Modelling Affective Forecasts: Comparing the Kalman Filter to the Multilevel Autoregressive Model in an Ecological Momentary Assessment Study

Marieke Boerendonk

s5233224

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

GR26

Supervisor: Fridtjof Petersen

Second evaluator: Dr. Oliver Weigelt

In collaboration with: Fabienne Liepelt, Henrike Nebel, Martino Avalle, and Miora Haslacher

June 30, 2025

Abstract

Making the wrong decision can be very upsetting, but could you have anticipated how you would later feel about this decision? The current study investigated affective forecasting accuracy of two different statistical models. More specifically, the predictive accuracy of the multilevel autoregressive model was compared to that of the Kalman filter. Participants (N = 30) rated and predicted their affect for 14 days in an ecological momentary assessment study, and provided point and interval predictions at the current time, three hours ahead, and six hours ahead. Consequently, one-step predictions (three hours ahead) were computed by the multilevel autoregressive model and the Kalman filter. We hypothesized that, compared to the Kalman filter, the multilevel autoregressive model will perform better due to its extensive use in affect dynamics and its potential to capture the development of emotions. Moreover, the multilevel autoregressive model was expected to specifically outperform the Kalman filter for negative affect (NA), as emotional inertia is suggested to be more strongly related to NA than positive affect (PA). Linear mixed models showed a significant difference between the statistical models in NA, where the multilevel autoregressive model outperformed the Kalman filter. However, no significant difference was found for PA. These findings are important when we consider the biases that humans are subjected to in making affective forecasts. This study might serve as a foundation for bridging statistical models with real-world affective forecasting, as further research could investigate how combining human judgments with statistical forecasts could enhance the overall accuracy.

Keywords: affective forecasting, EMA, multilevel AR model, Kalman filter

Modelling Affective Forecasts: Comparing the Kalman Filter to the Multilevel Autoregressive Model in an Ecological Momentary Assessment Study

People are often required to make decisions in which they have to take their future emotions into account. For example, when deciding to purchase a slice of pie, one must anticipate their satisfaction of eating it. *Affective forecasting*, described as making a prediction about how one feels in the future (American Psychological Association, 2018), has been a major topic of interest in research. Studies show that people are competent at making predictions about whether they will feel positively or negatively and which specific emotions they will likely experience in the future. However, people are less accurate when asked to predict the intensity and duration (Wilson & Gilbert, 2003). You might accurately predict that eating that slice of pie will result in overall positive feelings, and more specifically you might experience happiness, but it is likely that you are mistaken about *how* happy you will feel and how long exactly you will remain happy when eating that slice of pie. This is shown in the *impact bias*, described as the tendency to overestimate how strongly future events will affect us (Wilson & Gilbert, 2003). This bias is reflected in a study by Dunn et al. (2003), who investigated forecasting accuracy by asking how participants would feel if they were living in either a desirable or undesirable house. After one year and then two years, they were asked to describe their actual feelings of living in either house. The impact bias was supported in their results, showing that participants overestimated their level of happiness living in a desirable house, and overestimated their misery living in an undesirable house. Their study reinforces the finding that people are subject to biases when predicting future emotional states. In addition, individuals have also been found to differ in their forecasting accuracy for positive affect (PA) and negative affect (NA). Thompson et al. (2017) investigated this difference in predicting PA and NA as part of a larger study where they compared predictive accuracy of affective forecasts between individuals diagnosed with major depressive disorder, those with bipolar disorder, and healthy individuals. Their results showed that the healthy control group was more accurate in predicting PA than NA, highlighting the importance of investigating forecasting accuracy for PA and NA separately. Altogether, research shows that humans are not perfectly accurate in making affective forecasts and are subjected to biases that distract them from making objective estimates.

It is of importance to highlight that people are generally bad at affective forecasting when we focus on the negative consequences of this inaccuracy. For example, Peters et al. (2014) suggest that people might become more risk-averse when they mistakenly predict the intensity and duration of anticipated NA about medical interventions. The researchers explain that people who overestimate the negative consequences of medical procedures may decide to refrain from these potentially life-enhancing/life-saving procedures. It is reasonable to assume that this might have lasting consequences on peoples' medical health, mental health, and/or quality of life (QoL) and underscores the associated negative consequences of the impact bias. Additionally, Loewenstein (2005) states that people who are asked to make decisions about these life-saving procedures are often already experiencing emotional distress about their current situation. These people overweigh their current affective state in making these decisions, taking their current emotional distress into account. Where the impact bias shows the negative consequences by stating that people overestimate their future negative feelings, Loewenstein (2005) argues that people, for example, fail to take simple preventative measures (e.g. flossing, medication, eating healthy food, etc.) because they are not able to correctly envision how they would feel if they were sick as a result.

Loewenstein (2005) states that it is the *projection bias*, described as a phenomenon in which we project our current affective state onto the future, that leads people to abstain from these preventative measures. A scenario of this bias is apparent when you are invited to go running with your friend the next day, however, you have just come home from a busy day and do not feel like running. Thus, because you are feeling tired now, you decline your friend's invitation to go running the next day, when in fact you could be feeling as good as new tomorrow after a good night of sleep. Your current state is affecting your decision for the future. Furthermore, research done by Meyvis et al. (2010) found that people's inaccurate forecasts are also partly due to the misremembering of past forecasts. Meaning that we do not learn how far off our prior predictions were as we fail to remember what exactly we predicted. As a result, future forecasts are unlikely to become more accurate, because we do not take into account the prediction error. This line of reasoning was further expanded upon in a study by Takano and Ehring (2024), who examined how humans make and update their affective forecasts with the use of an ecological momentary assessment (EMA) approach. They argue that people compare their current experience to their forecasted experience, referred to as the prediction error, and update their new forecast based on this difference. Furthermore, if people mistake their past forecast as being closer to their current experience, they will project their current experience onto the next forecast. By measuring participants' affective states and monitoring their affective forecasts, they showed that people use both the past forecast and current experience to predict their next emotional state. Participants exhibiting the projection bias showed more inaccuracies in their affective forecasts. However, when making predictions for events in the near future (minute-long forecasts), relying on one's current affective state proved to be less inaccurate.

The aforementioned negative consequences, as a result from inaccurate affective forecasts, might be mitigated by using statistical models instead of relying on human prediction. Human forecasting is a more general term to describe the process of making predictions by people, for example in fields ranging from the weather (e.g., Doswell, 2004) to the economy (e.g., Fritsche et al., 2020) to the outcome of an election (e.g., Ojo et al., 2019), among many other things. Certain statistical models have recently been employed in human forecasting, however, research reveals that these statistical models continue to make errors, and are sometimes just as accurate as human prediction (e.g. Abolghasemi et al., 2025; Grove et al., 2000). Whereas other research suggests that combining human prediction and statistical prediction could enhance forecasts, compared to predictions made by humans or statistical models alone (e.g. Ibrahim et al., 2020; Mellers et al., 2024). Thus, the results about whether statistical models outperform human forecasts are conflicting.

Within the area of affective forecasting, Takano and Ehring (2024) are the first to employ a statistical model in predicting people's affective states, namely the Kalman filter. At each step, the filter predicts the next affective state from all the available past estimates, then updates that prediction by incorporating the observed affect and weighing the resulting forecast error by the Kalman gain (see Method section or Takano and Ehring, 2024, for more detail). Unlike human forecasters, who often forget or misremember earlier predictions, the filter systematically uses every prior forecast as the past forecast is always available. In addition, it places the optimal weight on evaluating the current observed emotion which should protect against the projection bias that humans often show. In their study they compared participants' predictions to those made by the Kalman filter, and found that the forecasting accuracy of participants was less accurate when they employed the projection bias in hour-long forecasts (study 1 and study 2b). Thus, the Kalman filter's predictions outperformed those of the participants. However, in study 2a (minute-long forecasts) there was no difference found in predictive accuracy between the Kalman filter and the participants. Nevertheless, while Takano and Ehring (2024) use the Kalman filter to predict people's future emotional states from a remembering perspective, there has also been research on how emotions evolve in the area of affect dynamics.

Affect Dynamics

The study of affect and how it changes over time is referred to as affect dynamics (Kuppens, 2015). A method that is often used in measuring these dynamics is the EMA method (e.g., Dejonckheere et al., 2019; Krieke et al., 2016; McKone & Silk, 2022; Shiffman et al., 2008). In this method, also referred to as experience sampling method (ESM), participants can report on their experiences in the real world through their phone and are often asked to fill out multiple daily questionnaires over a set period. An advantage of this method is that it can capture moment-to-moment fluctuations in mood, context, and behavior (Moskowitz & Young, 2006; Shiffman et al., 2008). An example item that a participant might be asked in an EMA study is: "How happy are you feeling right now on a scale from 1-10?"

Since EMA methods are measuring participants' affect multiple times throughout the study with high frequency, the arising data is time series data as it is creating inherently time-dependent measurements. This temporal dependence, referred to as the autocorrelation, quantifies the correlation between a variable's current value and the value at a previous time point (Jebb et al., 2015). The autocorrelation is often referred to as emotional inertia, defined in the field of affect dynamics as "the degree to which a person's current emotional state can be predicted by his or her emotional state at a previous moment" (Kuppens et al., 2010, p. 985; Houben et al., 2015). Simply put, high emotional inertia reflects emotions that are carried over from one time point to the next (Kuppens & Verduyn, 2015).

Autocorrelation, or emotional inertia, is related to affective forecasting, as people take the previous affective state into account when making their forecast. However there is a nuance, in affective forecasting people often overestimate the impact of the current moment (as mentioned, the projection bias). Whereas, the emotional inertia captures the actual observed carry over from one moment to the next. Thus, whereas affective forecasts reflect subjective, often inflated predictions, autocorrelation/emotional inertia provides an empirical indicator of how much past affective states truly influence the next time point. It is of significance to investigate the level of emotional inertia in predicting how one feels in the future, as recent research has shown that emotional inertia can be linked to psychological well-being. More specifically, researchers found a relationship between emotions that are more inert and low psychological well-being, especially for NA (e.g., Houben et al., 2015; McKone & Silk, 2022; Maciejewski et al., 2023). Research by Maciejewski et al. (2023) further shows that emotional inertia is more strongly associated with NA, than with PA. In other words, people's negative emotions carry over more than positive emotions, likely making them more stable. Thus, we can argue that a statistical model that can directly capture these emotion dynamics of individuals, would be a good fit to explain affective forecasting data.

Autoregressive (AR) models capture this autocorrelation present in affective dynamics research. This model regresses the experienced affect at previous time points against experienced affect at the current time points (Jebb et al., 2015). A multilevel AR model is useful for capturing the between- and within-person differences in terms of how strongly affect measurements depend on each other over time (Schuurman, 2016). The between-person differences refer to how the affect dynamics of one individual differs from another, whereas the within-person differences refer to how a person's individual affect changes over time. The individual differences are modeled using random effects where every person gets their own set of parameters, including an autoregressive coefficient. Accordingly, a multilevel AR model allows us to model the data that arises from doing an EMA study on affective forecasting. Lastly, many recent studies have employed an autoregressive model when investigating affect dynamics (e.g. De Haan-Rietdijk, Gottman, et al., 2016; De Haan-Rietdijk, Kuppens, et al., 2016; Hamaker et al., 2018; Li et al., 2022; Pooseh et al., 2024).

Present Study

All in all, biases have been identified in the field of affective forecasting, and inaccuracies in predicting affect have been linked to negative consequences in, for example, the medical field. To address these limitations, statistical models have been proposed to enhance forecasting accuracy. While Takano and Ehring (2024) introduced a Kalman filter-based approach designed to emulate the human process of forecasting and updating anticipated emotions, we argue that a multilevel AR model more appropriately handles the temporal structure of EMA data. More specifically, the multilevel AR model directly captures the autocorrelation/emotional inertia present in repeated affective observations. Therefore, we expect that using this model instead of the Kalman filter will enhance the predictive accuracy of affective forecasts.

The present study aims to, in part, replicate the study done by Takano and Ehring (2024) by comparing the predictions made by the Kalman filter to those made by the multilevel AR model, rather than human judgments. Based on the reasoning above, we hypothesize that the multilevel AR model will outperform the Kalman filter when predicting people's affective states. Moreover, we will investigate whether there is a difference between the accuracy of the statistical models in predicting PA versus NA. We expect that the multilevel AR model will perform better in predicting negative emotions than the Kalman filter, due the aforementioned correlation between negative emotions and emotional inertia.

Methods

The aim of this study is to capture affective ratings of participants in an EMA design to compare the affective forecasts made by the Kalman filter and those made by the multilevel AR model to discern which model better fits the data. This study is building on the study conducted by Takano and Ehring (2024). Similar to their study, we also employ the EMA method to collect data on affective forecasting and we make use of the Kalman filter, however, not the human predictions. What further sets our study apart is that we are extending their study by including a two-step ahead prediction as well as an interval prediction. By comparing the predictions to the level of emotions they later reported experiencing, we measured prediction accuracy or prediction error (i.e, how well their predicted emotions matched their actual emotions). Additionally, the use of point and interval predictions was important to assess uncertainty regarding self-ratings and predicted emotions. However, these two-step ahead and interval predictions will not be used for the particular research question investigated in this paper, as the focus is on comparing the one-step ahead point predictions of both models for PA and NA.

Participants

Participants predict and report their real-time emotional experiences, as done in EMA studies, and these responses are compared with each other to determine forecasting accuracy (Takano & Ehring, 2024). We aimed to replicate and extend the study of Takano and Ehring (2024) and accordingly aimed for a sample size of at least 68 participants. However, the final sample consisted of 30 students from the University of Amsterdam (20 women, 10 men) with a mean age of 19.97 (*SD* = 1.83) who received course credits for participating (see Table 1).

Table 1

Sample Characteristics

Variable	Category	*n*	%	*M*	*SD*
Sex	Female	20	66.7	-	-
	Male	10	33.3	-	-
Age		-	-	19.97	1.83

Note. *N* = 30. *n* = number of participants in category. *M* = mean. *SD* = standard deviation.

Procedure

Data collection methods

The study was advertised via flyers on campus, social media, and The University Research Pool. This approach makes the sample that we collected a convenience sample. To be eligible for this research project, students had to (a) own a smartphone, (b) understand the English language and (c) not have been diagnosed with depression or anxiety. Data was collected using EMA via the m-Path app (Mestdagh et al., 2023), which allows participants to complete short questionnaires on their phones while going about their daily activities. The study was ethically approved (FMG-12534_2025) and individuals provided their informed consent before participating in the EMA study, which was received through a Qualtrics questionnaire. This questionnaire also included instructions for the EMA study and assessed participants demographic characteristics. Upon enrollment participants received a link to the m-Path questionnaire via email.

Participants received five beeps per day over a 14-day period to complete the EMA questionnaire. These were sent at fixed times, starting at 9:00 am, and then every three hours until 9:00 pm, thus in total the questionnaire can be filled out five times per day. The questionnaire takes around 5-10 minutes to complete. Upon receiving a notification, participants have a 30-minute time window to respond and are reminded to fill out the questionnaire after the first 15 minutes. The prompt expires after 30 minutes to ensure that the time between the questionnaires remains great enough to make affective forecasts for the next beep. The items for each emotion, concerning their current and future affect, were presented in the same order for all participants, however, the order in which each emotion was presented was randomized at each prompt.

Materials

Ecological Momentary Assessment (EMA)

This study utilizes an EMA design and is therefore a longitudinal study. We asked participants to make predictions about their emotions at two different time horizons: three hours and six hours into the future. The questionnaires assess four emotions: happy, relaxed, sad, and anxious. These emotions were chosen as PA and NA indicators in line with the study of Takano and Ehring (2024), where they also used these four emotions. The emotions happy and relaxed together form PA, and the emotions sad and anxious are combined to form NA. To further justify the choice of these emotions, other studies (e.g., Maciejewski et al., 2023) also use happy and relaxed for measuring PA, and sad and anxious for measuring NA, although other emotions are used as well. Each emotion contains six questions assessing participants' current affective state and their forecasted level of affect. The Visual Analog Scale (VAS) was used to assess participants' level of affect. Weigl and Forstner (2021) described this scale as a linear continuum assessing a psychological construct, accompanied with a written statement on each end indicating what the numbers signify. Participants are first asked to rate their current experience of the assessed emotion (e.g. "How happy do you feel at the moment?") on a VAS from 0 (*not at all*) to 100 (*extremely*). Subsequently, they are asked to make a point prediction for their affective state at two following time points with a three hour interval. The point prediction assesses their best guess of the emotion under assessment (e.g. "My best guess is that I will be … happy") on a VAS from 0 (*not at all*) to 100 (*extremely*). The remaining three questions focus on the credible interval prediction that participants are required to make, however, this is not relevant for the research question under investigation.

Models

As the aim of the current study is to investigate the accuracy of two statistical models, the Kalman filter and the multilevel AR model, we will describe these models in more detail. For both models we compute the one-step prediction (three hours ahead) based on previous data.

Kalman filter. The Kalman filter is used in many fields and is described as a model that uses the past and the present to make estimations for the future (Khodarahmi & Maihami, 2022). Moreover, the Kalman filter consists of a process of predicting and updating, in which predicting is based on analysing past observations and updating involves evaluating prior estimations with current observations. In relation to affective forecasting, the Kalman filter is a statistical model that predicts people's future affective states and continuously updates forecasts by analysing the prediction error (the difference between the actual and predicted affective state). Takano and Ehring (2024) state that this updating is controlled by the Kalman gain (k_t), which determines how strongly the past forecast and current emotion have an impact on the next forecast. See the researchers' article for a more detailed explanation of the Kalman filter (Takano & Ehring, 2024).

Multilevel Autoregressive model. Autoregressive (AR) models are regression models that use previous time points of affect as the predictors for current time points. In an AR model with only one predictor, one regresses the values of the time series against the values of the time series at the previous time point, called lag 1 (Jebb et al., 2015). The regression equation of an autoregressive model with one predictor can be written as follows:

$$y_t = \alpha + \phi(y_{t-1}) + \varepsilon_t \tag{1}$$

where α indicates the intercept, ϕ the autoregressive coefficient of the autoregressive model, y_{t-1} refers to the previous time point, and ε_t describes the innovation. The intercept α quantifies an individual's baseline affective state. The autoregressive coefficient ϕ is also indicated as the autocorrelation. As this value increases, emotions are carried over more strongly from moment to moment, and it takes longer before for affective states to return to their baseline. Lastly, the innovation or random shock ε_t refers to unexpected factors that influence the affective states of individuals, and assumes a normal distribution ($N(0, \sigma^2)$; Schuurman, 2016).

However, an AR(1) model does not take into account that people might experience different emotional processes, as this model assumes the same model parameters for every individual (Schuurman, 2016). In contrast, a multilevel AR model does not assume the same model parameters and allows the model parameters to vary across individuals. This is a useful model, because much of research is interested in discerning why and how people differ from each other. People experience emotions in different manners, and the multilevel AR model can therefore account for these individual differences. The equation for the multilevel AR(1) model takes the form of:

$$y_{t,i} = \alpha_i + \phi_i(y_{t-1,i}) + \varepsilon_{t,i} \tag{2}$$

where α_i is the intercept for subject i, ϕ_i is the regression coefficient of the autoregressive model for subject i, $y_{t-1,i}$ refers to the previous time point of subject i, and ε_t refers to the innovation of subject i, where the innovation has a normal distribution ($N(0, \sigma^2)$; Schuurman, 2016).

Data Analysis

Data Preprocessing and Diagnostics

Data preprocessing was done in R (v4.4.2; R Core Team, 2024; see Appendix A for packages used) and excluded participants with less than 30% compliance (i.e. less than 21 beeps) from the dataset to ensure that our data is in line with the data from Takano and Ehring (2024). The overall completion rate of the questionnaires in our dataset is 72.4% (50.7 questionnaires out of 70 sent questionnaires) with a standard deviation of 19.9 (13.9 questionnaires). The individual compliance ranges from 20 to 67 questionnaires. Moverover, 29 of the total number of participants ($N = 30$) filled out 21 or more questionnaires. After preprocessing the data, variables were computed for the prediction error of PA and NA, and the type of model that made the predictions.

Analysis

The variables PA error and NA error are expressed as the absolute value of the difference between the experienced affect and the predicted affect, as we are interested in the magnitude of the error and not the direction. ModelType has two categories: the multilevel AR model and the Kalman filter.

First and foremost, some descriptive statistics of the variables will be calculated, such as the mean, standard deviation, range, and the autocorrelation. Furthermore, a linear mixed models analysis will be done to test our main hypotheses. Linear mixed models are able to account for individual differences in the prediction error, using random effects and fixed effects, and are often used in analyzing longitudinal data (Liu, 2016). The dependent variable is the prediction error, PA error or NA error. The fixed effects variable is which model is making the predictions, the Kalman filter or the multilevel AR model, using dummy coding (0 = multilevel AR model, 1 = Kalman filter). The random effects grouping factor is the participant (ParticipantID). We include a random effect for the intercept to account for the fact that some participants might have higher or lower prediction error on average. The linear mixed models analysis will be performed twice, once for comparing the Kalman filter and the multilevel AR model on PA and once for NA. The linear mixed models analysis will be done in JASP (Version 0.18).

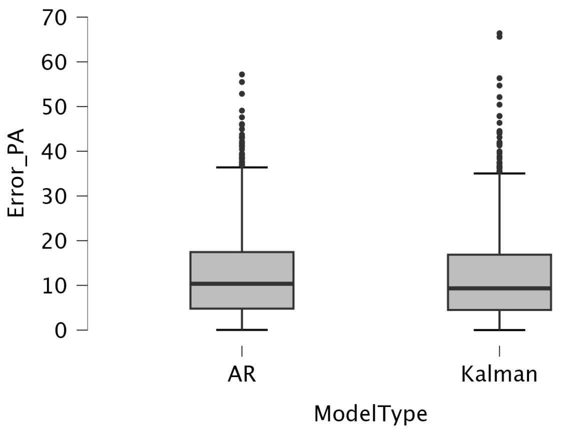
Results

Descriptive Statistics

The mean, standard deviation, and range for the dependent variables (PA error & NA error) are reported, while also splitting them according to the type of model, multilevel AR model and the Kalman filter. See Figure 1 for a boxplot of PA error for the multilevel AR model and the Kalman filter, and Figure 2 for a boxplot of NA error. The error in PA prediction for the multilevel AR model shows a mean of 12.37 (*SD* = 9.69) and for the Kalman filter it shows a mean of 11.73 (*SD* = 9.66). For NA, the prediction error of the multilevel AR model has a mean of 11.11 (*SD* = 9.60) and for the Kalman filter it shows a mean of 33.75 (*SD* = 22.48). These statistics indicate substantial variability in prediction error, especially for the Kalman filter's predictions for NA. Moreover, the range for these variables is also quite high (see Figure 1 and Figure 2), showing that there is a large difference between the highest and lowest value of the prediction error. Moreover, the autocorrelation (combined for PA and NA) ranges from -0.27 to 0.27, with a mean of almost 0 (*SD* = 0.11).[1] This indicates that current affective states do not strongly carry-over to the next moment for the composite score of PA and NA. See Figure 3 for a histogram of the autocorrelation for all participants.

Figure 1

Boxplot of Prediction Error in PA against ModelType

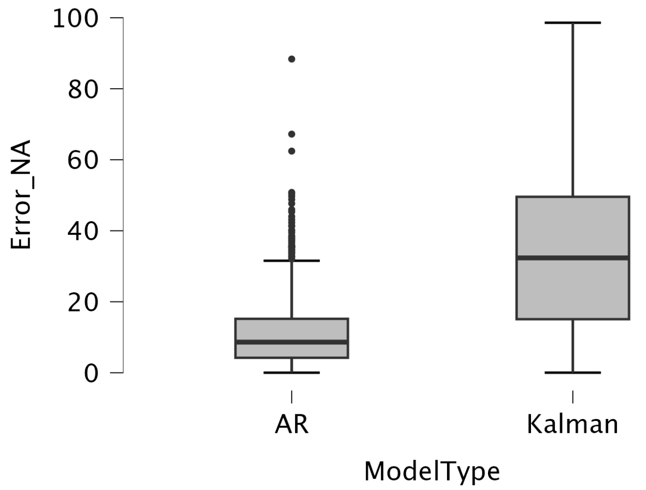


Note. M_{AR} = 12.37. M_{Kalman} = 11.73. The values of the prediction error are centered around the lower values, as also shown by the means. Moreover, it shows that there is high variation in prediction error as the boxplot for the multilevel AR ranges from 0.04 to 57.18 and for the Kalman filter it ranges from 0.00 to 66.40.

[1] *M* = 0.00 is the rounded value, actual value is *M* = -0.0002364

Figure 2

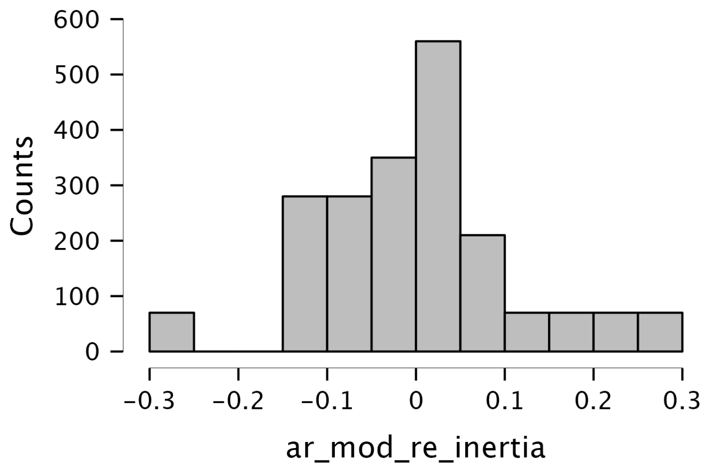
Boxplot of Prediction Error in NA against ModelType



Note. M_{AR} = 11.11. M_{Kalman} = 33.75. This shows that there is more variation in prediction error for the Kalman filter than for the multilevel AR model. Moreover, the boxplot for multilevel AR shows that most values are below 40, with some outliers present. Lastly, it shows that the prediction error for the multilevel AR ranges from 0.02 to 88.37 and for the Kalman filter it ranges from 0.03 to 98.59.

Figure 3

Histogram of the Autocorrelation for Every Participant



Note. *M* = -0.0002364. *SD* = 0.11

Inferential Statistics

Linear Mixed Models Analysis for Positive Affect (PA)

To test the hypothesis that the multilevel AR model is a better predictor of the level of affect than the Kalman filter, we performed a linear mixed models analysis for the prediction error of PA. The results showed that the fixed effect estimate for the effect for model type was non-significant at an alpha level of .05 (b = -0.64, t(2306.71) = -1.67, p = 0.095). This implies that there was no significant difference in predictive accuracy between the multilevel AR model and the Kalman filter. Consequently, the estimated marginal mean of prediction error for the multilevel AR model (b_0) is 12.22 (95% CI [11.00, 13.44]). The estimated marginal mean for the Kalman filter is 11.58 (95% CI [10.37, 12.80]), which is the difference between the estimated marginal mean for the multilevel AR model and the slope estimate. These values indicate the average prediction error for PA when either using the multilevel AR model or the Kalman filter as the predictor. Moreover, it shows that the averages are relatively proximate in value and therefore we cannot confidently say that there is a difference in the predictive accuracy of the models. The results of the random effects show us that the variance of the intercept for all individuals is 8.85 (SD = 2.97). That is to say, participants varied by 2.97 units in prediction error around the overall mean. See Table 2 for the fixed effects of the linear mixed model analysis and see Figure 4 for the plot of Error_PA against the type of model.

Table 2

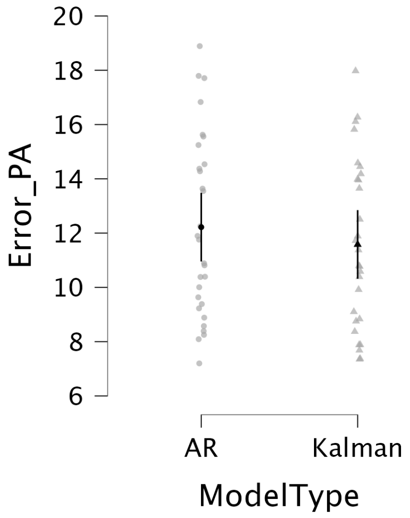
Fixed Effects of Linear Mixed Model for PA

	Estimate	*SE*	*t*	*p*
(Intercept)	12.22	0.62	19.66	<.001
ModelType_Kalman	-0.64	0.38	-1.67	0.095

Note. This analysis was performed with dummy coding.

Figure 4

Error of PA Plotted Against the Type of Model

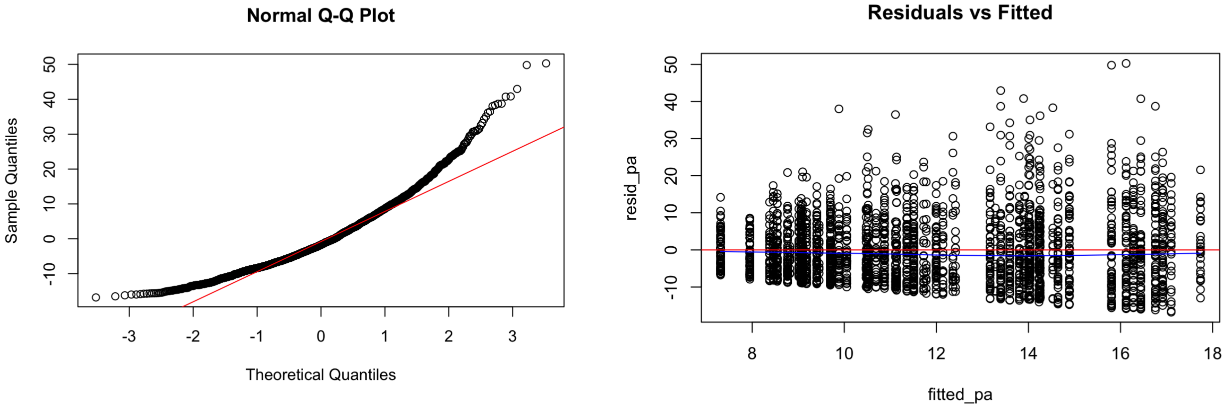


Assumptions Positive Affect

Furthermore, we checked the assumptions for the linear mixed models analysis of PA. Figure 5 shows the Q-Q Plot of the residuals and reveals that the normality assumption is violated. There are clear deviations from the diagonal line drawn across. Moreover, Figure 5 also shows the residuals versus the fitted values plot and reveals slight violation of the homoscedasticity assumption as the dots are scattered more away from the line when the value of the x-axis (fitted_pa) increases. Lastly, linearity is met as the dots are scattered in a linear line around the horizontal line.

Figure 5

Q-Q Plot and Residuals vs. Fitted Plot



To account for the violated assumptions of the linear mixed models analysis, a generalized linear mixed model analysis was performed, as this allows us to handle non-normal data (Ng & Cribbie, 2017). It allows us to model right-skewed data, with the use of a gamma distribution. The results showed that the parameter estimates were similar in direction. Moreover, the *p*-values were similar in terms of significance as well, showing a non-significant effect for model type (see Table 3). The assumption checks of the generalized linear mixed model are shown in Figure 6, and indicates that all assumptions are met. Homoscedasticity and linearity are met as plot B (Homogeneity of Variance) shows that all the data points are equally scattered around the horizontal line. Moreover, normality is met as plot C (Uniformity of Residuals) shows that all the data points are scattered closely around the diagonal line. Moreover, plot D (Normality of Random Effects) shows an approximately normal distribution for the random effects (ParticipantID).

Figure 6

Assumption Plots of the Generalized Linear Mixed Model

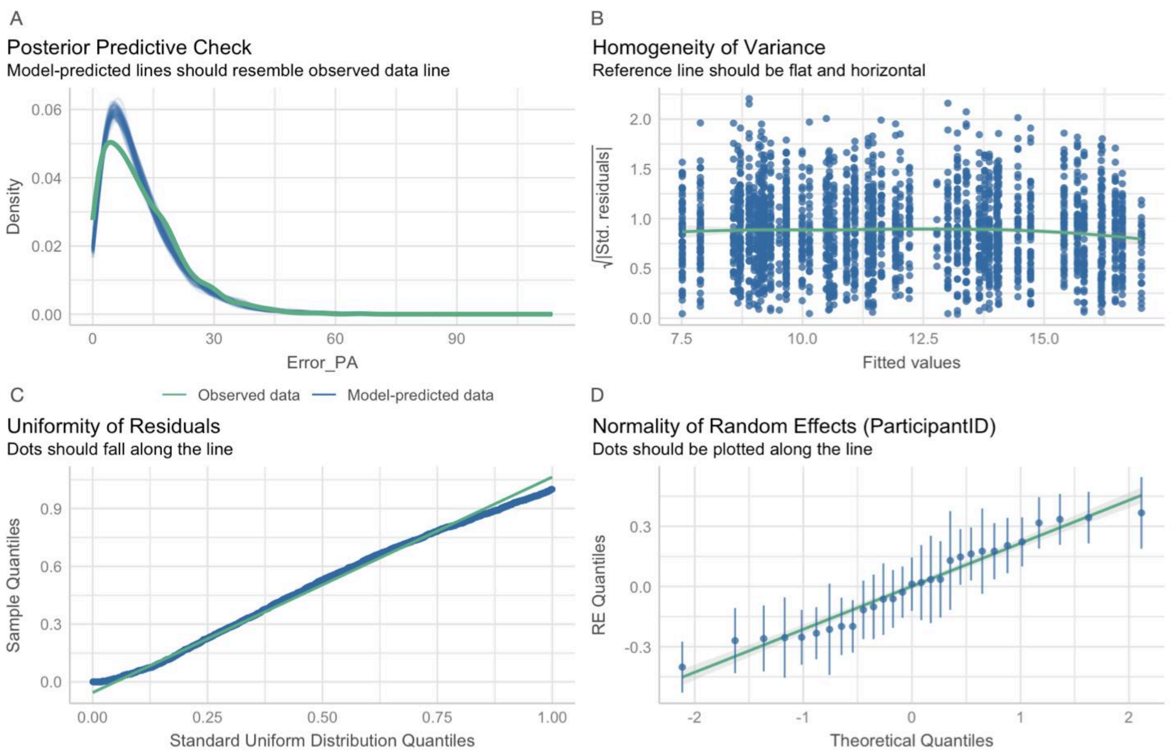


Table 3

Fixed Effects of Generalized Linear Mixed Model for PA

	Estimate	*SE*	*t*	*p*
(Intercept)	2.47	0.05	46.15	<.001
ModelType_Kalman	-0.05	0.03	-1.40	0.163

Note. This analysis was performed with dummy coding.

Linear Mixed Models Analysis for Negative Affect (NA)

The same analysis was performed for prediction error of NA to see whether the multilevel AR model is a better predictor of the level of affect than the Kalman filter. These results show that the effect for model type is significant at an alpha level of .05 (b = 22.64, t(2306.87) = 34.65, $p < .001$). Indicating a statistically significant difference in predictive accuracy between the multilevel AR model and the Kalman filter. The estimated marginal mean of prediction error for the multilevel AR model is 10.85 (95% CI [8.24, 13.46]), and shows the average prediction error when using the multilevel AR model. The estimated marginal mean for the Kalman filter is 33.49 (calculated by adding the estimated marginal mean for the multilevel AR and the slope estimate), and shows the average prediction error in predicting NA for the Kalman filter (95% CI [30.88, 36.11]). The results of the random effects show us that the variance of the intercept for all individuals is 44.54 (SD = 6.67), and suggests that there is a lot of variability of prediction error for each individual. Thus, participants varied by 6.67 units in prediction error around the overall mean. See Table 4 for the fixed effects of the linear mixed model analysis and see Figure 7 for the plot of NA error against the type of model.

Table 4

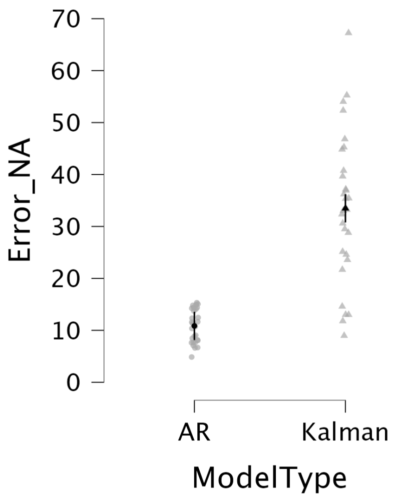
Fixed Effects of Linear Mixed Model for NA

	Estimate	*SE*	*t*	*p*
(Intercept)	10.85	1.33	8.14	<.001
ModelType_Kalman	22.64	0.65	36.65	<.001

Note. This analysis was performed with dummy coding.

Figure 7

Error of NA Plotted Against the Type of Model



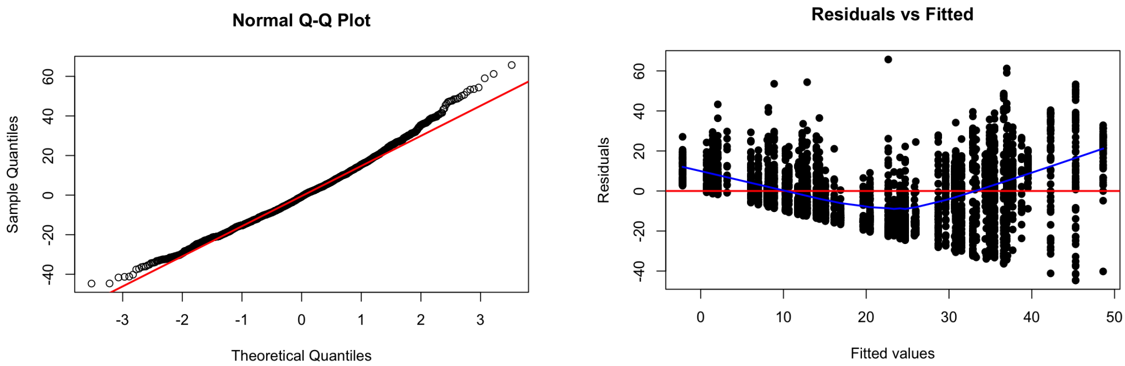
Note. There is a much higher variability in NA prediction error for the Kalman filter.

Assumptions Negative Affect

We additionally checked the assumptions for the linear mixed models analysis of NA. Figure 8 shows the Q-Q Plot of the residuals and reveals a few outliers at the end of the diagonal line. However, these deviations are not too drastic and therefore this is not a major problem for our analysis on prediction error of NA. Moreover, Figure 8 shows the Residuals versus the Fitted values plot and reveals that linearity and homoscedasticity is violated. This can be observed by the data points, which are scattered farther away from the horizontal line as the value of the x-axis increases. Additionally, the data points take on the form of a parabolic line more than a linear line.

Figure 8

Q-Q Plot and Residuals vs. Fitted Plot



To account for the violation of the homoscedasticity and the linearity, a generalized linear mixed model analysis was performed (Ng & Cribbie, 2017). The results of this analysis furthermore showed that the parameter estimates were similar as the linear mixed model analysis, and the *p*-values remained significant (see Table 5). Figure 9 shows assumption plots and indicates that all assumptions are met. Plot B (Homogeneity of Variance) shows that homoscedasticity and linearity are met, as all data points are scattered equally around the horizontal line. In addition, plot C (Uniformity of Residuals) shows that all data points are closely scattered around the diagonal line. Lastly, plot D (Normality of Random Effects) shows an approximately normal distribution for the random effects (ParticipantID).

Figure 9

Assumption Plots of the Generalized Linear Mixed Model

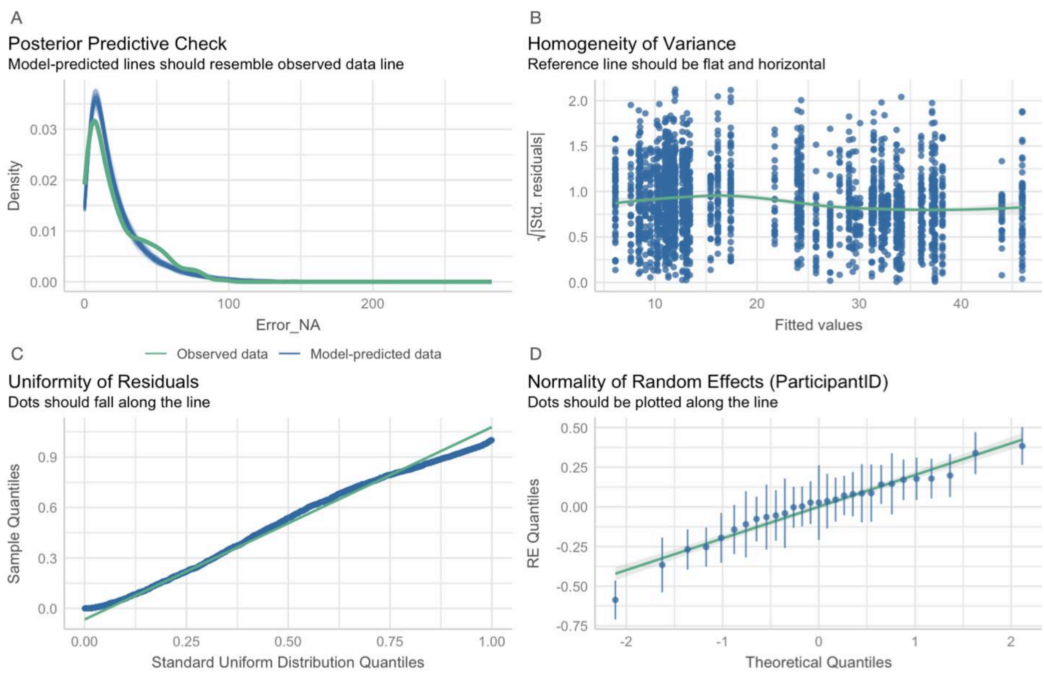


Table 5

Fixed Effects of Generalized Linear Mixed Model for NA

	Estimate	*SE*	*t*	*p*
(Intercept)	2.40	0.05	46.78	<.001
ModelType_Kalman	1.05	0.04	29.87	<.001

Note. This analysis was performed with dummy coding.

Discussion

The central aim of the current study was to, in part, replicate the research done by Takano and Ehring (2024) and extend their research by including a multilevel AR model to predict people's future affective states. Takano and Ehring (2024) found that the Kalman filter outperformed participants' predictions in study 1 and study 2b (hour-long forecasts), but no differences were found in study 2a (using minute-long forecasts). More specifically, in study 2a they found a significant difference between participants and the Kalman filter for relative errors, but not for absolute errors. Which suggests an optimism bias of participants, as stated by the authors. In study 2b they found the opposite, showing a significant difference for absolute errors between the participants and the Kalman filter, but not for relative errors.

In our study, however, we compared the predictive accuracy of the Kalman filter with the performance of the multilevel AR model and hypothesized that the latter will outperform the former, due to the frequent use of multilevel AR models in investigating affect dynamics. Moreover, considering that the multilevel AR model is able to capture the amount of carry-over and some research has shown that negative emotions present more emotional inertia (Maciejewski et al., 2023), we expected that the multilevel AR model will specifically outperform the Kalman filter in predicting negative emotions. Our results partly support our hypothesis that the multilevel AR model will outperform the Kalman filter in predictive accuracy. No significant difference was found in the predictive accuracy between the statistical models when analyzing the predictions for PA. However, a significant difference between the Kalman filter and the multilevel AR model showed that the multilevel AR model is better at predicting people's negative emotions than the Kalman filter. Therefore, our hypothesis that the multilevel AR model is a better predictor for NA was supported.

The finding that there was a significant difference in examining NA, but not in examining PA, might be explained by emotional instability, described as the amount of change in emotions from moment to moment (Houben et al., 2015). For example, if PA is more unstable it would be harder to make forecasts on these emotions. However, results about the instability of positive and NA are inconclusive. Spindler et al. (2016) noted that there was higher instability present in PA than in NA, meaning, people's positive emotions show larger changes from moment to moment than negative emotions. In contrast, Houben et al. (2015) showed that instability is positively related to NA, indicating that as the magnitude of people's changes in emotions increases, NA increases alongside. Therefore, further research should be done on the relationship between instability and PA and NA to investigate whether this difference has an influence on the predictive accuracy on PA and NA of statistical models.

Furthermore, it is important to evaluate the predictive accuracy between different models, not only in our study but also in other fields. Much research has been devoted to comparing how well different models fit the data and how accurate they are in predicting unobserved data. One way to do this is with the use of cross-validation, which estimates the predictive accuracy of models. In this process, data is split into a number of training and test sets. Then the model is fitted to the train sets and the predictive performance is evaluated by comparing it to the assigned test set (Bulteel et al., 2018). By repeating this step, one gets a more accurate estimate of the prediction error. In the study of Bulteel et al. (2018), they compared the predictive accuracy of the simpler autoregressive model (AR) to the vector autoregressive model (VAR) using cross-validation. Their overarching message was that it is of good value to evaluate the predictive accuracy of different models to avoid models from overfitting the data. Moreover, they found that the more complex multilevel VAR(1) model did not consistently outperform the simpler multilevel AR(1) model. In our analysis, the multilevel AR model showed comparable predictive performance to the simpler Kalman filter model for PA, but outperformed this simpler model for NA. We, however, used only a train/test split where the data is only tested once, instead of cross-validation. As a consideration, future research could look at comparing the predictive accuracy of the multilevel AR model and the Kalman filter with the use of cross-validation. Moreover, the VAR model was not utilized in this study and could potentially be an even better fit to the data. Therefore, this could also be useful to investigate in affective forecasting in future research.

Limitations and Strengths

Despite the many new insights this study has provided, it is also confronted with various limitations. Firstly, we employed a convenience sample to investigate our research topic. Because of this, our sample mainly consisted of students from the University of Amsterdam, who are most likely similar in age and education level. The use of such a sample raises the question whether the collected sample is truly representative of the population, or if it reflects a skewed representation. Moreover, the overall sample size was already quite low, as we only had access to the data of 30 participants. Therefore, we might have too little data to obtain reliable parameter estimates of the multilevel AR model or the Kalman filter. For instance, the non-significant difference in predictive accuracy for PA may result from the multilevel AR model underfitting the data, capturing a relationship that is too simplistic of the truth, due to the limited amount of data. A sufficient sample size is needed to obtain more accurate predictions (Simon & Aliferis, 2024). As too little data might be associated with larger confidence intervals and reflects uncertainty, and therefore less accurate predictions.

Besides the small sample size, many data points were also missing due to missed prompts by the participants. In our study 579 questionnaires were not filled out from the 2100 sent prompts (27.6%). Moreover, due to the dependency of measurements in our study—as we need both the previous forecast and actual experience to compute the prediction error—this missingness might further reduce the amount of usable data to evaluate model performance. Therefore, this might contribute to the argumentation of too little data for accurate predictions. Nevertheless, this missingness is also related to examining possible observable trends on missing data. For example, participants might systematically fail to complete the questionnaire when they are in a low mood, potentially leading to data missing at random or data missing not at random (Fritz et al., 2024). As we did not perform an analysis on missinginess in our study, we have to be cautious of immediately disregarding this missing data.

Lastly, some limitations that were discussed by Takano and Ehring (2024) could be of influence on the research design we employed. They mention that asking participants to first rate their current affect and consequently ask them to make affective forecasts could encourage them to act on their projection bias, entailing that they use their current affective state to predict their future affective state. As we also used this approach with the goal of replicating their study, this limitation could also apply to the current investigation. Therefore, further research could look at randomizing these questions. Another limitation proposed by Takano and Ehring (2024) is the lack of qualitative data that examines, for example, participants' knowledge about upcoming events and how this might influence predictive accuracy. There was little investigation to which future events people anticipated, however, much research on affective forecasting has focused on a specific future event to analyse people's predictive accuracy (e.g. Dunn et al., 2003; Hughes et al., 2022; Meyvis et al., 2010). Our investigation did not link participants' forecasts to a specific event, as we did not analyse what happens in our one-step predictions.

While there were several limitations present, this investigation also provides valuable knowledge and insight. This study might be the first step in improving the accuracy of affective forecasts, as it investigated which model better predicted the participants' affective states. Moreover, we compared the multilevel AR model and the Kalman filter not only statistically, but also theoretically, by emphasising the differences in theoretical framework and evaluating the predictive accuracy of both models. Furthermore, another strength of our study is that we made use of EMA data, which is a powerful tool to assess participants' affective states in a real-life setting. EMA data is expected to have strong ecological validity and to provide more reliable data (Shiffman et al., 2008).

Future Directions and Implications

To correct for the noted limitations present, we put forth directions for future research. A greater sample is necessary to increase the statistical power of the current investigation, and to avoid capturing a relationship that is either too simplistic or too complex due to limited observations. Moreover, as mentioned, 27.6% of our data is missing due to missed prompts. To analyse if any observable trend is present in the missing data in our study, a Little's test could be performed (Little, 1988). Little's MCAR test examines the null hypothesis that the data is missing completely at random (MCAR). Consequently, rejecting the null hypothesis, means that the missing data is likely not MCAR, and thus either missing at random (MAR) or missing not at random (MNAR). Furthermore, a qualitative design could be implemented to account for the fact some participants might provide insightful rationale for their predictive forecasts. For instance, we could have added an open-ended question asking participants what they think will happen in three hours, and consequently ask them later to state what is actually happening. For example, it could be the case that people's inaccurate forecasts are due to future events they did not anticipate. This way we can take into account the uncertainty people have about future events.

The main takeaway from analysing the results is that the multilevel AR model outperformed the Kalman filter in making affective forecasts when we focus on NA, but not for PA. This study expands our knowledge on the use of statistical models in the field of affective forecasting. Where Takano and Ehring (2024) were the first to use a statistical model in predicting future affect, the current paper uses a model that is closer aligned to the theory of how emotions develop over time. This research contributes to our knowledge, as statistical models might enhance decision-making since research shows that humans are subjected to many biases when making affective forecasts. Moreover, the negative consequences of biased forecasts have also been stated, and by deepening our knowledge on how people make affective forecasts and how these models are able to contribute to enhance the predictive accuracy, we can make the first step to mitigate these negative consequences. Correspondingly, further research could focus on investigating the predictive accuracy of forecasts when predictions of statistical models are combined with human affective forecasts. However, some people might not embrace the idea of using statistical predictions for affective forecasts. A term that can be referred to as algorithmic aversion (Jussupow et al., 2020). This resistance to the use of algorithms might block the path to the effective use of the multilevel AR model in affective forecasting. Mahmud et al. (2022) state in their systematic literature review that there are several factors that influence people's aversion to algorithms, with the four main categories being: algorithm factors, individual factors, task factors, and high-level factors (see Mahmud et al., 2022). For example, if an older person was presented with an unfamiliar algorithm and was told that this algorithm would predict their feelings, they might not be very accepting of this algorithm. However, this is a small example and several more factors, within these main categories, are present that influence whether or not people are aversive towards algorithms. This might thus stand in the way of using affective forecasts made by the multilevel AR model or the Kalman filter.

Another important consideration is how affective forecasting interacts with mental health problems, such as depression. In a meta-analysis on the relationship between affective forecasting and psychopathology, Rizeq (2024) showed that in people suffering with mental health disorders, affective forecasts showed increased negativity and these predictions overestimated the amount of NA actually experienced. This implies that treatment of mental health disorders (e.g. depression) might focus on addressing these biases in affective forecasting, with the use of statistical predictions.

Conclusion

To conclude, the current investigation is important in reviewing how different statistical models operate to predict future emotional states. Furthermore, comparing, in particular, the multilevel AR model with the Kalman filter gives us insight into how emotions might develop over time. This is an important topic as research continues to show biases present in human affective forecasting, where statistical models might overcome these biases.

References

Abolghasemi, M., Ganbold, O., & Rotaru, K. (2025). Humans vs. large language models: Judgmental forecasting in an era of advanced AI. *International Journal of Forecasting*, *41*(2), 631–648. https://doi.org/10.1016/j.ijforecast.2024.07.003

American Psychological Association. (2018). *APA Dictionary of Psychology*. https://dictionary.apa.org/affective-forecasting

Bulteel, K., Mestdagh, M., Tuerlinckx, F., & Ceulemans, E. (2018). VAR(1) based models do not always outpredict AR(1) models in typical psychological applications. *Psychological Methods*, *23*(4), 740–756. https://doi.org/10.1037/met0000178

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*(1), 217–241. https://doi.org/10.1007/s11336-014-9417-x

De Haan-Rietdijk, S., Kuppens, P., & Hamaker, E. L. (2016). What's in a day? A guide to decomposing the variance in intensive longitudinal data. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00891

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour*, *3*(5), 478–491. https://doi.org/10.1038/s41562-019-0555-0

Doswell, C. A. (2004). Weather Forecasting by Humans-Heuristics and Decision Making. *Weather and Forecasting*, *19*(6), 1115–1126.

Dunn, E. W., Wilson, T. D., & Gilbert, D. T. (2003). Location, location, location: the misprediction of satisfaction in housing lotteries. *Personality & Social Psychology Bulletin*, *29*(11), 1421–1432.

Fox, J., Weisberg, S. (2019). *An R Companion to Applied Regression*, 3rd Edition. Sage.

Fritsche, U., Köster, R., & Lenel, L. (2020). *Futures past : economic forecasting in the 20th and 21st century*. Peter Lang. https://doi.org/10.3726/b16817

Fritz, J., Piccirillo, M. L., Cohen, Z. D., Frumkin, M., Kirtley, O., Moeller, J., Neubauer, A. B., Norris, L. A., Schuurman, N. K., Snippe, E., & Bringmann, L. F. (2024). So You Want to Do ESM? 10 Essential Topics for Implementing the Experience-Sampling Method. *Advances in Methods and Practices in Psychological Science*, *7*(3). https://doi.org/10.1177/25152459241267912

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*(1), 19–30. https://doi.org/10.1037/1040-3590.12.1.19

Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the Frontiers of Modeling Intensive Longitudinal Data: Dynamic Structural Equation Models for the Affective Measurements from the COGITO Study. *Multivariate Behavioral Research*, *53*(6), 820–841. https://doi.org/10.1080/00273171.2018.1446819

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*(4), 901–930. https://doi.org/10.1037/a0038822

Hughes, C. D., King, A. M., Bailey, K., Alba, M. C., Hoelscher, E., & Rizvi, S. L. (2022). How Will You Feel on Valentine's Day? Affective Forecasting and Recall Biases as a Function of Anxiety, Depression, and Borderline Personality Disorder Features. *Journal of social and clinical psychology*, *41*(5), 491–516. https://doi.org/10.1521/jscp.2022.41.5.491

Ibrahim, R., Kim, S., & Tong, J. (2020). Eliciting Human Judgment for Prediction Algorithms. *ERN: Statistical Decision Theory; Operations Research (Topic)*. https://doi.org/10.2139/ssrn.3606633.

Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: Examining and forecasting change. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00727

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards Algorithms? A comprehensive literature Review on Algorithm aversion. *European Conference on Information Systems*.

Khodarahmi, M., & Maihami, V. (2022). A Review on Kalman Filter Models. *Archives of Computational Methods in Engineering : State of the Art Reviews*, *30*(1), 727–747. https://doi.org/10.1007/s11831-022-09815-7

Krieke, L. V. D., Jeronimus, B. F., Blaauw, F. J., Wanders, R. B. K., Emerencia, A. C., Schenk, H. M., Vos, S. D., Snippe, E., Wichers, M., Wigman, J. T. W., Bos, E. H., Wardenaar, K. J., & Jonge, P. D. (2016). HowNutsAreTheDutch (HoeGekIsNL): A crowdsourcing study of mental symptoms and strengths. *International Journal of Methods in Psychiatric Research*, *25*(2), 123–144. https://doi.org/10.1002/mpr.1495

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional Inertia and Psychological Maladjustment. *Psychological Science*, *21*(7), 984-991. https://doi.org/10.1177/0956797610372634

Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Review*, *7*(4), 297–300. https://doi-org.proxy-ub.rug.nl/10.1177/1754073915590947

Kuppens, P., & Verduyn, P. (2015). Looking at Emotion Regulation Through the Window of Emotion Dynamics. *Psychological Inquiry*, *26*(1), 72–79. https://doi.org/10.1080/1047840X.2015.960505

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). “lmerTest Package: Tests in Linear Mixed Effects Models.” _Journal of Statistical Software_, *82*(13), 1-26. doi:10.18637/jss.v082.i13 <https://doi.org/10.18637/jss.v082.i13>

Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. https://doi.org/10.2307/2290157

Li, Y., Wood, J., Ji, L., Chow, S.-M., & Oravecz, Z. (2022). Fitting Multilevel Vector Autoregressive Models in Stan, JAGS, and Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(3), 452–475. https://doi.org/10.1080/10705511.2021.1911657

Liu, X. (2016). *Methods and applications of longitudinal data analysis*. Academic Press is an imprint of Elsevier. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1059445

Loewenstein, G. (2005). Projection Bias in Medical Decision Making. *Medical Decision Making*, *25*(1), 96–105. https://doi.org/10.1177/0272989X04273799

Lüdecke D (2024). _sjPlot: Data Visualization for Statistics in Social Science_. R package version 2.8.17, <https://CRAN.R-project.org/package=sjPlot>

Maciejewski, D. F., Roekel, E. V., Ha, T., France, K. D., Lin, L., Lennarz, H., Trompetter, H., Meeus, W., Lichtwarck-Aschoff, A., Branje, S., Hollenstein, T., & Verhagen, M. (2023). Beyond main effects? Affect level as a moderator in the relation between affect dynamics and depressive symptoms. *Journal of Emotion and Psychopathology*, *1*(1), 356-372. https://doi.org/10.55913/joep.v1i1.52

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion.

Technological Forecasting & Social Change, *175*. https://doi.org/10.1016/j.techfore.2021.121390

McKone, K. M. P., & Silk, J. S. (2022). The Emotion Dynamics Conundrum in Developmental Psychopathology: Similarities, Distinctions, and Adaptiveness of Affective Variability and Socioaffective Flexibility. *Clinical Child and Family Psychology Review*, *25*(1), 44–74. https://doi.org/10.1007/s10567-022-00382-8

Mellers, B. A., McCoy, J. P., Lu, L., & Tetlock, P. E. (2024). Human and algorithmic predictions in geopolitical forecasting: Quantifying uncertainty in hard-to-quantify domains. *Perspectives on Psychological Science*, *19*(5), 711–721. https://doi.org/10.1177/17456916231185339

Mestdagh, M., Verdonck, S., Piot, M., Niemeijer, K., Kilani, G., Tuerlinckx, F., Kuppens, P., & Dejonckheere, E. (2023). m-Path: an easy-to-use and highly tailorable platform for ecological momentary assessment and intervention in behavioral research and clinical practice. *Frontiers in Digital Health*, 5. https://doi.org/10.3389/fdgth.2023.1182175

Meyvis, T., Ratner, R. K., & Levav, J. (2010). Why Don't We Learn to Accurately Forecast Feelings? How Misremembering Our Predictions Blinds Us to Past Forecasting Errors. *Journal of Experimental Psychology: General*, *139*(4), 579–589.

Moskowitz, D. S., & Young, S. N. (2006). Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. *Journal of psychiatry & neuroscience : JPN*, *31*(1), 13–20.

Ng, V. K. Y., & Cribbie, R. A. (2017). Using the Gamma Generalized Linear Model for Modeling Continuous, Skewed and Heteroscedastic Outcomes in Psychology. *Current Psychology : A Journal for Diverse Perspectives on Diverse Psychological Issues*, *36*(2), 225–235. https://doi.org/10.1007/s12144-015-9404-0

Ojo, A., Ibeh, S. C., & Kieghe, D. (2019). How Nigeria's 2015 presidential election outcome was forecasted with geodemographics and public sentiment analytics. *African Geographical Review*, *38*(4), 343–360. https://doi.org/10.1080/19376812.2018.1447976

Peters, S. A., Laham, S. M., Pachter, N., & Winship, I. M. (2014). The future in clinical genetics: affective forecasting biases in patient and clinician decision making. *Clinical Genetics*, *85*(4), 312–317. https://doi.org/10.1111/cge.12255

Pooseh, S., Kalisch, R., Köber, G., Binder, H., & Timmer, J. (2024). Intraindividual time-varying dynamic network of affects: linear autoregressive mixed-effects models for ecological momentary assessment. *Frontiers in Psychiatry*, *15*. https://doi.org/10.3389/fpsyt.2024.1213863

R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Rizeq, J. (2024). Affective forecasting and psychopathology: A scoping review. *Clinical Psychology Review*, *108*. https://doi.org/10.1016/j.cpr.2024.102392

Schuurman, N. K. (2016). Multilevel Autoregressive Modeling in Psychology: Snags and Solutions (D. Borsboom, E. Ceulemans, P. G. M. Van Der Heijden, P. De Jonge, & D. T. D. De Ridder, Interviewers). In Proefschrift Universiteit Utrecht, *Proefschrift Universiteit Utrecht*. GVO drukkers & vormgeving. https://www.nkschuurman.com/NKSchuurman_dissertation.pdf

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, *4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Simon, G. J., & Aliferis, C. F. (2024). *Artificial intelligence and machine learning in health care and medical sciences : best practices and pitfalls*. Springer. https://doi.org/10.1007/978-3-031-39355-6

Spindler, G., Stopsack, M., Aldinger, M., Grabe, H. J., & Barnow, S. (2016). What about the "ups and downs" in our daily life? The influence of affective instability on mental health. *Motivation and Emotion*, *40*(1), 148–161. https://doi.org/10.1007/s11031-015-9509-7

Takano, K., & Ehring, T. (2024). Affective forecasting as an adaptive learning process. *Emotion (Washington, D.C.)*, *24*(3), 795–807. https://doi.org/10.1037/emo0001303

Thompson, R. J., Spectre, A., S. Insel, P., Mennin, D., Gotlib, I. H., & Gruber, J. (2017). Positive and Negative Affective Forecasting in Remitted Individuals with Bipolar I Disorder, and Major Depressive Disorder, and Healthy Controls. *Cognitive Therapy and Research*, *41*(5), 673–685. https://doi.org/10.1007/s10608-017-9840-2

Weigl, K., & Forstner, T. (2021). Design of Paper-Based Visual Analogue Scale Items. *Educational and Psychological Measurement*, *81*(3), 595–611. https://doi.org/10.1177/0013164420952118

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). _dplyr: A Grammar of Data Manipulation_. R package version 1.1.4, <https://CRAN.R-project.org/package=dplyr>

Wickham H, Hester J, Bryan J (2024). _readr: Read Rectangular Text Data_. R package version 2.1.5, <https://CRAN.R-project.org/package=readr>

Wickham H, Vaughan D, Girlich M (2024). _tidyr: Tidy Messy Data_. R package version 1.3.1, <https://CRAN.R-project.org/package=tidyr>

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. *Advances in Experimental Social Psychology*, *35*, 345–411. https://doi.org/10.1016/S0065-2601(03)01006-2

Appendix A

R Packages

The following packages were used in R: dplyr (Wickham, François, et al., 2023), tidyr (Wickham, Vaughan, et al., 2024), readr (Wickham, Hester ,et al., 2024), lmerTest (Kuznetsova et al., 2017), effects (Fox & Weisberg, 2019), sjPlot (Lüdecke, 2024). The references for each of the packages can be found in the reference list. The packages were used in Version 4.4.2.

Appendix B

I acknowledge the use of ChatGPT (https://chatgpt.com/) to generate materials for background research and self-study in the drafting of this assessment. I used it mainly for computing R code for the analysis and fixing errors that occurred. Moreover, clarification of passages and improvement of readability was asked. No content generated by AI technologies has been presented as my own work.

Example prompts that were input into ChatGPT:

(1) Can I argue this: the kalman filter, a statistical model explaining affective forecasts by the fact that people update their forecasts by misremembering, whereas the multilevel AR(1) model might better explain the data of people's forecasts as this emphasizes emotional inertia (as the autocorrelation), so my argument is then, we don't need to focus on updating based on misremembering and prediction error, but we need to place more focus on emotional inertia

This prompt was written for the drafting of a research question. I wanted to clarify, with the limited existing knowledge I had at the beginning of the semester, that this was an angle I could take for the research question. The response confirmed what I already thought, that this perspective could be supported by doing extensive research on this line of reasoning.

(2) Error in group_by(): ! Must group by variables found in .data. Column Model is not found. Column Target is not found. Run rlang::last_trace() to see where the error occurred, how do I fix this error?

I used ChatGPT to write code for the analysis of the data. Where errors occurred, I furthermore used ChatGPT to resolve these.

(3) I don't really understand this line of reasoning, could you clarify what this person meant here?

I imported the original article and provided the necessary context to ask what the original author meant. Of course, extreme caution should be taken when asking ChatGPT to interpret things. Therefore, I critically read the answer to the prompt and made sure that ChatGPT did not provide information that was not supported, by thoroughly reading the original passage to make sense of the argumentation. I used the answer to the prompt to better understand the article, however, I did not (!) use the interpretation provided by ChatGPT in my own article.

(4) How can I improve this sentence in terms of readability?

The original passage was a bit weirdly phrased with awkward transitions. I used ChatGPT to draft some transition words for better flow of the passage.