrix. Simulation results show that stability and spread are inversely related but unique metrics, and the simulation results point to recommended guidelines for setting mathematically and theoretically principled minimum parameter values for: i) the number of observations to be included in a given transition matrix ( $W \geq 5$ ); ii) the number of timeseries variables to be included in a given transition matrix ( $k \geq 4$ ); and iii) the minimum length of the overall timeseries data, assuming a windowing lag of one and interest in time-varying stability and spread questions ( $L \geq 9$ ).

As an applied example, stability and spread were calculated for a sample of socially anxious people. This complex dataset is ideally suited to the proposed method because participants repeatedly reported their use of 20 different ER strategies and individual differences in how people transition between strategies over time is expected to be meaningful for psychosocial functioning. Indeed, we found that stability and spread in short-term ER are unique constructs that each predict next timepoint affect and stability also predicted next timepoint anxiety, even after accounting for auto-regressive and cross-lagged effects of affect and anxiety. This suggests that transition matrices can be used to extract meaningful quantifications of how switches in ER strategies are associated with affect and anxiety dynamics in daily life.

**Acknowledgments** We would like to thank the action editor and two reviewers for their helpful feedback.

**Funding** This work was supported by the National Institute of Mental Health under Grant R01MH113752 awarded to BAT; University of Virginia Hobby Postdoctoral and Predoctoral Fellowship Grant; and University of Virginia Jefferson Fellowship Grant awarded to KED.

**Data availability** R scripts used to simulate the data are available on the Open Science Framework (<https://www.osf.io/xqdk5/>), as are the overall descriptives for the simulated data. Due to storage limitations, data for individual simulation runs are available upon request to the corresponding author. ER data and associated R scripts are available on the Open Science Framework (<https://www.osf.io/xqdk5/>).

## Declarations

**Disclosure of interest** The authors report no conflicts of interest.

## References

- Aldao, A., Sheppes, G., & Gross, J. J. (2015). Emotion regulation flexibility. *Cognitive Therapy and Research*. <https://doi.org/10.1007/s10608-014-9662-4>
- Anderson, T. W., & Goodman, L. A. (1957). Statistical inference about Markov chain. *Annals of Mathematical Statistics*, 28, 89–110.
- Birk, J. L., & Bonanno, G. A. (2016). When to throw the switch: The adaptiveness of modifying emotion regulation strategies based on affective and physiological feedback. *Emotion*, 16(5), 657–670. <https://doi.org/10.1037/emo0000157>
- Blanke, E. S., Brose, A., Kalokerinos, E. K., Erbas, Y., Riediger, M., & Kuppens, P. (2019). Mix it to fix it: Emotion regulation variability in daily life. *Emotion*. <https://doi.org/10.1037/emo0000566>
- Boker, S. M., Rotondo, J. L., Xu, M., & King, K. (2002). Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychological Methods*, 7(3), 338. <https://doi.org/10.1037/1082-989x.7.3.338>
- Boker, S. M., Tiberio, S. S., & Moulder, R. G. (2018). Robustness of time delay embedding to sampling interval misspecification. In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous Time Modeling in the Behavioral and Related Sciences*. Springer. [https://doi.org/10.1007/978-3-319-77219-6\\_10](https://doi.org/10.1007/978-3-319-77219-6_10)
- Bonanno, G. A., & Burton, C. L. (2013). Regulatory flexibility: An individual differences perspective on coping and emotion regulation. *Perspectives on Psychological Science*. <https://doi.org/10.1177/1745691613504116>
- Cheng, C. (2001). Assessing coping flexibility in real-life and laboratory settings: a multimethod approach. *Journal of Personality and Social Psychology*, 80(5), 814–833. <https://doi.org/10.1037/0022-3514.80.5.814>
- Coco, M.I., Monster, D., Leonardi, G., Dale, R., Wallot, S., Dixon, J.D., Nash, J.C. (2022). Recurrence Quantification Analysis for Categorical and Continuous Time-Series. Package version 2.0.3. <https://www.cran.r-project.org/web/packages/crqa/crqa.pdf>
- Daniel, K.E., & Moulder, R.G. (2020). transitionMetrics: R package to calculate characteristics of transitions in multivariate binary time series data. Version 0.1.0. <https://www.github.com/KatharineDaniel/transitionMetrics>
- Daniel, K. E., Daros, A. R., Beltzer, M. L., Boukhechba, M., Barnes, L. E., & Teachman, B. A. (2020). How anxious are you right now? Using ecological momentary assessment to evaluate the effects of cognitive bias modification for social threat interpretations. *Cognitive Therapy and Research*, 44(3), 538–556. <https://doi.org/10.1007/s10608-020-10088-2>
- Eckmann, J. P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *EPL*, 4, 973. <https://doi.org/10.1209/0295-5075/4/9/004>
- Eldesouky, L., & English, T. (2021). Always in flux: The role of attentional focus in emotion regulation dynamics. *Cognition & Emotion*, 1–7. Advance online publication. <https://doi.org/10.1080/02699931.2021.1993146>
- Ford, B. Q., Gross, J. J., & Gruber, J. (2019). Broadening our field of view: The role of emotion polyregulation. *Emotion Review*, 11(3), 197–208. <https://doi.org/10.1177/1754073919850314>
- Heiy, J. E., & Cheavens, J. S. (2014). Back to basics: a naturalistic assessment of the experience and regulation of emotion. *Emotion (Washington, D.C.)*. <https://doi.org/10.1037/a0037231>
- Jazaieri, H., Morrison, A. S., Goldin, P. R., & Gross, J. J. (2014). The role of emotion and emotion regulation in social anxiety disorder. *Current Psychiatry Reports*. <https://doi.org/10.1007/s11920-014-0531-3>
- Kashdan, T. B., Weeks, J. W., & Savostyanova, A. A. (2011). Whether, how, and when social anxiety shapes positive experiences and events: A self-regulatory framework and treatment implications. *Clinical Psychology Review*, 31(5), 786–799. <https://doi.org/10.1016/j.cpr.2011.03.012>
- Lougheed, J. P., & Hollenstein, T. (2012). A limited repertoire of emotion regulation strategies is associated with internalizing problems in adolescence. *Social Development*, 21(4), 704–721. <https://doi.org/10.1111/j.1467-9507.2012.00663.x>
- Marwan, N., & Kurths, J. (2002). Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters A*, 302(5–6), 299–307. [https://doi.org/10.1016/S0375-9601\(02\)01170-2](https://doi.org/10.1016/S0375-9601(02)01170-2)
- Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, 36(4), 455–470.

- Molenaar, P. C. M., Huizenga, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology* (pp. 339–360). Kluwer Academic Publishers. [https://doi.org/10.1007/978-1-4615-0357-6\\_15](https://doi.org/10.1007/978-1-4615-0357-6_15)
- Moulder, R.G., Daniel, K.E., Teachman, B.A., & Boker, S.M. (2021). Tangle: A metric for quantifying complexity and erratic behavior in short time series. *Psych Methods*. Advance online publication. <https://doi.org/10.1037/met0000386>
- Muenz, L. R., & Rubinstein, L. V. (1985). Markov models for covariate dependence of binary sequences. *Biometrics*, *41*, 91–101.
- Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, *81*(2), 535–549. <https://doi.org/10.1007/s11336-014-9435-8>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing <https://www.R-project.org/>
- Rajaram, R., Castellani, B., & Wilson, A. N. (2017). Advancing Shannon entropy for measuring diversity in systems. *Complexity*. <https://doi.org/10.1155/2017/8715605>
- Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University. R package version 2.2.3. <https://www.CRAN.R-project.org/package=psych>
- Romano, M. C., Thiel, M., Kurths, J., & von Bloh, M. (2004). Multivariate recurrence plots. *Physics Letters A*, *330*(3–4), 214–223.
- Rusch, S., Westermann, S., & Lincoln, T. M. (2012). Specificity of emotion regulation deficits in social anxiety: an internet study. *Psychology and Psychotherapy: Theory, Research and Practice*, *85*(3), 268–277. <https://doi.org/10.1111/j.2044-8341.2011.02029.x>
- Southward, M. W., Altenburger, E. M., Moss, S. A., Cregg, D. R., & Cheavens, J. S. (2018). Flexible, yet firm: A model of healthy emotion regulation. *Journal of Social and Clinical Psychology*, *37*(4), 231–251. <https://doi.org/10.1521/jscp.2018.37.4.231>
- Tian, W., & Anderson, S. J. (2000). Markov chain models for multivariate repeated binary data analysis. *Communications in Statistics: Simulation and Computation*, *29*(4), 1001–1019. <https://doi.org/10.1080/03610910008813651>
- Wallot, S. (2019). Multidimensional Cross-Recurrence Quantification Analysis (MdCRQA) - A Method for Quantifying Correlation between Multivariate Time-Series. *Multivariate Behavioral Research*, *54*(2), 173–191. <https://doi.org/10.1080/00273171.2018.1512846>
- Webber, C. L., & Marwan, N. (2015). *Recurrence Quantification Analysis: Theory and Best Practices*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-07155-8>
- Webber, C. L., & Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, *76*(2), 965–973. <https://doi.org/10.1152/jap.1994.76.2.965>
- Wen, A., Quigley, L., Yoon, K. L., & Dobson, K. S. (2021). Emotion regulation diversity in current and remitted depression. *Clinical Psychological Science*, *9*(4), 563–578. <https://doi.org/10.1177/2167702620978616>
- Zbilut, J. P., Giuliani, A., & Webber, C. (1998). Recurrence quantification analysis and principal components in the detection of short complex signals. *Physics Letters A*, *3*(237), 131–135. [https://doi.org/10.1016/S0375-9601\(97\)00843-8](https://doi.org/10.1016/S0375-9601(97)00843-8)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.