Vector Autoregression (VAR) of Longitudinal Sleep and Self-report Mood Data

Jeff Brozena brozena@psu.edu

Editor: Could use one

Abstract

Self-tracking is one of many behaviors involved in the long-term self-management of chronic illnesses. As consumer-grade wearable sensors have made the collection of health-related behaviors commonplace, the quality, volume, and availability of such data has dramatically improved. This exploratory longitudinal N-of-1 study quantitatively assesses four years of sleep data captured via the Oura Ring, a consumer-grade sleep tracking device, along with self-reported mood data logged using eMood Tracker for iOS. After assessing the data for stationarity and computing the appropriate lag-length selection, a vector autoregressive (VAR) model was fit along with Granger causality tests to assess causal mechanisms within this multivariate time series. Oura's nightly sleep quality score shown to Granger-cause self-reported presence of depressed and anxious moods using a VAR(2) model.

1 Introduction

Long-term self-management of chronic illnesses such as bipolar disorder require persistent awareness of illness state over long periods of time and at varying time scales (Murnane et al., 2016; Morton et al., 2018; Majid et al., 2022). Remaining consistently aware of key indicators signalling the onset of a chronic condition allow individuals a chance at early intervention to reduce the severity of a givene episode. For example, an individual may modify behavior, engage their health practitioners, or adjust medication dosage. However nuanced, bipolar disorder is an illness that often degrades an individual's self-awareness and capacity for self-monitoring during symptomatic periods.

In the context of this specific illness, a volume of prior work has demonstrated the vital role of sleep in order to promote mood stability and prevent symptomatic episodes (Harvey et al., 2009; Murray and Harvey, 2010; Gruber et al., 2011). Although the particulars of this topic fall beyond the scope of this paper, these nuanced relationships may in fact be self-reinforcing and bidirectional — poor sleep may lead to episodic onset, which may also lead to worsening (or shortening) sleep bouts.

Given the importance of sleep in the ongoing management of this illness, accurate consumer-grade alternatives to polysomnography (considered the gold standard of sleep tracking) have emerged over the last few years. Indeed, comparatively inexpensive sleep tracking technologies like the Oura Ring have dramatically improved the quality of information that can be used to augment and inform these self-monitoring activities. Objective sensor-based tracking technology can be complemented with subjective self-report measures in order to form a more complete picture of physical and mental health across time. Given the aforementioned interplay of sleep and mood, this combination of subjective and objective tracking creates the possibility of longitudinal analysis — and potentially deepens one's capacity for self-awareness.

Following four years of consistent sleep and mood tracking, I sought to more formally interpret the data I had collected to quantify what I had previously intuited: that certain mood states could be understood (and potentially even predicted) by recent sleep trends. Indeed, this intuition has been demonstrated quantitatively in existing literature (Bose et al., 2017; Moshe et al., 2021; Jafarlou et al., 2023). As this work also demonstrates, combining data from consumer wearable technology and subjective self-report logs allows for a more comprehensive picture of health.

I will first describe the vector autoregression (VAR) method and subsequent tests, namely the Granger causality test and an impulse response analysis, that were performed to achieve these goals.

First, I will describe the methods used to achieve these goals, providing an overview of vector autoregression, Granger causality, and impulse response functions. Next, I will detail the findings of these methods on the dataset. This work concludes with a discussion of the methods and their potential applications in future work.

2 Problem setup

A multivariate time series analysis was performed using a vector autoregressive (VAR) model fit using ordinary least squares. An optimal lag order was first obtained using a combination of Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Hannan-Quinn Information Criterion (HQIC), and final prediction error (FPE). After fitting a VAR(2) model on the multiple time series data (outlined below), a Granger causality test was performed in order to assess the predictive relationships between variables. Finally, an impulse response analysis was plotted to further explore the temporal relationships between variables, specifically between sleep, heartrate variability, and self-reported mood. I will outline these analysis steps in greater detail in the sections that follow.

2.1 Vector Autoregression

A VAR(p) model for a multivariate time series is a regression model for outcomes at time t and time lagged predictors, with p indicating the lag. Given $p = 1$, the model would be concerned with one observation prior to t. As noted by Lütkepohl (2005) (as cited in Seabold and Perktold 2010), a $T \times K$ multivariate time series (where T is the number of observations and K is the number of variables) can be modeled using a p-lag VAR model, notated as

$$\begin{aligned} Y_t = \nu + A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + u_t \\ u_t \sim \mathsf{Normal}(0, \Sigma_u) \end{aligned} \tag{1}$$

where A_i is a $K \times K$ coefficient matrix.

Intercept terms are included in ν and regression coefficients are included as the subscripted A values. This equation is solved using ordinary least squares (OLS) estimation. The vector autoregressive (VAR) model is a flexible method for the analysis of causality in this setting.

2.2 Granger Causality Testing

In order to better assess the predictive capacity of the Oura sleep score on self-reported mood states, I incorporated Granger causality tests. Granger causality defines one type of relationship between time series (Granger, 1969) and states that a variable *Granger causes* another variable if "the prediction of one time series is improved by incorporating the knowledge of a second time series" (Bose et al., 2017).

Here, two autoregressive models are fit to the first time series, once with and once without the inclusion of the second time series. The improvement of the prediction is measured as the ratio of variance of the error terms. The null hypothesis states that the first variable *does not* Granger cause the second variable and is rejected if the coefficients for the lagged values of the first variable are significant.

For the purposes of this study, Granger causation tests were applied using sleep scores as a single predictor and each mood state as outcome variables.

2.3 Impulse Response Function Visualization

An impulse response function (IRF) is the "reaction of a dynamic system in response to an external change" (de Vries et al., 2023). Plotting an IRF allows for the interpretation of the impulse of a predictor on other variables on subsequent days. Given sleep scoring as a predictor, an IRF visualization was created to better understand its impact on mood state. Figure 3 displays the results of this analysis over a 10-day period.

3 Experimental Results

3.1 Dataset Description

The sleep score dataset was created using the second- and third-generation Oura Ring. The proprietary Oura sleep score is on a scale of 1 to 100 and incorporates a variety of sensor-based measures (i.e., heartrate variability, resting heartrate, body temperature) across time. Although the specifics of this algorithm are not public, the Oura Ring has been found to produce accurate measures of sleep timing and heartrate variability when compared against polysomnography (de Zambotti et al., 2019). As detailed in Table 1, my use of the Oura Ring was consistent across time. The dataset contains 1,455 nights of sleep bout data occuring between February, 2019 and March, 2023.

	Value
Total nights	1455
Missing nights	1
Mean	73.82
SD	12.36
Max	97.00
Min	30.00

Table 1: Descriptive statistics of Oura Ring sleep score data

Each day at 4:30pm I received a notification prompting me to log my subjective state in eMood Tracker, a mobile application for iOS. eMood Tracker is "recommended by psychologists, therapists, and social workers" and is intended to "track symptom data relating to Bipolar I and II disorders" (eMo, 2023). The version used through this period contains preset mood categories (depressed, irritable, anxious, and elevated) and allow users to log the presence and intensity on a scale of 0 to 3, where 0 is "not present" and 3 is "severe". The resulting dataset contains the most severe mood state per day. The contents of this dataset are outlined in Table 2.

3.2 Data Analysis

All analysis were performed in Python version 3.11.0 (Pyt) using `pandas` 1.5.3 (The pandas development team, 2020) for data preprocessing and `statsmodels` 0.13.5 (Seabold and Perktold, 2010) for

EMA Categories	**Count**
irritable	100
anxious	88
depressed	103
elevated	48

Table 2: Count of days where EMA item contains a non-zero value

modeling. Dickey-Fuller tests of stationarity were performed using `statsmodels` and the `pymdarima` library (Smith and others, 2017), a clone of R's `auto.arima`.

3.3 Stationarity, Decomposition, and Autocorrelation

A stationary time series contains no periodic fluctuation ("trend"). Without stationarity, the means and correlations given by a model will not accurately describe a time series' true signal (Bose et al., 2017). If a time series is found not to be stationary, an approach known as differencing can be applied to achieve stationarity. The Dickey-Fuller test is one mechanism to determine whether a time series is stationary.

Two Dickey-Fuller tests were performed on the sleep score time series, first via `statsmodels` and then, additionally, using `pymdarima` to assess the need for differencing. The `statsmodels` approach **?**, an Augmented Dickey-Fuller test (ADF), yielded a significant p-value of .001 indicating support for the null hypothesis that the time series is not stationary. However, the `pymdarima` approach **?**, which performed ADF using an alpha value of 0.05, yielded a non-significant p-value of 0.01 indicating that no differencing was required in order to produce a stationary time series. For the purposes of this study, I followed the results of the `pymdarima` library and assumed stationarity.

An exploratory time series decomposition visualization was created to better understand the presence of trend in the sleep score dataset. Figure 1 contains these results. Additionally, a partial autocorrelation function was plotted using `statsmodels`, displayed in Figure 2. Notably, partial autocorrelation appears to drop to zero for lag values greater than 2.

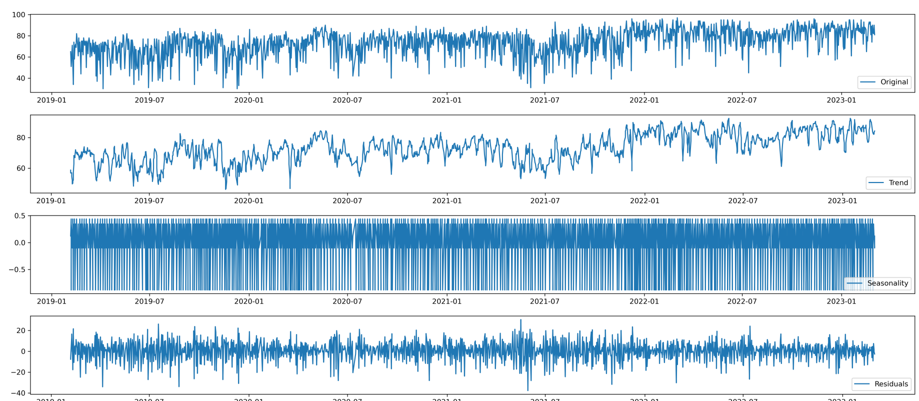


Figure 1: Decomposition of sleep time series

3.4 Lag Order Selection

In congue tristique fermentum. Morbi eleifend tortor justo, sed iaculis ante posuere vitae. Quisque sagittis ex sit amet mi sollicitudin tempor vel in ligula. Morbi porta tincidunt diam, non accumsan lorem maximus eget. Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Sed ut tristique leo, eu facilisis massa. Maecenas sed nibh hendrerit, rutrum ipsum sit amet, tempus velit. Ut venenatis dolor nisi, vel egestas est convallis ut. Nulla facilisi. Aenean vestibulum malesuada eleifend. Sed pulvinar elementum maximus. Fusce vulputate justo vel est accumsan luctus. Nulla rhoncus nisi at lacus scelerisque, et pharetra elit interdum. Donec non ante ac nulla commodo euismod a et neque.

Cras pretium imperdiet sem id euismod. In blandit pharetra urna. Integer molestie eleifend ex, nec tincidunt sapien ullamcorper sit amet. Fusce dictum purus elit, ut sagittis diam viverra nec. Vestibulum erat felis, placerat ac ligula eu, dictum lacinia lacus. Maecenas mollis augue sem,

	AIC	BIC	FPE	HQIC
0	1.688	1.708	5.408	1.695
1	-0.02482	0.09412*	0.9755	0.01979*
2	-0.03545*	0.1826	0.9652*	0.04635
3	-0.03417	0.2830	0.9664	0.08481
4	-0.02907	0.3872	0.9714	0.1271
5	-0.02537	0.4900	0.9750	0.1680
6	-0.01701	0.5975	0.9832	0.2135
7	-0.01940	0.6943	0.9809	0.2483
8	-0.01014	0.8026	0.9900	0.2948
9	-0.0008049	0.9111	0.9993	0.3413
10	0.01201	1.023	1.012	0.3913
11	0.02510	1.135	1.026	0.4415
12	0.03723	1.246	1.038	0.4909
13	0.04867	1.357	1.050	0.5395
14	0.06022	1.468	1.063	0.5882
15	0.07076	1.577	1.074	0.6359

Table 3: VAR Order Selection (* highlights the minimum)

sit amet tincidunt ex fringilla ac. Pellentesque vel metus ipsum. Sed sit amet nisi vitae tellus consectetur vulputate. Proin aliquam quam at purus tincidunt, ut pharetra ante posuere. Morbi viverra lacinia leo, ac malesuada eros aliquam vitae. Vivamus vitae nisl dictum, laoreet nisi sed, dapibus odio.

3.5 Vector Autoregression Model

Sed id feugiat tortor. Duis arcu lorem, rutrum ut nunc quis, bibendum mollis metus. In pulvinar ut mauris in malesuada. Curabitur eu massa aliquam, laoreet dui maximus, porta tellus. Aliquam erat volutpat. Proin quis lectus nisl. Mauris sagittis bibendum elit eu accumsan. Aliquam ut dui faucibus, dictum magna id, venenatis metus. Mauris ex mi, tempor vel nunc eu, placerat placerat nunc. Cras facilisis varius bibendum. Sed sollicitudin vel sem in maximus. Aenean aliquam sem ac arcu lacinia, vitae bibendum mauris malesuada. Praesent ac imperdiet erat. Maecenas tincidunt sed sem ac porta. Phasellus in justo orci.

	coefficient	std. error	t-stat	prob
L1.score	0.633262	0.027574	22.966	0.000
L1.anxious	0.153275	0.446110	0.344	0.731
L1.depressed	0.477164	0.409130	1.166	0.243
L1.irritable	-0.282988	0.412509	-0.686	0.493
L1.elevated	-0.220198	0.655784	-0.336	0.737
L2.score	-0.003080	0.027452	-0.112	0.911
L2.anxious	0.353528	0.445359	0.794	0.427
L2.depressed	1.241873	0.409667	3.031	0.002
L2.irritable	-0.080069	0.412341	-0.194	0.846
L2.elevated	-0.499540	0.657230	-0.760	0.447

Table 4: VAR results for equation score

Vestibulum sit amet faucibus purus. Proin ornare nisi et purus pellentesque mollis. Mauris aliquet metus sed sem blandit, quis accumsan justo blandit. Donec vitae lectus commodo, placerat tellus sed, interdum ex. Suspendisse potenti. Suspendisse tincidunt, enim in convallis sollicitudin, metus orci facilisis ex, non tristique nunc est ac magna. Aenean ut lorem interdum dui dictum venenatis. Nullam hendrerit placerat mauris, et tempor neque convallis vitae. Pellentesque vitae ligula id elit hendrerit dapibus. Suspendisse lobortis, dui id ultrices viverra, lorem arcu ornare dui, faucibus vehicula tortor erat sed ex. Morbi ac dolor ut nunc fringilla facilisis nec lobortis urna. Nullam ac tempus turpis.

3.6 Granger Causality

Causal Variable	Variable	Test statistic	Critical value	p-value	df
sleepscore	**depressed**	**5.384**	2.997	**0.005**	(2, 6535)
sleepscore	**anxious**	**3.294**	2.997	**0.037**	(2, 6535)
sleepscore	irritable	1.347	2.997	0.260	(2, 6535)
sleepscore	elevated	1.203	2.997	0.500	(2, 6535)

Table 5: Granger Causality Test for Sleep Score

3.7 Impulse Response Analysis

4 Discussion

Fusce consectetur accumsan tincidunt. Donec mollis odio at purus convallis tristique. Phasellus maximus quis tortor quis vehicula. Vivamus eget aliquam odio. Etiam efficitur feugiat aliquet. Cras sit amet turpis id nunc interdum placerat. Duis neque nibh, auctor nec eleifend in, pellentesque eget massa. Duis efficitur urna urna, at commodo massa efficitur sit amet. Curabitur tincidunt justo sem, sit amet hendrerit nisi dictum vitae. Ut sed sagittis dolor, vel aliquam urna. Ut placerat lorem non vehicula vehicula. Pellentesque sed justo sodales, blandit orci eget, porta sapien.

Donec pellentesque ut ipsum rhoncus venenatis. Aliquam eu nisi vel urna pharetra mollis. Nulla eget tempus odio. Sed iaculis diam sit amet accumsan suscipit. Fusce finibus arcu a purus feugiat, at vehicula sapien venenatis. Etiam fermentum lacus nisl, dictum efficitur enim fermentum id. Donec accumsan id arcu eu rhoncus. Donec sed magna posuere, efficitur est eleifend, suscipit nisl. Cras semper sollicitudin condimentum.

Nulla eleifend eros sodales elit mattis varius vel sed felis. Sed eu elit ligula. Pellentesque consectetur arcu at tortor hendrerit venenatis. Praesent luctus dolor eros, mattis pellentesque neque consectetur eu. Sed iaculis porttitor quam, varius blandit ligula. Sed justo ante, laoreet in mattis sit amet, fringilla id lectus. Integer egestas sem ut tortor gravida, a suscipit orci semper.

Nam placerat pellentesque lorem vel iaculis. Aenean congue, nibh ut mattis pharetra, eros velit mattis enim, consequat rutrum velit justo sed leo. Curabitur pretium molestie iaculis. Sed malesuada malesuada nisl nec feugiat. Quisque eu congue erat, id interdum justo. Nam fringilla condimentum tempus. Ut nec eros hendrerit, pulvinar sapien ac, molestie nunc. Maecenas mattis ultricies augue, quis lacinia massa laoreet et. Etiam ut massa vel metus mollis ultrices. Mauris bibendum neque et lectus vestibulum dapibus. Mauris quis lorem eu est tincidunt lacinia.

Integer felis massa, rhoncus quis sagittis quis, fringilla non lorem. Pellentesque feugiat eu nibh eu luctus. Duis faucibus hendrerit justo, ac condimentum turpis viverra non. Suspendisse egestas et risus sit amet luctus. In varius quam efficitur nulla efficitur sodales eget in purus. Quisque sagittis erat nec bibendum eleifend. Integer gravida nibh a commodo finibus. Nam semper sapien at mauris efficitur pretium. Phasellus quis egestas nunc.

References

The Python Language Reference. URL https://docs.python.org/3/reference/index.html.

eMoods, 2023. URL https://emoodtracker.com.

Eliezer Bose, Marilyn Hravnak, and Susan M. Sereika. Vector Autoregressive (VAR) Models and Granger Causality in Time Series Analysis in Nursing Research: Dynamic Changes Among Vital Signs Prior to Cardiorespiratory Instability Events as an Example. *Nursing research*, 66(1):12–19, 2017. ISSN 0029-6562. doi: 10.1097/NNR.0000000000000193. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5161241/.

Herman J. de Vries, Helena J. M. Pennings, Cees P. van der Schans, Robbert Sanderman, Hilbrand K. E. Oldenhuis, and Wim Kamphuis. Wearable-Measured Sleep and Resting Heart Rate Variability as an Outcome of and Predictor for Subjective Stress Measures: A Multiple N-of-1 Observational Study. *Sensors*, 23(1):332, January 2023. ISSN 1424-8220. doi: 10.3390/s23010332. URL https://www.mdpi.com/1424-8220/23/1/332. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

Massimiliano de Zambotti, Leonardo Rosas, Ian M. Colrain, and Fiona C. Baker. The Sleep of the Ring: Comparison of the ŌURA Sleep Tracker Against Polysomnography. *Behavioral Sleep Medicine*, 17(2):124–136, March 2019. ISSN 1540-2002. doi: 10.1080/15402002.2017.1300587. URL https://doi.org/10.1080/15402002.2017.1300587. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15402002.2017.1300587.

C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica : journal of the Econometric Society*, 37(3):424–438, 1969. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1912791. Publisher: [Wiley, Econometric Society].

June Gruber, David J. Miklowitz, Allison G. Harvey, Ellen Frank, David Kupfer, Michael E. Thase, Gary S. Sachs, and Terence A. Ketter. Sleep matters: Sleep functioning and course of illness in bipolar disorder. *Journal of Affective Disorders*, 134(1):416–420, November 2011. ISSN 0165-0327. doi: 10.1016/j.jad.2011.05.016. URL https://www.sciencedirect.com/science/article/pii/S016503271100262X.

Allison G. Harvey, Lisa S. Talbot, and Anda Gershon. Sleep Disturbance in Bipolar Disorder Across the Lifespan. *Clinical Psychology: Science and Practice*, 16 (2):256–277, 2009. ISSN 1468-2850. doi: 10.1111/j.1468-2850.2009.01164.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2850.2009.01164.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-2850.2009.01164.x.

Salar Jafarlou, Jocelyn Lai, Iman Azimi, Zahra Mousavi, Sina Labbaf, Ramesh C. Jain, Nikil Dutt, Jessica L. Borelli, and Amir Rahmani. Objective Prediction of Next-Day's Affect Using Multimodal Physiological and Behavioral Data: Algorithm Development and Validation Study. *JMIR Formative Research*, 7(1):e39425, March 2023. doi: 10.2196/39425. URL https://formative.jmir.org/2023/1/e39425. Company: JMIR Formative Research Distributor: JMIR Formative Research Institution: JMIR Formative Research Label: JMIR Formative Research Publisher: JMIR Publications Inc., Toronto, Canada.

Helmut Lütkepohl. *New introduction to multiple time series analysis*. New York : Springer, Berlin, 2005. ISBN 978-3-540-40172-8. OCLC: ocm61028971.

Shazmin Majid, Richard Morriss, Grazziela Figueredo, and Stuart Reeves. Exploring self-tracking practices for those with lived experience of bipolar disorder: Learning from combined principles of Patient and Public Involvement and HCI. In *Designing Interactive Systems Conference*, pages 1907–1920, Virtual Event Australia, June 2022. ACM. ISBN 978-1-4503-9358-4. doi: 10.1145/3532106.3533531. URL https://dl.acm.org/doi/10.1145/3532106.3533531.

Emma Morton, Erin E. Michalak, Rachelle Hole, Simone Buzwell, and Greg Murray. ‘Taking back the reins’ – A qualitative study of the meaning and experience of self-management in bipolar disorder. *Journal of Affective Disorders*, 228:160–165, March 2018. ISSN 0165-0327. doi: 10.1016/j.jad.2017.12.018. URL https://www.sciencedirect.com/science/article/pii/S0165032717317913.

Isaac Moshe, Yannik Terhorst, Kennedy Opoku Asare, Lasse Bosse Sander, Denzil Ferreira, Harald Baumeister, David C. Mohr, and Laura Pulkki-Råback. Predicting Symptoms of Depression and Anxiety Using Smartphone and Wearable Data. *Frontiers in Psychiatry*, 12, 2021. ISSN 1664-0640. URL https://www.frontiersin.org/articles/10.3389/fpsyt.2021.625247.

Elizabeth L Murnane, Dan Cosley, Pamara Chang, Shion Guha, Ellen Frank, Geri Gay, and Mark Matthews. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. *Journal of the American Medical Informatics Association*, 23(3):477–484, May 2016. ISSN 1067-5027. doi: 10.1093/jamia/ocv165. URL https://doi.org/10.1093/jamia/ocv165.

Greg Murray and Allison Harvey. Circadian rhythms and sleep in bipolar disorder. *Bipolar Disorders*, 12(5):459–472, 2010. ISSN 1399-5618. doi: 10.1111/j.1399-5618.2010.00843.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1399-5618.2010.00843.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1399-5618.2010.00843.x.

Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th python in science conference*, 2010.

Taylor G. Smith and others. pmdarima: ARIMA estimators for Python, 2017. URL http://www.alkaline-ml.com/pmdarima.

The pandas development team. pandas-dev/pandas: Pandas, February 2020. URL https://doi.org/10.5281/zenodo.3509134.

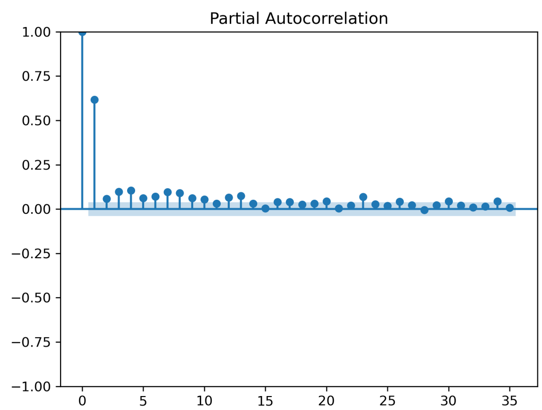
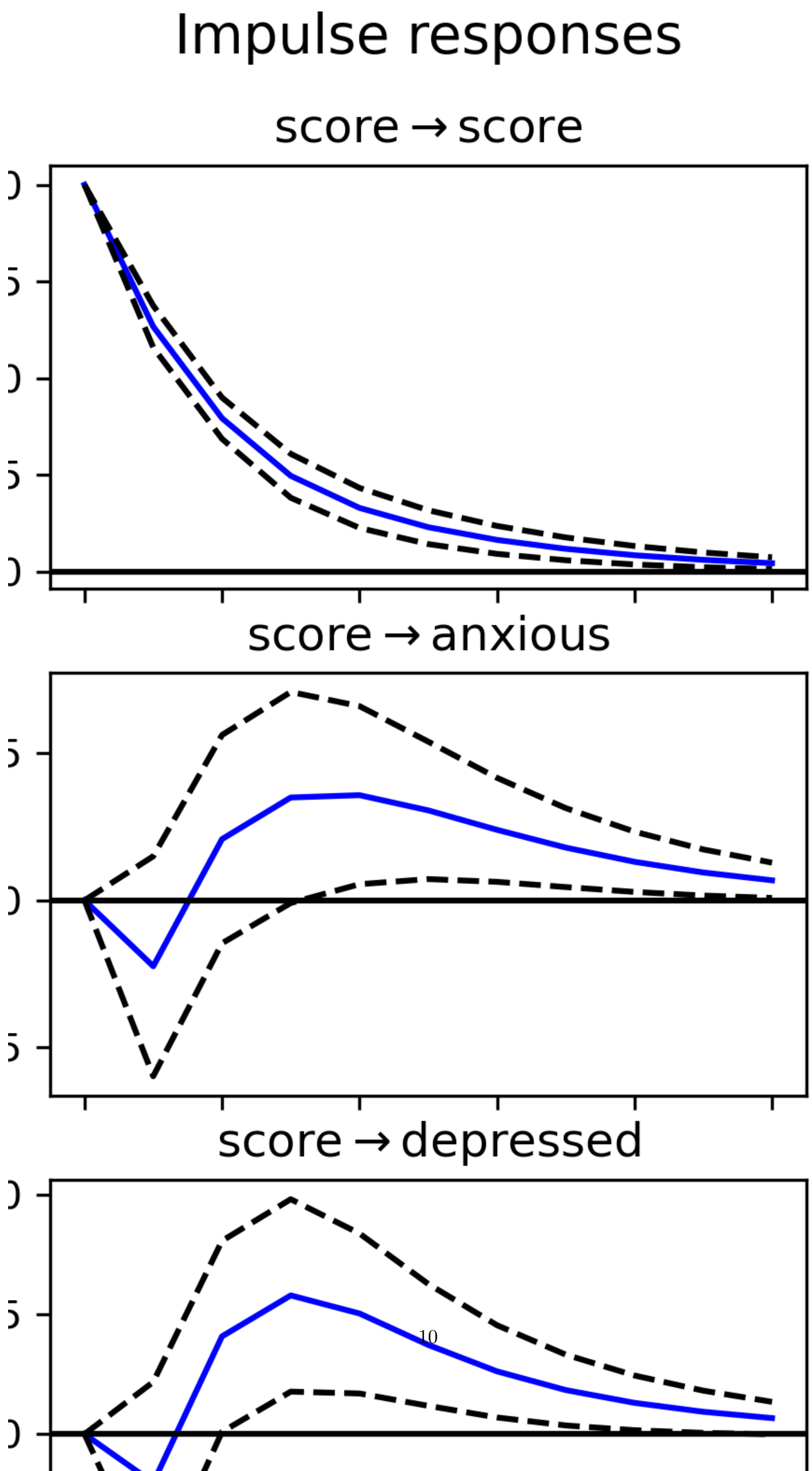


Figure 2: Partial autocorrelation of sleep time series


Impulse responses
score → score
score → anxious
score → depressed