Detecting hysteresis in psychological processes with the hysteretic threshold autoregressive (HysTAR) model

Daan de Jong, Oisín Ryan, Han L.J. van der Maas, Ellen L. Hamaker

December 15, 2023

Abstract: *Many psychological processes consist of different states or regimes, which are characterized by, for instance, different means and/or variances. Switches between these regimes are sometimes characterized by hysteresis, which makes a switch from one regime to another regime difficult to reverse. In the current study, we present the hysteretic threshold autoregressive (HysTAR) model, which can be used to detect hysteresis in a bivariate time series consisting of a control variable and an outcome variable. We created a user-friendly package in R to simulate data the HysTAR model and estimate its parameters. Estimation and model selection performance is investigated in simulation studies. Additionally, we further demonstrate the use of the HysTAR model with a time series of depression scores generated from a network of depression symptoms and with an empirical example on hysteresis in the speed-accuracy trade-off.*
Keywords: *hysteresis, regime switching, non-lineairity, bimodality, bistability, dynamical systems, threshold autoregression, intensive longitudinal data, time series analysis*

1 Introduction

Individuals are complex dynamical systems, whose emotions, cognitions and behaviors constantly change over time. To study the dynamics of their psychological processes, reseachers often measure these individuals at relatively many occasions that are typically closely spaced in time (Bolger & Laurenceau, 2013; Walls & Schafer, 2006). The data resulting from these measurements are called *intensive longitudinal data* (ILD) or *time series data*, roughly defined to consist of at least 20 measurement occasions. Technological advances have caused a enourmous growth in the collection of ILD with the use of ecogical momentary assessments, experience sampling and daily diaries (Hamaker & Wichers, 2017), resulting in a body of data containing a rich source of information about individual processes. However, this source of information remains largely untapped when using conventional statistical models like the autoregressive model of order 1 (AR(1), Box et al., 2015; Chatfield & Xing, 2019). These models are often too restrictive to capture the inherently complex dynamics underpinning psychological phenomena, since they imply a symmetric distribution with a constant mean, variance and temporal associations between measurements (Haslbeck et al., 2022; Bringmann et al., 2017). However, it is evident that many psychological time series data exhibit nonstandard features, such as skewness (Haslbeck et al., 2023; Peeters et al., 2006; Von Klipstein et al., 2022), time-varying temporal associations (Koval & Kuppens, 2012; Bringmann et al., 2017) and bimodality (Haslbeck et al., 2023; Van der Maas & Molenaar, 1992). The inadequacy of AR models in capturing these features has underscored the necessity for alternative modeling approaches.

In the current paper, we focus on modeling approaches to account for bimodality and jumps in psychological processes. By bimodality, we simply mean that the distribution of the variable of interest has two modes. From a dynamical systems perspective, these two modes of the variable of interest correspond to two *stable equilibrium states* or two *regimes* of the underlying system. A stable equilibrium state is a state of the system at which it settles in the long run and returns to after small perturbations. A regime is a closely related concept and refers to a specific pattern of behavior exhibited by the system over an extended period of time, and can be characterized by the stable equilibrium state that the system moves around and/or the rate at which this happens. Jumps refer to large sudden changes in the variable of interest. These changes occur at a state transition or regime switch, when the system "jumps" from one stable equilibrium state to another.

One class of models that was developed to account for bimodality and large sudden changes is formed by threshold autoregressive (TAR) models (Tong & Lim, 1980). In these time series models, the variable of interest falls within a certain regime at each point in time. The value of a *control variable* determines which regime this is. When the control variable crosses a certain *threshold*, the system shifts to another regime. For example, the control variable temperature can push a system of water molecules from a solid into a liquid regime, when it crosses a threshold of 0 degrees Celsius.

But TAR models fail to account for an interesting regime-swichting phenomenon, however, the phenomenon of hysteresis. Hysteresis is present when the switches between two regimes are associated with two different thresholds, one for each direction of the state transition (Gilmore, 1993; Strogatz, 2018). As a simple example, under high pressure, water freezes at -4 degrees Celsius and melts at 0 degrees Celcius, so there is a different temperature threshold for the transition from solid to liquid. As a result, a regime switch cannot be reversed as easy as would be the case when there is a single threshold (when water would have frozen and melted at 0 degrees Celsius, for example). This irreversibility of hysteresic switches has intrigued researchers from fields like magnetics, biology, ecology, economy, sociology and psychology for a long time (Mayergoyz, 2003).

Currently, hysteresis is often modeled with the (stochastic) cusp catastrophe model in psychological applications. Since this model uses differential equations to describe the system's dynamics, it is quite a challenge to fit this model to intensive longitudinal data (Cobb, 2010; Hartelman et al., 1998; Wagenmakers et al., 2005; Grasman et al., 2009). Instead, the cusp catastrophe model is often fit to cross-sectional data (e.g., Clair, 1998; Hardy, 1999; Ceja & Navarro, 2012), which heavily impedes any conclusion about dynamics at the individual level (Molenaar, 2004; Hamaker, 2012). As a result, not much is known about hysteresis in psychological processes, yet.

Fortunately, a straightforward adaptation to the TAR model that incorporates the hysteresis effect was recently developed in the field of econometrics (G. Li et al., 2015), resulting in the hysteretic TAR (HysTAR) model. In the current paper, we bring the HysTAR model to the psychological realm in an accessible manner, opening up the possibility to study how hysteretic processes unfold over time, using intensive longitudinal data. For example, hysteresis may be present in the sudden transitions between depressed and non-depressed periods, that is, the transitions take place at different threshold values of external stressors.

This paper is structured as follows. In the next section, we introduce the concept of hysteresis. In Section 3, we present the conditional least squares method to estimate the HysTAR model parameters, which was implemented by the first author in the accompanying R-package `hystar` (R Core Team, 2022; De Jong, 2022). In Section 4, we conduct a simulation study to investigate whether hysteresis can be detected with model selection tools. In Section 5, we present another

simulation study, focussed on estimation of the threshold parameters. In Section 6, we generate time series data from an empirical network of depression symptoms. Here, the research question is whether the HysTAR model can detect the hysteresis effect in the regime switches of this network. An empricial example is given in Section 7, where the use of the HysTAR model is demonstrated in an analysis of the speed-accuracy trade-off phenomenon. We conclude with a discussion section where we highlight current and future extensions to the HysTAR model.

2 Models

In this section, we will first introduce the standard threshold autoregressive (TAR) model. Subsequently, we will explain how the TAR model can be expanded to the hysteretic TAR model to account for the hysteresis effect.

2.1 TAR model

The TAR model is a discrete time series model using piece-wise linearization to approximate non-linear features in the data, like bimodality and sudden jumps (Tong & Lim, 1980). This approach yields a very flexible, interpretable but still accurate description of the process at hand, which is one of the reasons why TAR models are so popular and widely applied in economics (Hansen, 2011). In psychology, the TAR model has been applied to capture sudden shifts in affect regulation within individuals (Hamaker et al., 2010; de Groot, 2022) and dyads (De Haan-Rietdijk et al., 2016; Hamaker et al., 2009). In the two-regime TAR model Tong & Lim (1980), an outcome variable y follows an autoregressive process according to one of the regimes R_t at each time $t = 0, 1, \ldots, T$:

$$y_t = \begin{cases} \phi_{00} + \phi_{01}y_{t-1} + \cdots + \phi_{0p_0}y_{t-p_0} + \sigma_0\varepsilon_t & \text{if } R_t = 0 \\ \phi_{10} + \phi_{11}y_{t-1} + \cdots + \phi_{1p_1}y_{t-p_1} + \sigma_1\varepsilon_t & \text{if } R_t = 1, \end{cases} \quad (1)$$

where the residuals ε_t have a mean of 0 and a variance of 1, and are all uncorrelated with each other. Note that all AR parameters are allowed to differ between regimes: the intercept ϕ_{j0}, the autoregressive coefficients $\phi_{j1}, \ldots, \phi_{jp_j}$ and the residual standard deviations σ_j, where j denotes the regime and p_j denotes the order of the AR process, that is, the largest lag that is used in the prediction of y_t. In this way, regimes can differ in their mean, temporal dynamics, and residual variance. The long-run mean of y in regime j is given by $\phi_{j0}/(1 - \sum_{i=1}^{p_j} \phi_{ji})$. The dynamics are described by the autoregressive coefficients, that indicate the extent to which current and past values of y are carried over to future observations. Additionally, regime differences in σ can indicate different associations between y and unmeasured variables, or different amounts of measurement error.

How the process switches between regimes is dictated by the following equation:

$$R_t = \begin{cases} 0 & \text{if } z_{t-d} \leq r \\ 1 & \text{if } z_{t-d} > r, \end{cases} \quad (2)$$

where z denotes the control variable, d is zero or a positive integer allowing for a delayed effect of z and r is the threshold parameter. Note that the delay parameter (also called threshold lag) is associated with the problem of identifying the control variable z, since the value of d changes the control variable. Note that the control variable can also be y itself, in which case the model
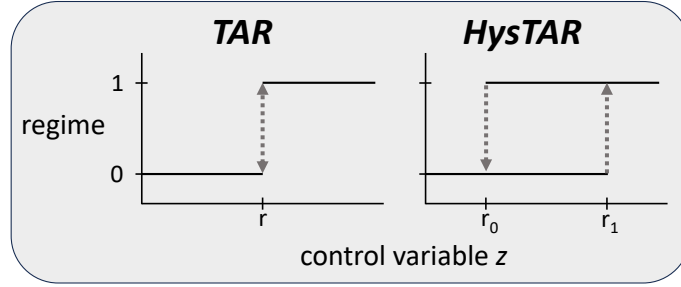
Figure 1: Regime switching in the TAR model (left) and in the HysTAR model (right). On the y-axis, we see the regimes, which can be 0 or 1. On the x-axis, we see the control variable. The black lines represents which regime is active for different levels of the control variable. The dotted arrows represent the sudden jumps between regimes. For simple switches between regime 0 and 1, there is one threshold, r. For hysteretic switching, the switch from 0 to 1 has a threshold of r_1 and the switch from 1 to 0 has a threshold of r_0. Note that both regimes are possible when the control variable is between r_0 and r_1, which is called the hysteresis zone.

is referred to as a self-exciting TAR (SETAR) model. The left panel of Figure 1 illustrates how the regime is related to the control variable.

The switches between regime 0 and 1 have the same threshold, regardless whether the switch is from regime 0 to 1 or from 1 to 0. This may not always be realistic: Sometimes, a switch from 0 to 1 cannot be reversed directly, because the control variable has to move to a different threshold for a switch from 1 to 0. For example, a switch to a regime characterized by high levels of alcohol consumption (a relapse) may be triggered by increasing levels of negative affect. However, the restoration of negative affect will not necessarily bring about a switch back to a low drinking regime again, but may only be possible after decreasing negative affect to an even lower value (Hufford et al., 2003). Or as another example, consider the amplifying feedback loop between perceived threat and arousal producing panic attacks in individuals with panic disorder (D. Robinaugh et al., 2019). When perceived threat is elevated by some external event, an individual might experience a panic attack, and the feedback loops of the system will retain this state. To end the panic attack, perceived threat may have to decrease to a lower level than the level that triggered the attack in the first place. This phenomenon is called hysteresis. Unfortunately, the TAR model cannot capture this, because it has only one threshold parameter. To transfer the benefits of the TAR model into a model that can also account for hysteresis, G. Li et al. (2015) proposed the hysteretic TAR model.

2.2 HysTAR model

In the hysteretic TAR model (HysTAR, also: buffered AR or hysteretic AR model), the dynamics of the outcome variable y in each regime are modeled in the same way as in the TAR model (see Equation 1). The difference is in the regime switching equation, which can be expressed now as

$$
R_t = \begin{cases} 0 & \text{if } z_{t-d} \leq r_0 \\ R_{t-1} & \text{if } r_0 < z_{t-d} \leq r_1 \\ 1 & \text{if } z_{t-d} > r_1. \end{cases}
\tag{3}
$$

In this equation, the regime R_t remains unchanged when z_{t-d} enters the *hysteresis zone*

$(r_0, r_1]$, which leads to the hysteresis effect. In the Figure 1, it is illustrated how the HysTAR model differs from the TAR model. That is, a switch from regime 0 to regime 1 happens when z_{t-d} crosses r_1, but for a switch back from regime 1 to regime 0, z_{t-d} has to cross the lower threshold r_0. Note that the HysTAR model is equivalent to the TAR model if $r_0 = r_1$; hence, the TAR model is nested under the HysTAR model.

3   Model estimation and inference

We take it as a starting point that we already have established that there is threshold non-linearity in the observed data, for example by following the procedure proposed by Tsay (1989) as implemented in the R-package NTS (Tsay et al., 2020). Now, we want to determine whether there is hysteresis in the mechanics underlying the regime switches. Additionally, we assume that the control variable z is known. Although this involves knowing the value of the delay parameter d, we will discuss estimation of d for the sake of completeness. Since the TAR model and HysTAR model only differ in the hysteresis effect, we can compare the model fit of the TAR model to the fit of the HysTAR model to detect hysteresis. This is described in Section 3.5 and demonstrated in Section 4. But first, we provide some tools for parameter estimation and inference.

We implemented the least squares method proposed by G. Li et al. (2015) to estimate the parameters of the HysTAR model in R (R Core Team, 2022; De Jong, 2022), which we will briefly describe here. For conciseness, we introduce the following notation. Let $\phi = (\phi_{00}, \ldots, \phi_{0p_{0\max}}, \phi_{10}, \ldots, \phi_{1p_{1\max}})$ denote the vector of AR coefficients, where $p_{0\max}$ and $p_{1\max}$ are the maximum orders we want to consider. Let $d_{\max}$ denote the maximum delay we want to consider. Let $r = (r_0, r_1)$ denote the vector of thresholds, where $r_0 \le r_1$. Further, let $\sigma = (\sigma_0, \sigma_1)$ denote the vector of residual standard deviations.

For a given $p_0 \le p_{0\max}$ and $p_1 \le p_{1\max}$, our estimates of ϕ, r, and d are the values that minimize the sum of squared errors

$$L(\phi, r, d) := \sum_{t=k}^{T-1} (y_t - \hat{y}_t)^2, \tag{4}$$

where

$$\hat{y}_t = \begin{cases} \phi_{00} + \phi_{01}y_{t-1} + \cdots + \phi_{0p_0}y_{t-p_0} & \text{if } R_t = 0 \\ \phi_{10} + \phi_{11}y_{t-1} + \cdots + \phi_{1p_1}y_{t-p_1} & \text{if } R_t = 1. \end{cases} \tag{5}$$

To select the appropriate orders p_0 and p_1, we cannot use least squares estimation, because the different orders have different numbers of ϕ parameters, and more ϕ parameters will always lead to lower values of $L(\phi, r, d)$. For this purpose, we use information criteria (Section 3.4). Also note that we start the summation of Equation 4 at $t = k := \max\{d_{\max}, p_{0\max}, p_{1\max}\}$; the reason for this is that earlier observations cannot be modeled, because these would need lagged observations y_{t-p_0}, y_{t-p_1}, or z_{t-d} that are unobserved.

3.1   Parameter estimation

Different values of the thresholds r_0 and r_1 can result in different assignments of the time points to the regimes. In turn, this leads to a different sum of squares. However, different regimes will only be assigned when r_0 or r_1 crossed an observed value of z. As a result, $L(\phi, r, d)$ is not continuous in r, so we cannot use the derivative to find the value of r that minimizes

the sum of squared errors. This is solved by minimizing $L(\phi, r, d)$ over a discrete set. In the `hystar` package, this set consists of values that lie between the ordered observed values of z. For example, when $z = (2, 4, 8, 6)$, these values are $\{3, 5, 7\}$. Then, the search set of $\hat{r} = (\hat{r}_0, \hat{r}_1)$ would be $\{(3, 3), (3, 5), (3, 7), (5, 5), (5, 7), (7, 7)\}$, since $\hat{r}_0 \leq \hat{r}_1$ by design. Note that an estimate of 3 for r_0 is actually equivalent to all other values in the range $[2, 4]$ in this case. In practice, often only the 80% inner sample range of z is used, to make sure that there are enough time points assigned to both regimes.

The sum of squared residuals is also not differentiable with respect to the delay parameter d, because d takes on integer values. Therefore, the delay is optimized over the set $\{0, 1, \ldots, d_{\max}\}$, where the maximum delay that is considered $d_{\max}$ is specified by the researcher.

Fortunately, $L(\phi, r, d)$ is differentiable with respect to ϕ given r and d. A closed form solution exists for the estimator $\hat{\phi}$ that minimizes the sum of squared errors (see Appendix A). In this way, we can compute $\hat{\phi}$ for the finite combinations of $\hat{r}$ and $\hat{d}$, and select the solution that minimizes the sum of squared errors.

Finally, we can obtain estimates of the residual standard deviations in the following way. Denote the number of observations $k, \ldots, T$ in regime j by n_j. Then, the estimate of σ_j is given by

$$\hat{\sigma}_j = \sqrt{\frac{1}{n_j} \sum_{t \in \{k, \ldots, T-1\} | R_t = j} \left(y_t - \hat{y}_t\right)^2}. \tag{6}$$

3.2 Estimation challenges

Estimating the HysTAR model comes with two challenges that are not present in the standard TAR model. First, the first regime R_k may be undetermined for certain values of $\hat{r}$ and $\hat{d}$. This can be the case when the observation that determines the first regime, z_{k-d}, falls in the hysteresis zone. A straightforward solution is to apply the estimation procedure to the cases where y starts in regime 0 and where y starts in regime 1, and choose the estimates with the smaller sum of squared errors. However, when one of the AR orders is larger than the delay, there might be observations before z_{k-d} that can inform us about the starting regime, if (one of) these are outside the hysteresis zone, see Figure 2.

Second, suppose we have two consecutive time points with the same regime, so $R_t = R_{t+1}$. This can result from two situations. First, z_{t-d} and z_{t-d+1} are both outside the hysteresis zone, or second, z_{t-d} is outside the hysteresis zone but z_{t-d+1} is inside the hysteresis zone. Both of these situations lead to the same sum of squared errors, because the series of regimes are exactly the same, but have different values for r. In practice, we could choose the more conservative estimate (with the smallest hysteresis zone) and report equivalent estimates. Similarly, different estimates of d can yield the same results (albeit with different threshold estimates associated with them). In that case, the convention is to choose the smallest d, which is more parsimonious (G. Li et al., 2015). In the current paper and the accompanying `hystar`-package, we also adopt this convention.

3.3 Parameter inference

To draw conclusions about the parameter estimates, we need standard errors of the (asymptotic) sampling distributions of the estimators. The estimator of the ϕ-parameters are asymptotically normal, and we can obtain their standard errors analytically (G. Li et al., 2015). The delay
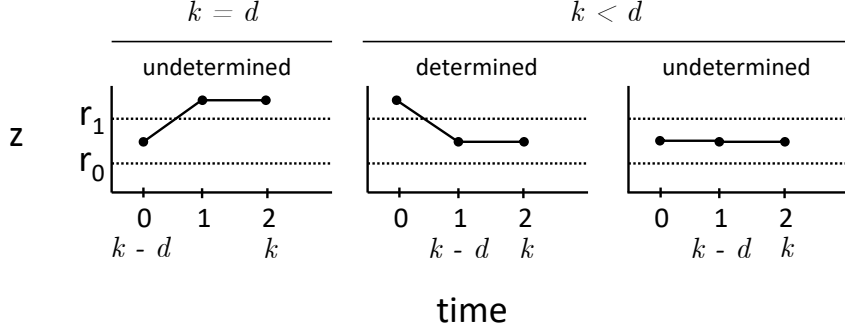
Figure 2: *Two situations where we want to determine the regime at time point k. If $k = d$, then $z_{k-d-1} = z_{-1}$ is unobserved. Hence, R_{k-1} cannot be determined, so R_k can also not be determined. We tackle this scenario by estimating the parameters twice, once while assigning $R_k = 0$, and a second time while assigning $R_k = 1$. Then, we choose the solution that yields the smallest sum of squared errors. If $k > d$, we can have an interesting situation, since we have observations before z_{k-d}. Those observations $z_0, \ldots, z_{k-d-1}$ could be used to determine the regime at time point k, if at least one of them falls outside the hysteresis zone (see Figure 2). This solution is implemented in the `hystar` package. However, if all $z_0, \ldots, z_{k-d-1}$ are in the hysteresis zone, we have to use the solution for the first scenario. (Note that $k < d$ is not possible by definition of k.)*

parameter d is an integer, so $\hat{d}$ will equal d when T is large (G. Li et al., 2015). For the threshold parameters, however, we face a challenge, that is also known from threshold inference in the TAR model. The limiting distribution of $\hat{r}$ does not have a closed form and numerical approximation of this distribution is difficult (G. Li et al., 2015; D. Li & Ling, 2012). As a result, standard errors are not provided for $\hat{r}$ in the `hystar` package (these are also not provided for the `mTAR`-function to estimate the standard TAR model in the R-package `NTS`, Tsay et al., 2020). To still gain some insight in how the threshold estimator behaves, we conducted a simulation study, which is described in Section 4.

3.4 Order selection

We can use information criteria to select the AR orders p_0 and p_1 (see Hamaker, 2009, for an overview of information criteria in the context of threshold models). Then, we can select the orders that minimize the Akaike information criterion (AIC)

$$AIC(\hat{p}_0, \hat{p}_1) = \sum_{j=0}^{1} \left(n_j \log \hat{\sigma}_j^2 + 2(\hat{p}_j + 2) \right),$$ (7)

the corrected AIC (AICc)

$$AIC_c(\hat{p}_0, \hat{p}_1) = \sum_{j=0}^{1} \left(n_j \log \hat{\sigma}_j^2 + \frac{2(\hat{p}_j + 2)}{n_j - \hat{p}_j - 3} \right)$$ (8)

7

or the Bayesian information criterion (BIC)

$$BIC(\hat{p}_0, \hat{p}_1) = \sum_{j=0}^{1} \left(n_j \log \hat{\sigma}_j^2 + (\hat{p}_j + 2) \log n_j \right). \tag{9}$$

Note that we use $\hat{p}_j + 2$ for the number of parameters within a regime, namely $\hat{p}_j$ autoregressive effects, one intercept and a residual standard deviation.

3.5 Model selection

Since the standard error of the threshold estimator is unknown, it cannot be used to choose between the TAR model and the HysTAR model. Otherwise, overlapping confidence intervals for the estimates of r_0 and r_1 could indicate a preference for the TAR model, where r_0 and r_1 are equal. Unfortunately, we can also not use a likelihood-ratio test to compare the HysTAR to the TAR model, since the threshold estimators do not have the same convergence rate in both models (G. Li, personal communication, June 1, 2023). So, we need an alternative approach to assess whether hysteresis is present in the dynamics of the system under investigation.

In the current paper, we propose to estimate information criteria for both models, and select the model with smaller values. Note that the information criteria presented in Section 3.4 do not penalize the threshold parameter(s), and that the HysTAR model has one threshold parameter more than the TAR model. Ignoring this difference may lead to overfitting of the HysTAR model. Therefore, we adopt an AIC-type information criterion that was derived for change point models (AICcp, Ninomiya, 2005). The TAR and HysTAR model are closely related to change point models, since change point models are TAR models where $z = (0, 1, \ldots, T-1)$. In the AICcp, the penalty of a threshold parameter is 6, that is,

$$AIC_{CP} = \sum_{j=0}^{1} \left(n_j \log \hat{\sigma}_j^2 + 2(\hat{p}_j + 2) \right) + 6a, \tag{10}$$

where a is the number of thresholds.

Additionally, we can assess whether the residuals of the model follow a white noise process, which should be the case according to both models. Since the observations of a white noise process are all uncorrelated with each other, we can check whether the autocorrelations of the residuals are not significantly different from zero. This can be done at once with the Ljung-Box test, for example, that tests the null-hypothesis that the autocorrelations are zero (Ljung & Box, 1978), but also by visualizing the autocorrelation function (ACF) of the residuals. We assess the performance of the model selection methods presented here in a simulation study 1 (Section 4).

4 Simulation study 1: Model selection

Detecting hysteresis in empirical time series data raises several practical questions. For example, does it matter how many regime switches have taken place? Does it matter how the regimes differ from each other? And how large does the hysteresis zone need to be? These questions are addressed in simulation study 1.

4.1 Simulation setup

In simulation study 1, we considered the following conditions: 50, 100, 200 and 400 time points, 2, 5, and 10 regime switches (denoted by N_{rs}), (un)equal means and (un)equal autoregression between the regimes, and hysteresis zones $(0, 0]$, $(-0.25, 0.25]$ and $(-0.5, 0.5]$. For each condition, we generated 500 replications of time series of the control variable z and the outcome variable y from the HysTAR model.

This results in a total of $4 \times 3 \times 3 \times 3 = 108$ conditions. To obtain the differences in regimes in means and autoregression, we use $\phi = (\phi_{00}, \phi_{01}, \phi_{10}, \phi_{11}) = (0, 0.6, 0, 0.2)$, $(0, 0.6, 3, 0.2)$ and $(0, 0.4, 3, 0.4)$. Note that the long-run mean of regime j is given by $\phi_{j0}/(1 - \phi_{j1})$. Although threshold non-linearity with equal means seems improbable and also does not imply a unimodal distribution, it is possible. We considered this situation as an edge case, but it is not the typical situation of interest. In all conditions, we set the delay parameter to $d = 0$ and use equal residual variances in the regimes, $\sigma_0^2 = \sigma_1^2 = 1$. Furthermore, to control the number of regime switches, the time series of the control variable is a sine wave, oscillating between -1 and 1 the desired number of times. Since hysteresis zones are centered around zero in the current setup, this results in an equal amount of observations in both regimes for 2 and 10 switches, and slightly more observations in regime 0 when there are 5 regime switches. In Figure 3, we give three examples of time series generated from the HysTAR model.

To see whether we can detect the hysteresis effect, we fit the HysTAR model and TAR model to each data set within a condition. The AIC, BIC, AICc AICcp are used to select the HysTAR or TAR model, by choosing the model with the smaller value. In case of equal values for an information criterion, we select the TAR model, because it is more parsimonious. Additionally, we conduct the Ljung-Box test on the residuals of both models, to test the hypotheses that the residuals are uncorrelated. A significant result from the Ljung-Box test implies that the residuals are correlated, and hence, a model does not capture the dependencies in the data well. We select a model when the p-value of its Ljung-Box test is non-significant (> 0.05) and for the other model it is significant. Otherwise, we consider neither model to be preferred over the other.

In fitting the models, we assume the orders of the regimes known, because we want to focus on model selection between the HysTAR and the TAR model rather than order selection, which is also typically done by comparing information criteria. We also assume the delay parameter known. In our simulations, the HysTAR model is equivalent to a TAR model with very large delay, since the the control variable moves symetrically around the hysteresis zone. Therefore, fixing the delay parameter to zero will lead to a clear distinction between the HysTAR and TAR model. Of course, this is not needed in practice, where z will behave less predictable and the thresholds are not always exactly located around the median of z.

4.2 Results

The code to reproduce this simulation study can be found in the online supplementary materials. [1] In Table 1, we present the proportion of correctly selected models according to the BIC and AICcp within each condition. Results of the AIC and AICc are omitted here, because their values are very similar to the BIC. In general, we see that information criteria can detect hysteresis if the regimes have different means and T is sufficiently large. The Ljung-Box test was not able to detect the HysTAR model very well. However, the test has a very high *precision* (also: positive predictive value), for many conditions 100%. So when there is hysteresis, comparing

---

[1]`https://github.com/daandejongen/hysteresis_paper`

Figure 3: Examples of generated data from the HysTAR model
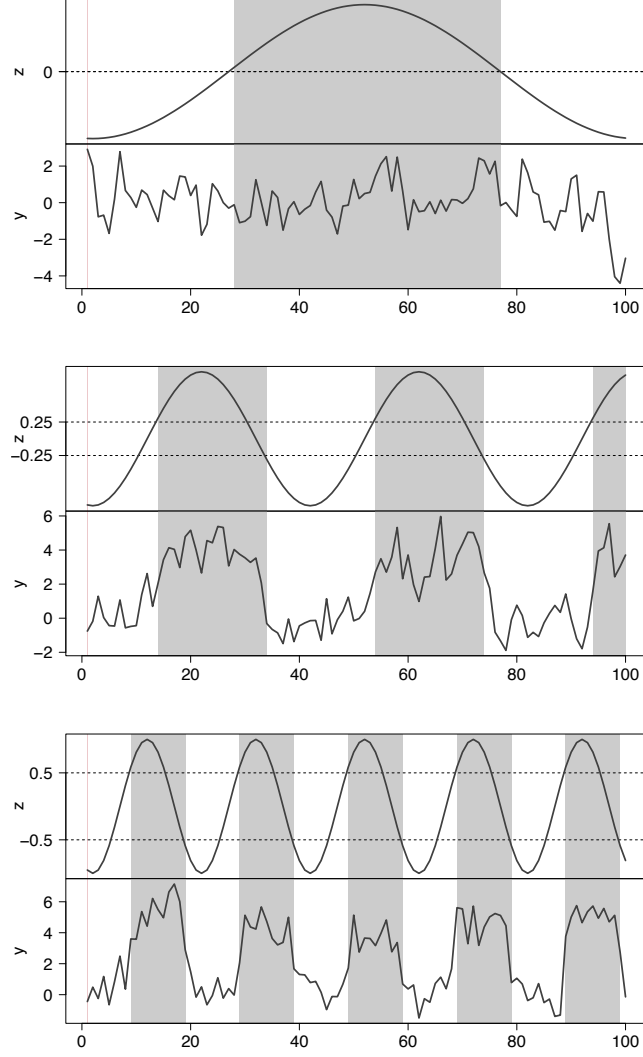
Note. White background = regime 0, grey background = regime 1. Each of the examples have a different number of regime switches, a different hysteresis zone and a different difference between the regimes, from top to bottom: 1) 2 switches, $r = (0, 0]$, equal mean - unequal autoregression. 2) 5 switches, $r = (-0.25, 0.25]$, unequal mean - unequal autoregression. 3) 10 switches, $r = (-0.5, 0.5]$, unequal mean - equal autoregression. For all examples, $T = 100$. Plots were generated with the hystar package in R, and code to create the plots can be found in the supplementary material.

the Ljung-Box results does not indicate hysteresis very often. But when the Ljung-Box comparison is positive, hysteresis is nearly always present (results can be found in the supplementary materials).

When we compare the BIC to the AICcp, we see that the AICcp performs better when the data are generated from a TAR model, and the BIC performs better when the data are generated from HysTAR model. This is unsurprising at first glance, since the AICcp penalizes the HysTAR model 6 points more than the TAR model. Interestingly, it does not seem to matter much how large the hysteresis zone is, when the HysTAR is the true model. This difference between the BIC and AICcp is especially pronounced in the case where the regimes have the same mean. Apparently, the BIC value of the HysTAR model is only slightly higher than that of the TAR model in this case, such that the threshold penalty in the AICcp leads to a preference for the TAR model. When the regimes do not have the same mean, they are much more easily disentangled. For $T \geq 200$, we see that both information criteria perform very well, although the BIC still performs not so well when the TAR model is true.

When the number of regime switches N_{rs} increases, model selection becomes better when the true model is a TAR, but worse when the true model is a HysTAR. This seems odd, but we suspect that this is an artifact from the way the control variable was simulated, as more switches leads to fewer unique values in z (denoted by $\#z$). In turn, this leads to a higher possibility that r_0 and r_1 have the same estimate, such that a TAR model is chosen. This makes it hard to disentangle the effects of T, N_{rs} and $\#z$. However, we can see that $\#z$ is not the only important variable. For example, when comparing conditions with equal values of $\#z$, for example comparing $T = 200$ and $N_{\text{rs}} = 5$ with $T = 400$ and $N_{\text{rs}} = 10$, model selection is better with larger T and N_{rs}. So, there is still additional value to higher T or N_{rs}. Furthermore, the z-values themselves play an imporant role. In the case with $r = (-0.25, 0.25]$, $T = 50$ and $N_{\text{rs}} = 10$, the observed z-values are not close to the true threshold, leading to a correct model selection of 0%. Therefore, it is important to observe enough z-values around the true threshold.

4.3   Conclusion

To detect hysteresis in a regime switching process, comparing information criteria can be a viable method, that compensate for the unavailability of direct statistical hypothesis tests. The AICcp is the preferred model selection criterion, because it outperforms the other ICs when the TAR model is the true model, and performs similar to the other ICs when the HysTAR model is the true model. Comparing Ljung-Box test results between the models has a high type-II error (not selecting the HysTAR when the HysTAR model is true), but a high precision (when the HysTAR is selected, it is often the true model).

5   Simulation Study 2: Parameter estimation

Although the asymptotic distribution of $\hat{\phi}$ is known, we don't know how well it performs in finite samples. For the threshold estimator, it is even more important to assess its finite sample performance. In simulation study 2, we investigate how well the parameters of the HysTAR model can be recovered, when the model is known (or selected) to be a HysTAR model. That is, we estimate the HysTAR model with data generated from the HysTAR model and assess how close the parameter estimates are to the true parameter values. For this study, we use the exact same generated data sets as in simulation study 1.

Table 1: Proportion correct model selection according to the BIC and AICcp

T	N_{rs}	$\#z$	eq.-uneq.		uneq.-uneq.		uneq.-eq.	
			BIC	AICcp	BIC	AICcp	BIC	AICcp
			$r = (0,0]$ (TAR model)					
50	2	19	0.39	0.94	0.64	0.95	0.84	0.99
50	5	8	0.51	0.97	0.80	0.98	0.98	0.99
50	10	3	0.71	0.99	0.95	1.00	1.00	1.00
100	2	40	0.29	0.95	0.63	0.96	0.86	0.99
100	5	16	0.39	0.96	0.80	0.98	0.97	1.00
100	10	8	0.53	0.96	0.95	0.99	1.00	1.00
200	2	80	0.26	0.93	0.63	0.98	0.87	0.99
200	5	32	0.33	0.94	0.81	0.98	0.98	1.00
200	10	16	0.42	0.97	0.95	0.99	1.00	1.00
400	2	160	0.26	0.96	0.66	0.96	0.86	0.99
400	5	64	0.28	0.96	0.85	1.00	0.98	1.00
400	10	32	0.39	0.97	0.96	1.00	1.00	1.00
			$r = (-0.25, 0.25]$					
50	2	19	0.65	0.06	0.97	0.60	0.98	0.88
50	5	8	0.52	0.04	0.81	0.37	0.95	0.74
50	10	3	0.30	0.01	0.05	0.01	0.00	0.00
100	2	40	0.76	0.10	1.00	0.93	1.00	0.99
100	5	16	0.65	0.09	0.98	0.80	1.00	0.99
100	10	8	0.57	0.06	0.92	0.69	1.00	0.98
200	2	80	0.86	0.14	1.00	1.00	1.00	1.00
200	5	32	0.76	0.11	1.00	0.99	1.00	1.00
200	10	16	0.71	0.10	1.00	0.98	1.00	1.00
400	2	160	0.93	0.32	1.00	1.00	1.00	1.00
400	5	64	0.91	0.32	1.00	1.00	1.00	1.00
400	10	32	0.88	0.32	1.00	1.00	1.00	1.00
			$r = (-0.5, 0.5]$					
50	2	19	0.64	0.08	0.99	0.86	1.00	0.95
50	5	8	0.56	0.05	0.98	0.77	1.00	0.97
50	10	3	0.46	0.05	0.98	0.89	1.00	0.99
100	2	40	0.77	0.16	1.00	0.99	1.00	1.00
100	5	16	0.74	0.15	1.00	0.98	1.00	1.00
100	10	8	0.61	0.10	1.00	0.98	1.00	1.00
200	2	80	0.88	0.29	1.00	1.00	1.00	1.00
200	5	32	0.84	0.24	1.00	1.00	1.00	1.00
200	10	16	0.87	0.33	1.00	1.00	1.00	1.00
400	2	160	0.94	0.57	1.00	1.00	1.00	1.00
400	5	64	0.96	0.57	1.00	1.00	1.00	1.00
400	10	32	0.93	0.58	1.00	1.00	1.00	1.00

Note. T = number of time points, N_{rs} = number of regime switches, $\#z$ = number of unique values in the 80% inner range of z. (un)eq.-(un)eq. = (un)equal means-(un)equal autoregression. BIC = Bayesian information criterion, AICcp = change point Akaike information criterion. r is the hysteresis zone.

5.1 Performance measures

Unbiasedness, efficiency and consistency are well-known desirable properties of estimators. Therefore, within each condition, we compute the empirical bias $N_R^{-1} \sum (\hat{\theta} - \theta)$, the empirical variance $N_R^{-1} \sum (\hat{\theta} - \bar{\theta})^2$ and the mean squared error $N_R^{-1} \sum (\hat{\theta} - \theta)^2$, where the summation is over the N_R replications in a condition, $\hat{\theta}$ is an estimate of the parameter θ, and $\bar{\theta}$ denotes the mean of the estimates. For an unbiased estimator, the bias is zero. An efficient estimator has low variance, and a consistent estimator is unbiased and has a variance that approaches zero as $T \rightarrow \infty$. Note that the estimator of ϕ is known to be asymptotically unbiased and consistent, so we should see a bias and a variance that approach zero as T increases.

We also check coverage rates of the confidence intervals, that is, the proportion of times the estimated 95% confidence interval of a ϕ-parameter contains the true value of that ϕ-parameter, which should be 95%. Additionally, the mean of the estimated standard errors of a parameter should be close to the empirical standard deviation of the estimates of that parameter.

5.2 Results

In Figure 4, the bias and variance of the threshold estimators are presented. Results for the ϕ-parameters are omitted here, since all estimators clearly show the desired properties, apart from $\hat{\phi}_{10}$, that converges somewhat slower than the other ϕ-parameters. All results not presented here can be found in the online supplementary materials, as well as the code to reproduce this simulation study.

In general, the threshold estimators seem to be consistent, because the variances converge to zero when T increases, which is in line with the analytical results of G. Li et al. (2015). However, we also need to consider the bias, which is not always zero. When the regimes have unequal means, we see a "mirroring" pattern around zero. The estimates of r are conservative, that is, the hysteresis zone is often underestimated. For $T \geq 200$, the biases become almost negligible. Also interesting is the fact that an equal or unequal autoregressive parameter does not seem to impact the amount of bias.

A different pattern can be observed when the means of the regimes are equal. When $r = (-0.25, 0.25]$, we see that both estimators are negatively biased. For equal means and $r = (-0.5, 0.5]$, $\hat{r}_0$ has a positive bias and $\hat{r}_1$ has a negative bias, implying that the size of the hysteresis zone is estimated conservatively (too small). However, the variance of the estimators show that the model has difficulty recognizing the threshold(s) when the regimes have equal means, which was also clear from the model selection results in Section 4.2. With such large variances, differences in bias may be less informative and individual parameter estimates are harder to interpret, regardless of whether the correct type of model is selected or not.

We note two more observations about estimation of the thresholds. First, when the autoregression in regime 0 is larger than in regime 1, a regime switch from 1 to 0 (caused by threshold r_0) will be "slower" than a switch from 0 to 1. That is, it takes longer to arrive at the mean of regime 0, because of the larger autoregression. As a result, the first few observations in regime 0 will also fit well in regime 1. Therefore, estimation of r_0 is more difficult than estimation of r_1. This is reflected by the fact that the variance of $\hat{r}_0$ is somewhat larger than the variance of $\hat{r}_1$ when the regimes have unequal autoregression, and is most visible for low T. Second, we can see small peaks in the variances where $N_{rs} = 5$ for $\hat{r}_0$, and not for $\hat{r}_1$. When $N_{rs} = 5$, there will be three switches by which to identify r_1 and two switches for r_0, because the generated data always start in regime 0. As a result, r_1 is easier to estimate in that case.

13

Figure 4: Empirical bias and variance of $r = (-0.25, 0.25]$ (top pane) and $r = (-0.5, 0.5]$ (bottom pane)
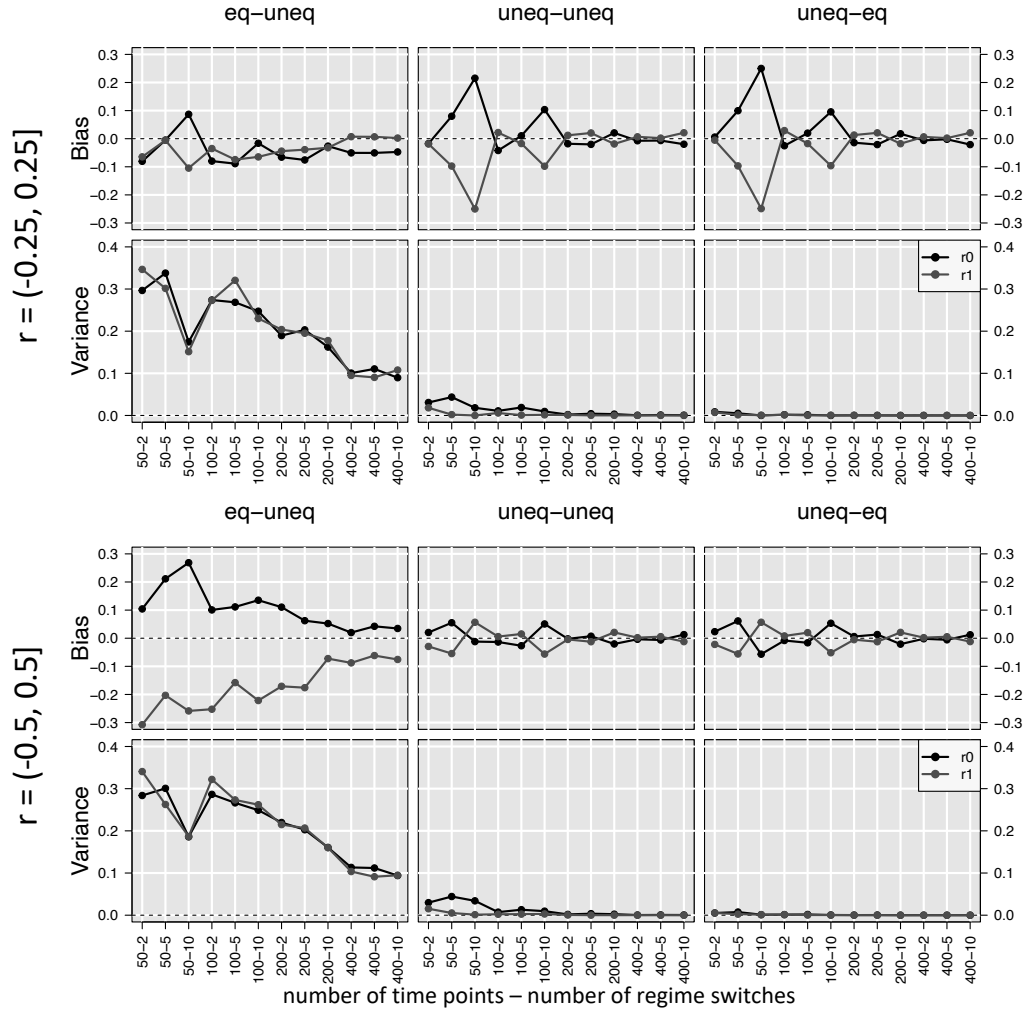
6   Simulation study 3: Depression

In the preceding simulation studies, the data generating mechanism was the HysTAR model. But there are also other models that produce the hysteresis effect. Therefore, in the current simulation study, we generate data from a different mechanism that gives rise to hysteretic regime-switching, which should still be picked up by the HysTAR model.

One research area where there has been particular interest in the phenomenon of hysteresis is the network approach to psychopathology (Borsboom, 2008). We follow an example of a network model of major depression as studied by Cramer et al. (2016). This network consist of 14 connected depression symptoms that can be switched on or off, where the sum of symptoms that are switched on is the level of depression. The symptoms trigger their neighboring symptoms to switch on or off, over time. For example, fatigue may lead to depressed mood, which in turn leads to loss of interest, which again leads to fatigue, et cetera. With the existence of such feedback loops, two stable regimes can emerge: one where most symptoms are switched on, representing a depression, and one where most symtoms are switched off, representing a healthy regime.

Depression networks in which the symptoms are highly connected are said to be vulnarable to external stress. That is, when external stress triggers a symptom to switch on, this will quickly result in a "cascade" of symptoms switching on. In the study by Cramer et al. (2016), external stress triggering the symptoms was simulated to go up an down over time. In this way, stress is a control variable that causes switches between the healthy and depressed regime. The authors found support for a hysteresis effect in these switches by fitting the cusp catastrophe model with the `cusp` package in R (R Core Team, 2022; Grasman et al., 2009).

In the current study, we replicated the simulation by Cramer et al. (2016) with 1000 time points. We used the same network (same nodes and connections), which was based on empirical data (Prescott et al., 2000). The stress levels were varied between -1 and 1. Since the control variable is completely known here, we do not estimate d here, but fix it to zero. We further estimated the HysTAR model with fixed orders of 1 and a threshold search space in the 80% inner sample range of stress, which were stress levels between -0.947 and 0.947. The R-code that was used to run the simulation can be found in the supplementary materials.

6.1   Results

Figure 5 shows the simulated time series of stress and depression. In Table 2, the HysTAR and TAR model estimates are shown. According to the AIC, AICc and BIC, the HysTAR model is preferable. The Ljung-Box results were indecisive, in the sense that these indicated that both models fit well. The difference between the threshold estimates indicates a substantial hysteresis effect, $\hat{r}_1 - \hat{r}_0 = 0.55 - (-0.33) = 0.88$. This shows that the hysTAR model is able to capture the core qualitative phenomenon of hysteretic regime switching, even when the generative model is quite different in form to an auto-regressive type model. Additionally, the ϕ-parameters show a clear distinction between the regimes. The mean depression score is 0.85 regime 0 and 9.09 in regime 1, such that we can interpret regime 0 as the healthy regime and regime 1 as the depression regime. We also see that the depression regime has a stronger autoregression, which is supported by empirical evidence (Kuppens et al., 2010; Koval et al., 2012). Notably, these dynamic differences as captured by the autoregressive parameters were not found in the original study by Cramer et al. (2016), because the cusp catastrophe model does not allow for dynamic differences between regimes.
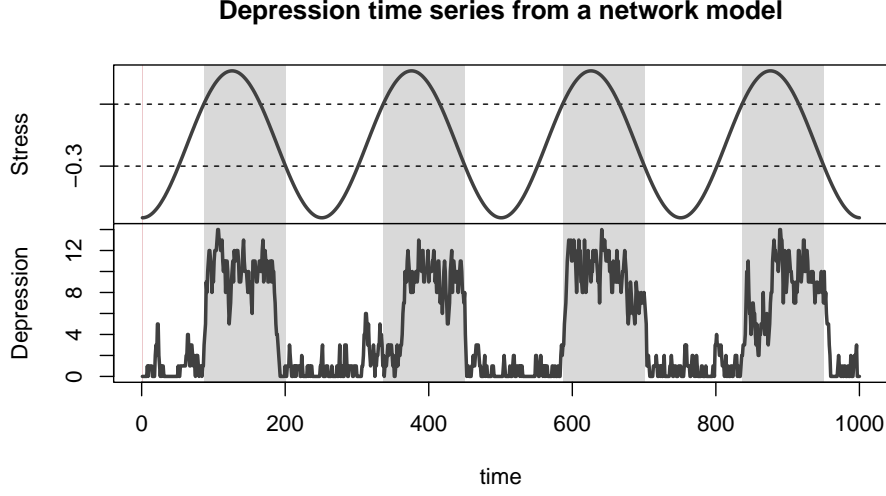
15

Figure 5: *Time series plots of stress and depression, simulated from a network model. The dashed lines represent the estimates threshold values of the HysTAR model. The grey background represents the depressed regime and the white background represents the healthy regime.*

Estimand	HysTAR	TAR
ϕ_{00}	0.23 $[0.14, 0.31]$	0.21 $[0.12, 0.29]$
ϕ_{01}	0.73 $[0.68, 0.78]$	0.67 $[0.61, 0.73]$
ϕ_{10}	1.00 $[0.64, 1.36]$	0.40 $[0.20, 0.60]$
ϕ_{11}	0.89 $[0.85, 0.93]$	0.95 $[0.92, 0.97]$
r	$(-0.30, 0.55]$	$(-0.32, -0.32]$
σ	$(0.70, 1.93)$	$(0.58, 1.78)$
AIC	116	141
AICc	116	141
BIC	141	166
AICcp	128	147
$\chi^2(1)$	3.35	2.48
p	0.067	0.115

Table 2: Parameter estimates of the HysTAR model and TAR model on data simulated from an empirical depression model, $T = 1000$. In brackets are the 95% confidence intervals of the ϕ-parameters. The delay parameter d is not estimated, but fixed to zero. $\chi^2(1)$ is result of the Ljung-Box test on the model residuals.

7   Empirical example: Speed-accuracy Trade-off

The HysTAR model has been succesfully applied to economic time series, such as growth rates of stocks and prices (Tsay, 2021), daily number of stock trades (Liu et al., 2020), air pollution (Chen et al., 2019) and quarterly GDP growth (G. Li et al., 2015). However, the HysTAR model

has not been used to investigate hysteresis in psychological processes, so far.

Therefore, we reanalyzed data from the speed-accuracy trade-off experiment by Dutilh et al. (2011, experiment 1b), which were originally analyzed with a hidden Markov model, with the HysTAR model [2]. In this hidden Markov model, hysteretic switches can be inferred from time series data, but the dynamics within regimes are not modeled. We selected participants F and I, corresponding to the participant labels used in the original study for illustrative purposes, because their data showed clear hysteresis effects. Participant F performed a lexical decision task where they had to discriminate word from non-word stimuli, and participant I had to judge whether a horizontal line crossing a vertical line extended more to the right or to the left. The researchers measured the response time (RT) and response accuracy (RA) of the participants at each trial. For each response, the participants received a payoff for speed and a payoff for accuracy, always summing op to 24 units of total payoff. When participants receive 24 units for a fast answer (so zero for an accurate answer), we expect them to behave according to a guess (G) regime with low response time and low accuracy. On the other hand, if there is only payoff for accurate answers, participants will follow a stimulus-controlled (SC) regime with high response time and high accuracy. The focus in the study was on the transitions between the two regimes under gradual changes of payoff for accuracy P_t^{AC}, which was under the control of the researchers. The researchers expected that the regime switches would happen at different values of P_t^{AC}, which indicates hysteresis. For further details about the experimental setup, see Dutilh et al. (2011).

7.1   Data analysis

Both participants completed two sessions (0 and 1), with 671 and 497 trials and 420 and 553 trials, respectively. We fit the HysTAR model to the log response times from each session separately, with P_t^{AC} as the control variable. We considered an autoregressive model of order 1 for both regimes and did not estimate the delay parameter. The search space of the thresholds was restricted to the inner 80% sample range of P_t^{AC}.

7.2   Results

We found support of a hysteresis effect for both participants F and I in both sessions, which is in line with the conclusions of Dutilh et al. (2011). Based on the information criteria, the HysTAR model is preferred in nearly all cases, see Appendix B. Here, we will discuss the results of participant I, during the first session, because it shows the largest hysteresis effect $(r_1 - r_0 = 6)$. The information criteria estimates for this case were contradictory, due to the 6-point penalty of the AICcp, but the Ljung-Box test was significant for the TAR model $(\chi^2 = 4.37, \mathrm{df} = 1, p = .037)$ and not for the HysTAR model $(\chi^2 = 1.38, \mathrm{df} = 1, p = .240)$, which indicates the presence of hysteresis.

Figure 6 shows the time series of P_t^{AC} together with the log response times. Visually, in the time series of response times itself, we can distinguish two distinct processes representing the guess regime with low response times and the stimulus-controlled regime with high response times. Note that this is supported by the white and grey background, representing the regimes as estimated in the HysTAR model. The estimated HysTAR model with y_t denoting the log

---

[2]We received permission to reuse these data from the Ethics Committee of the Faculty of Social and Behavioural Sciences of Utrecht University (23-0107).
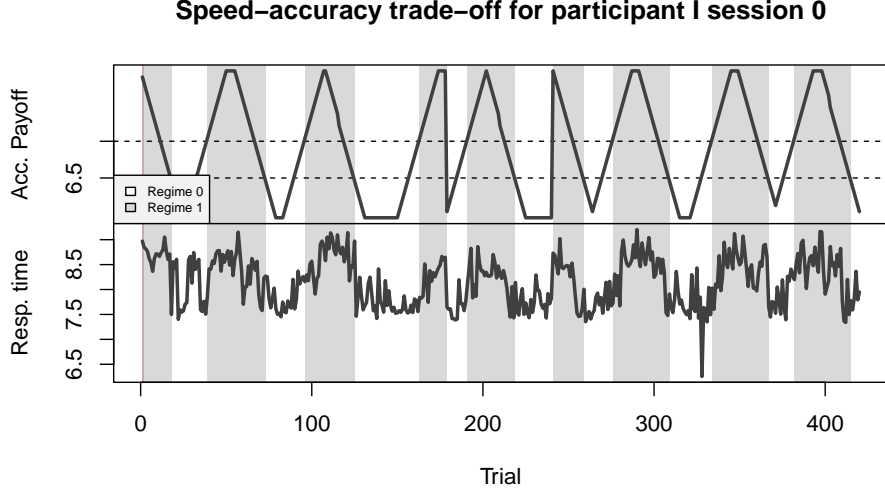
Figure 6: *Time series plot of response time as the outcome variable and payoff for accuracy as the control variable in a speed-accuracy trade-off experiment. The guess regime is in white and the stimulus-controlled regime is in grey.*

response time at time t is

$$y_t = \begin{cases} 5.98 + 0.23y_{t-1} + 0.11\varepsilon_t & \text{if } R_t = \text{G} \\ 3.54 + 0.58y_{t-1} + 0.10\varepsilon_t & \text{if } R_t = \text{SC}, \end{cases} \tag{11}$$

$$R_t = \begin{cases} \text{G} & \text{if } P_t^{\text{AC}} \in \{0, \ldots, 6\} \\ R_{t-1} & \text{if } P_t^{\text{AC}} \in \{7, \ldots, 12\} \\ \text{SC} & \text{if } P_t^{\text{AC}} \in \{13, \ldots, 24\}. \end{cases} \tag{12}$$

For all ϕ-estimates, $p < .05$. Note that the values of the intercepts might suggest that response time is higher in the guessing regime, however, intercepts have a different interpretation in autoregressive models than they do in standard regression models. To compare the regimes, we need to look at the means. The mean log response time is $5.98/(1 - 0.23) = 7.77$ in the guessing regime ($R_t = \text{G}$) and $3.54/(1-0.58) = 8.51$ in the stimulus-controlled regime ($R_t = \text{SC}$). Note that the autoregressive effect of response times from one moment to the next is larger in the stimulus-controlled regime, indicating that the regimes are qualitatively different in their dynamics, additional to the difference in their means. This could be the case because guessing behavior is more basic and reflexive, so its variability may be more random. In contrast, stimulus controlled behavior is characterized by cognitive processes that may be influenced more by past trials (Laming, 1979).

8 Discussion

We presented the hysteretic threshold autoregressive (HysTAR) model (Zhu et al., 2014; G. Li et al., 2015) to study hysteresis in univariate psychological processes characterized by two regimes.

This model can be estimated with the accompanying R-package `hystar` (R Core Team, 2022; De Jong, 2022). To the authors' knowledge, this is the first study to apply the HysTAR model to psychological time series. With simulations, it is demonstrated that information criteria can distinguish the HysTAR model from a TAR model. Additionally, the Ljung-Box test can be conducted on the residuals of both models, and show a high positive predictive value. Dynamic differences were found between simulated depression network regimes and between guessing and empirical stimulus-controlled regimes of response times, captured by the autoregressive parameters of the HysTAR model. These differences are not modeled in existing methods to study hysteresis, like the cusp catastrophe model or the hidden Markov model (HMM). Additionally, the HysTAR model is well suited for time series data, whereas the cusp model is not applicable to time series data in its current implementations (Grasman et al., 2009; Chow et al., 2014).

While the HysTAR model is promising for psychological applications, the model as presented here also has a number of limitations which are relevant for empirical psychological applications. First, it does not account for measurement error, in contrast to the mixture structural equation model with regime switching (MSEM-RS) approximation to the cusp catastrophe model (Chow et al., 2014). The MSEM-RS model has the additional benefit of being nested in the SEM framework, which has a longstanding tradition and is implemented in many software programs. Second, the regimes only define the process of one outcome variable, whereas multiple outcomes are easily incorporated in the MSEM-RS, cusp model and HMM. Therefore, it should be noted that various extensions of this model exist. For example, the Hysteretic VAR model proposed by Chen et al. (2019) allows for multiple outcome variables. When there are multiple control variables, an alternative to the HysTAR model is the multivariate Hysteretic AR (MHAR) model proposed by Tsay (2021). In the MHAR model, multiple control variables all have their own threshold and the regime of y_t only switches when all control variables have crossed their threshold.

Other extensions to the HysTAR model could be developed in the future. First, it would be interesting to incorporate a second control variable in the HysTAR model, analogous to the splitting variable of the cusp catastrophe model that controls whether hysteresis present. Such a variable could be straightforwardly implemented by adding an additional "layer" to the regime switching rule. Second, a multilevel extension can be developed to investigate individual differences; such an extension already exists for the TAR model (De Haan-Rietdijk et al., 2016). Third, it should be noted that the hysteretic HMM could be extended to include autoregression as well. Then, the HMM and the HysTAR model would only differ in their switching mechanism, which is probabilistic in the HMM, and deterministic in the HysTAR model. Depending on the theory under examination, one or the other may be preferred. Whether one or the other is preferable from a statistical point of view (e.g., properties of estimators, complexity of the estimation algorithm) should be investigated in future research.

Finally, we like to stress that, in our view, the HysTAR model should be primarily used to detect hysteresis. As we saw in simulation 3, one advantage of the model is that it can detect hystersis in different generative models. However, although we detected this phenomenon, the model did not give a description of the mechanism through which hysteretic regime switching took place. For this purpose, we need to formulate and test formal models that represent the systems underlying the hysteretic process (Borsboom et al., 2021; Haslbeck et al., 2022; Ryan et al., 2023; D. J. Robinaugh et al., 2021). Then, the HysTAR model can be useful to confirm that a formal model can produce the hysteresis effect.

References

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford press.

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*, 1089–1108. doi: 10.1002/jclp.20503

Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, *16*(4), 756–766.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons, Inc.

Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, *22*, 409–425. doi: 10.1037/met0000085

Ceja, L., & Navarro, J. (2012). 'suddenly i get into the zone': Examining discontinuities and nonlinear changes in flow experiences at work. *Human Relations*, *65*, 1101–1147. doi: 10.1177/0018726712447116

Chatfield, C., & Xing, H. (2019). *The analysis of time series: An introduction with r*. CRC press.

Chen, C. W., Than-Thi, H., & So, M. K. (2019). On hysteretic vector autoregressive model with applications. *Journal of Statistical Computation and Simulation*, *89*, 191–210. doi: 10.1080/00949655.2018.1540619

Chow, S.-M., Witkiewitz, K., Grasman, R., Hutton, R. S., & Maisto, S. A. (2014). A regime-switching longitudinal model of alcohol lapse-relapse. *Handbook of developmental systems theory and methodology*, 397–422.

Clair, S. (1998). A cusp catastrophe model for adolescent alcohol use: An empirical test. *,* 217–241.

Cobb, L. (2010). Estimation theory for the cusp catastrophe model. revised edition. In *Proceedings of the american statistical association section on survey research methods* (pp. 772–776).

Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLOS ONE*, *11*, e0167490. doi: 10.1371/journal.pone.0167490

de Groot, J. (2022). *Uncovering change in emotion regulation in clinical practice with the threshold autoregressive model* (Doctoral dissertation). Retrieved from `https://fse.studenttheses.ub.rug.nl/27617/`

De Haan-Rietdijk, S., Gottman, J. M., Bergeman, C. S., & Hamaker, E. L. (2016). Get over it! a multilevel threshold autoregressive model for state-dependent affect regulation. *Psychometrika*, *81*, 217–241. doi: 10.1007/s11336-014-9417-x

De Jong, D. (2022). hystar: Fit the hysteretic tar model to time series data [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=hystar` (R package version 1.2.0)

Dutilh, G., Wagenmakers, E.-J., Visser, I., & van der Maas, H. L. J. (2011). A phase transition model for the speed-accuracy trade-off in response time experiments. *Cognitive Science*, *35*, 211–250. doi: 10.1111/j.1551-6709.2010.01147.x

Gilmore, R. (1993). *Catastrophe theory for scientists and engineers*. Courier Corporation.

Grasman, R. P. P. P., van der Maas, H. L. J., & Wagenmakers, E.-J. (2009). Fitting the cusp catastrophe in r : A cusp package primer. *Journal of Statistical Software*, *32*. doi: 10.18637/jss.v032.i08

Hamaker, E. L. (2009). Using information criteria to determine the number of regimes in threshold autoregressive models. *Journal of Mathematical Psychology*, *53*, 518–529. doi: 10.1016/j.jmp.2009.07.006

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). The Guilford Press.

Hamaker, E. L., Grasman, R. P., & Kamphuis, J. (2010). Regime-switching models to study psychological processes. In M. PCM & N. KM (Eds.), *Individual pathways of change: Statistical models for analyzing learning and development* (pp. 155–168). American Psychological Association. doi: 10.1037/12140-009

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*, 10–15. doi: 10.1177/0963721416666518

Hamaker, E. L., Zhang, Z., & van der Maas, H. L. (2009). Using threshold autoregressive models to study dyadic interactions. *Psychometrika*, *74*, 727–745. doi: 10.1007/s11336-009-9113-4

Hansen, B. E. (2011). Threshold autoregression in economics. *Statistics and its Interface*, *4*, 123–127. doi: 10.4310/SII.2011.v4.n2.a4

Hardy, L. (1999). Stress, anxiety and performance. *Journal of Science and Medicine in Sport*, *2*, 227–233. doi: 10.1016/S1440-2440(99)80175-3

Hartelman, P. A., van der Maas, H. L., & Molenaar, P. C. (1998). Detecting and modelling developmental transitions. *British Journal of Developmental Psychology*, *16*, 97–122. doi: 10.1111/j.2044-835X.1998.tb00751.x

Haslbeck, J., Ryan, O., & Dablander, F. (2023). Multimodality and skewness in emotion time series. *Emotion*, Advance online publication. doi: 10.1037/emo0001218

Haslbeck, J., Ryan, O., Robinaugh, D. J., Waldorp, L. J., & Borsboom, D. (2022). Modeling psychopathology: From data models to formal theories. *Psychological Methods*, *27*(6), 930. doi: 10.1037/met0000303

Hufford, M. R., Witkiewitz, K., Shields, A. L., Kodya, S., & Caruso, J. C. (2003). Relapse as a nonlinear dynamic system: application to patients with alcohol use disorders. *Journal of Abnormal Psychology*, *112*(2), 219. doi: 10.1037/0021-843X.112.2.219

Koval, P., & Kuppens, P. (2012). Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion*, *12*, 256–267. doi: 10.1037/a0024756

Koval, P., Kuppens, P., Allen, N. B., & Sheeber, L. (2012). Getting stuck in depression: The roles of rumination and emotional inertia. *Cognition and Emotion*, 1412–1427. doi: 10.1080/02699931.2012.667392

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychologicla maladjustment. *Psychological Science*, *21*, 984–991. doi: 10.1177/0956797610372634

Laming, D. (1979). Autocorrelation of choice-reaction times. *Acta Psychologica*, *43*(5), 381-412. Retrieved from `https://www.sciencedirect.com/science/article/pii/0001691879900325` doi: https://doi.org/10.1016/0001-6918(79)90032-5

Li, D., & Ling, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models. *Journal of Econometrics*, *167*, 240–253. doi: 10.1016/j.jeconom.2011.11.006

Li, G., Guan, B., Li, W. K., & Yu, P. L. (2015). Hysteretic autoregressive time series models. *Biometrika*, *102*, 717–723. doi: 10.1093/biomet/asv017

Liu, M., Li, Q., & Zhu, F. (2020). Self-excited hysteretic negative binomial autoregression. *AStA Advances in Statistical Analysis*, *104*, 385–415. doi: 10.1007/s10182-019-00360-6

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*, 297–303. doi: 10.2307/2335207

Mayergoyz, I. D. (2003). *Mathematical models of hysteresis and their applications*. Academic press.

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*, 201–218. doi: 10.1207/s15366359mea0204_1

Ninomiya, Y. (2005). Information criterion for gaussian change-point model. *Statistics & probability letters*, *72*, 237–247. doi: 10.1016/j.spl.2004.10.037

Peeters, F., Berkhof, J., Delespaul, P., Rottenberg, J., & Nicolson, N. A. (2006). Diurnal mood variation in major depressive disorder. *Emotion*, *6*(3), 383.

Prescott, C. A., Aggen, S. H., & Kendler, K. S. (2000). Sex-specific genetic influences on the comorbidity of alcoholism and major depression in a population-based sample of us twins. *Archives of general psychiatry*, *57*, 803–811.

R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Retrieved from https://www.R-project.org/

Robinaugh, D., Haslbeck, J., Waldorp, L., Kossakowski, J., Fried, E. I., Millner, A., ... others (2019). Advancing the network theory of mental disorders: A computational model of panic disorder.
doi: 10.31234/osf.io/km37w

Robinaugh, D. J., Haslbeck, J. M., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science*, *16*(4), 725–743.

Ryan, O., Dablander, F., & Haslbeck, J. (2023). Towards a generative model for emotion dynamics.

Strogatz, S. H. (2018). *Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering.* CRC press.

Tong, H., & Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *42*, 245–268.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American statistical association*, *84*(405), 231–240.

Tsay, R. S. (2021). Multivariate hysteretic autoregressive models. *Statistica Sinica*, *31*, 2257–2274.

Tsay, R. S., Chen, R., & Liu, X. (2020). Nts: Nonlinear time series analysis [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=NTS (R package version 1.1.2)

Van der Maas, H. L., & Molenaar, P. C. (1992). Stagewise cognitive development: an application of catastrophe theory. *Psychological review*, *99*(3), 395.

Von Klipstein, L., Servaas, M., Lamers, F., Schoevers, R., Wardenaar, K., & Riese, H. (2022). Can floor effects explain increased affective reactivity among depressed individuals found in experience sampling research? , Advance online publication. doi: 10.31234/osf.io/6vtew

Wagenmakers, E.-J., Molenaar, P. C. M., Grasman, R. P. P. P., Hartelman, P. A. I., & van der Maas, H. L. J. (2005). Transformation invariant stochastic catastrophe theory. *Physica D: Nonlinear Phenomena*, *211*, 263–276. doi: 10.1016/j.physd.2005.08.014

Walls, T. A., & Schafer, J. L. (2006). *Models for intensive longitudinal data.* Oxford University Press.

Zhu, K., Yu, P. L., & Li, W. K. (2014). Testing for the buffered autoregressive processes. *Statistica Sinica*, 971–984.

A   Least squares estimation of ϕ given r and d

Let $y = (y_k, \ldots, y_T)$. Let

$$
X := \begin{pmatrix}
1 - R_k & (1-R_k)y_{p_0} & \ldots & (1-R_k)y_1 & R_k & R_k y_{p_1} & \ldots & R_k y_1 \\
1 - R_{k+1} & (1-R_{k+1})y_{p_0+1} & \ldots & (1-R_{k+1})y_2 & R_{k+1} & R_{k+1}y_{p_1+1} & \ldots & R_{k+1}y_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 - R_T & (1-R_T)y_T & \ldots & (1-R_T)y_{T-p_0} & R_T & R_T y_T & \ldots & R_T y_{T-p_1}
\end{pmatrix}
$$

denote the design matrix and let $\hat{\phi} = (\hat{\phi}_{00}, \ldots, \hat{\phi}_{0p_0}, \hat{\phi}_{10}, \ldots, \hat{\phi}_{1p_1})'$ denote the intercepts and AR coefficients, where $'$ denotes the transpose.

Then, the sum of squared errors

$$
L(\hat{\phi}, r, d) = (y - X\hat{\phi})'(y - X\hat{\phi})
$$

is minimized by

$$
\hat{\phi} = \underset{\phi}{\operatorname{argmin}}\, L(\phi, r, d) = (X'X)^{-1}X'y.
$$

See G. Li et al. (2015) for details.

B   Results from SAT experiments

Estimand	Part. F, Session 0		Part. F, Session 1		Part. I, Session 0		Part. I, Session 1	
	Model selection							
	TAR	HysTAR	TAR	HysTAR	TAR	HysTAR	TAR	HysTAR
AIC	-1845	-1853	-1532	-1562	-919	-921	-1185	-1194
AICc	-1845	-1853	-1531	-1562	-919	-921	-1185	-1194
BIC	-1822	-1830	-1511	-1541	-899	-901	-1163	-1173
AICcp	-1839	-1841	-1526	-1540	-913	-909	-1179	-1182
$\chi^2(1)$	8.42	5.50	28.35	17.79	4.37	1.38	2.98	2.49
p	0.004	0.019	0.00	0.00	0.037	0.239	0.084	0.115
	Parameter estimation							
	Est.	95% CI	Est.	95% CI	Est.	95% CI	Est.	95% CI
ϕ_{00}	7.28	$[6.46, 8.11]$	4.82	$[4.05, 5.59]$	5.98	$[4.93, 7.03]$	6.26	$[5.27, 7.26]$
ϕ_{01}	0.08	$[-0.03, 0.18]$	0.38	$[0.28, 0.48]$	0.23	$[0.10, 0.37]$	0.20	$[0.07, 0.32]$
ϕ_{10}	5.21	$[4.44, 5.97]$	5.51	$[4.73, 6.29]$	3.54	$[2.68, 4.40]$	4.72	$[3.96, 5.48]$
ϕ_{11}	0.39	$[0.30, 0.48]$	0.35	$[0.26, 0.44]$	0.58	$[0.48, 0.68]$	0.44	$[0.35, 0.53]$
r_0	8.5	-	10.5	-	6.5	-	9.5	-
r_1	10.5	-	11.5	-	12.5	-	11.5	-
σ_0	0.06	-	0.05	-	0.11	-	0.13	-
σ_1	0.07	-	0.04	-	0.10	-	0.10	-

Table 3: Results of the (Hys)TAR analyses on speed-accuracy time series data, for two participants with two sessions each. The outcome variable is log response time and the control variable is payoff for accuracy. In brackets are the 95% confidence intervals of the ϕ-parameters. The delay parameter d is not estimated, but fixed to zero. $\chi^2(1)$ is result of the Ljung-Box test on the model residuals.