# TALKING CHORALLY ALTERS SPEECH RHYTHM AND INDUCES FLUENCY IN PEOPLE WHO STUTTER, BUT ARE THESE THINGS CONNECTED?

Sophie Meekings[1], Lotte Eijk[1], Santosh Maruthy[2] & Sophie Scott[3]

[1]University of York, [2]All-India Institute of Speech and Hearing, [3] University College London
sophie.meekings@york.ac.uk, lotte.eijk@york.ac.uk, santoshm@aiishmysore.in, sophiescott@ucl.ac.uk

## ABSTRACT

When people who stutter speak chorally (in unison with another talker), their speech becomes more fluent. One hypothesis is that stuttering results from difficulty generating rhythmic speech gestures, and choral speech induces fluency by providing an external 'pace-setter' for speech rhythm.

In this study, 20 participants who stutter read aloud, alone and chorally with an experimenter. We measured fluency (percent syllables stuttered) and used envelope modulation spectral (EMS) analysis to derive measures of speech rhythm. Participants stuttered significantly less in the choral condition compared to solo speech, and their speech rhythm also changed: linear discriminant analysis (LDA) identified a combination of 6 EMS metrics which, together, reliably differentiated between choral and solo speech.

However, while some individual EMS metrics correlated with stuttering frequency, the specific rhythmic signature of choral speech identified by the LDA did not predict fluency, suggesting a complex relationship between speech rhythm and stuttering in this task.

**Keywords**: stuttering, choral speech, speech rhythm

## 1. INTRODUCTION

Stuttering is a neurological speech production disorder that affects an estimated 1% of adults worldwide [1]. People who stutter (PWS) experience disruptions to their speech in the form of involuntary prolongation, repetition, and blocking of speech sounds. The disorder is highly variable- people vary in how severely they are affected [2], and their disfluency can fluctuate in seemingly unpredictable ways [3]. However, research has identified a handful of situations that temporarily cause a reduction in stuttering in nearly all PWS, including speaking in time with a metronome [4], talking with pitch-shifted or delayed auditory feedback [5], and speaking in unison with another talker (choral speech) [6]. Understanding why these conditions induce fluency may provide insight into the aetiology of the disorder. Here, we focus on choral speech, which has previously been identified as more effective at reducing stuttering than other fluency-inducing manipulations [7], [8].

Since many of these manipulations affect the rate and rhythm of speech [6], it has been hypothesised that stuttering results from a deficit in the initiation and control of rhythmic speech movement [9], [10]. Thus, manipulations such as choral speech and metronome-timed speech induce fluency by providing an external signal that PWS are able to use to time their speech gestures [11].

Envelope Modulation Spectral (EMS) analysis has emerged as a measure of speech rhythm that is well suited to the analysis of disordered speech as it is not affected by typical features of atypical speech, such as prolonged pauses, and can be applied automatically with no assumptions about linguistic content [12]. In this analysis, the amplitude envelope is extracted from the full-band signal and selected frequency bands, low-pass filtered, and Fourier transformed to quantify the amplitude modulation rates that dominate in the signal and in different frequency bands. Various metrics can be extracted from the resulting power spectra that represent different characteristics of speech rhythm.

Liss et al. [12] successfully used discriminant function analysis with EMS to create rhythmic profiles that reliably distinguished different subtypes of dysarthria. These authors identified six metrics that provide information on different aspects of the speech signal, provided in Table 1 below.

| | |
|---|---|
| ***Peak frequency*** | The frequency in Hz of the peak in the spectrum with the greatest amplitude. This identifies the dominant rate of modulation in the signal. |
| ***Peak amplitude*** | The amplitude of the peak frequency (normalised by dividing by the overall amplitude of the spectrum). This indicates how much the rhythm is dominated by the peak frequency. |
| ***Energy 3-6 Hz*** | The sum of energy between 3-6 Hz, normalised. Energy in this band is correlated with intelligibility [14] and captures syllable durations [15]. |
| ***Energy 0-4 Hz*** | The sum of energy between 0-4 Hz, normalised. Energy above and below 4Hz was identified by [12] as important predictors of dysarthria type. |
| ***Energy 4-10 Hz*** | The sum of energy between 4-10 Hz, normalised. |
| ***Energy Ratio*** | Energy 0-4 Hz divided by Energy 4-10 Hz. |

**Table 1.** EMS predictors of interest identified by Liss et al., and their meaning.

Applying this approach to stuttering, Dechamma & Maruthy [13] used these six metrics to analyse speech rhythm during choral and solo speech. They found that, when PWS spoke chorally, they stuttered less and their mean peak frequency was higher than in the solo reading condition, while mean peak amplitude was lower during choral speech than solo

reading. This suggests that choral speech does change some aspects of speech rhythm, but it is not yet clear whether these changes cause a reduction in stuttering frequency, or merely co-occur with it. We aim to build on this existing work, using the metrics and discriminant analysis approach identified by Liss et al. [12] to further explore the questions raised by Dechamma & Maruthy [13] by directly testing the relationship between speech rhythm and stuttering frequency.

To examine this, we use linear discriminant analysis to identify the combination of EMS metrics that make up the 'rhythmic signature' of choral speech, and then test whether these metrics also predict stuttering frequency.

## 2. METHODS

### 2.1. Participants

21 adults who self-identified as people who stutter took part. They underwent a hearing test using an Amplivox 116 Screening Audiometer with DD45 earphones. Normal hearing was defined as a four-frequency pure tone average of less than 20dB in each ear. One participant's threshold exceeded this level and they were excluded for this reason. Thus, 20 PWS (8 female, 13 male) took part in the study. All participants were adult native British English speakers (mean age: 40 years, s.d. 12 years).

### 2.2. Task

Participants were asked to speak spontaneously for three minutes, read a passage aloud on their own, and then read a second passage in synchrony with an experimenter. They sat in a soundproofed booth facing a RODE NT1-A one-inch cardoid condenser microphone and wearing Beyerdynamic DT770 Pro circumaural headphones. The microphone was connected to a Windows computer via a Fireface UC high-speed USB audio interface. Their voices were recorded at 44100Hz with 16-bit quantisation using Adobe Audacity 3.0.

For the choral speech task, the experimenter spoke with the participant from outside the booth into an AKG 190E cardoid dynamic microphone, and heard through AKG K240 Studio on-ear headphones. The participant and experimenter were unable to see each other, ensuring that the pair could only use auditory cues to synchronise with each other.

Experiment materials were taken from Riley's Stuttering Severity Instrument IV (SSI-IV) [16]. The order of the read passages was counterbalanced across participants.

### 2.3. Measures of stuttering severity

Participants' fluency was assessed by a rater who had not been involved in the original experiment and was naive to the experimental manipulation. They performed the evaluation using the methods described in [16]. This gave four measures:

*Stuttering frequency:* The number of stuttering incidents divided by the total number of syllables uttered.

*Stuttering duration:* The average length of the three longest stuttering events timed to the nearest tenth of a second.

*Naturalness:* Perceived speech naturalness, on a scale from 1 (highly natural sounding speech) to 9 (highly unnatural sounding speech).

*Baseline severity score:* A score out of 46, calculated by taking ratings of stuttering frequency, duration, and physical tics in the spontaneous and solo speech conditions, converting to scale scores and adding them together. As it is rarely possible to produce spontaneous speech chorally, this score can only be calculated for the non-choral condition.

### 2.4. Envelope Modulation Spectrum metrics

Measures of speech rhythm in the solo and choral reading conditions were derived from the Envelope Modulation Spectrum of the whole signal using a MATLAB script following the methods described in [12], [13]

The full signal was filtered into seven octave bands centred around 125, 250, 500, 2000, 4000, and 8000 Hz. The amplitude envelope was extracted from each of the octave bands and the full-band signal, half-wave rectified and low-pass filtered at 30Hz using a fourth-order Butterworth filter, before being downsampled (to 80Hz, mean subtracted). The power spectrum of each down-sampled envelope was calculated using the Goertzel algorithm and converted to decibels for frequencies up to 10 Hz.

For each of the seven octave bands, and the full-band signal, the MATLAB script calculated six EMS metrics, as described in Table 1: peak frequency, peak amplitude, energy 3-6 Hz, energy 0-4 Hz, energy 4-10 Hz and energy ratio.

## 4. ANALYSIS

Analysis was carried out in R 4.2.2 [17] using the tidyverse [18], MASS [19], caret [20], klaR [21] and lme4 [22] packages.

### 4.1. Linear discriminant analysis

The EMS analysis produced 48 variables (six different metrics for each of seven octave bands and the full signal). To identify which, if any, of these variables best characterised the rhythmic differences between choral and solo speech, we ran stepwise linear discriminant analysis. To meet the assumption of equal variance, variables were scaled to have a mean of 0 and a standard deviation of 1.

We identified collinear variables with a pair-wise absolute correlation of 0.8 or above, and removed the variable with the largest mean absolute correlation for each pair. This reduced the predictor set to 25 variables.

The remaining variables were entered into stepwise forward variable selection. Beginning with the variable that most separated the conditions (choral

vs solo), at each step, the variable that best minimised Wilk's lambda was selected, and included in the model if the F-statistic for the change was significant ($p<0.05$). If, after adding a new predictor, any variable no longer significantly contributed to the model ($p<0.2$), it was removed.

This process identified six predictors, which were entered into a linear discriminant analysis to find a weighted linear combination of the predictor variables capable of classifying observations into either the choral or solo speech condition.

The analysis was run across the whole dataset to generate discriminant scores for data visualisation and further analysis, and with leave-one-out cross-validation to estimate the model accuracy. Results for both analyses are reported as the percentage of observations successfully classified.

### 4.2. Linear mixed-effect regression

We used model comparison with linear mixed effects regression to evaluate main effects of condition on stuttering and speech rhythm, as well as the relationship between stuttering frequency and speech rhythm.

Models were constructed for stuttering frequency, duration, naturalness, mean peak frequency and mean peak amplitude with each variable modelled as a function of condition (contrast-coded as 0.5, -0.5). Then, to explore the relationship between stuttering and speech rhythm, we modelled stuttering frequency (percentage of syllables stuttered) as a function of the linear combination of variables identified by the LDA.

This model was intended to discern whether the specific changes in speech rhythm caused by choral speech also predict fluency. To look at the relationship between speech rhythm and stuttering more generally, we modelled stuttering frequency as a function of peak frequency and peak amplitude (collapsed across all bands). All models included random intercepts for each subject.

These results were compared to the null model (with condition taken out), using a likelihood ratio test.

(1) H1: DV ~ IV + (1 | Subject)
(2) H0: DV ~ 1 + (1 | Subject)

## 5. RESULTS

### 5.1. Effect of choral speech on fluency

Assessment for stuttering severity found a wide spread of SSI-IV scores at baseline, ranging from 6 to 44 out of a possible 46 (mean 23.6, s.d. 10.9). However, when participants spoke chorally, they all converged on a relatively fluent speaking style, regardless of baseline severity.

Statistical analysis revealed a significant effect of condition on stuttering frequency ($\beta = -3.96$, SE = 1.6, $\chi^2(1) = 5.9$, $p = 0.015$) and duration ($\beta = -3.71$, SE = 1.49, $\chi^2(1) = 6.07$ , $p = 0.01$): participants stuttered significantly less, and the duration of their stutters were shorter, during choral reading compared to solo reading. However, there was no significant effect of condition on perceived naturalness ($\beta = 0.7$, SE = 0.62, $\chi^2(1) = 1.46$ , $p = 0.23$); participants' choral speech did not sound any more natural than when they spoke on their own (Figure 1).

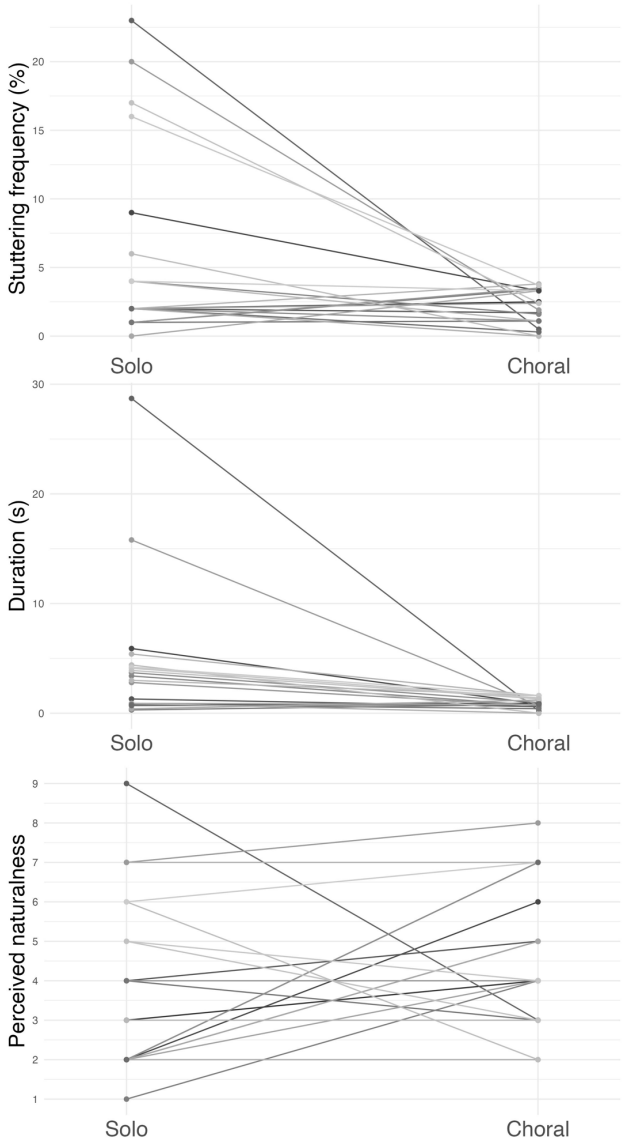


**Figure 1**: Individual scores for duration, stuttering frequency and perceived naturalness mapped between the solo and choral speaking conditions

### 5.2. Effects of choral speech on speech rhythm

There was a significant main effect of condition on peak frequency ($\beta = 0.22$, SE = 0.58, $\chi^2(1) = 13.67$ , $p = 0.0002$); peak frequency was higher in choral speech than in solo speech. However, there was no significant effect of condition on peak amplitude ($\beta = 0.39$, SE = 2.96, $\chi^2(1) = 0.0176$ , $p = 0.89$).

Stepwise variable selection identified six predictors: peak frequency in three octave bands (125, 2000 and 1000 Hz), energy between 4-10 Hz in the 4kHz octave band, full band energy between 3-6Hz and energy ratio in the 8kHz octave band, with

peak frequency in the 125Hz band as the most important predictor. Results are shown in Table 2.

| Metric | Band | Λ | F | p | F diff | p diff |
|---|---|---|---|---|---|---|
| **Peak frequency** | **125** | 0.88 | 5.38 | 0.026 | 5.38 | 0.026 |
| **Energy 4-10 Hz** | **4000** | 0.82 | 4.04 | 0.026 | 2.50 | 0.122 |
| **Peak frequency** | **2000** | 0.78 | 3.38 | 0.029 | 1.87 | 0.180 |
| **Energy 3-6 Hz** | **Full band** | 0.70 | 3.78 | 0.012 | 4.12 | 0.050 |
| **Energy ratio** | **8000** | 0.64 | 3.85 | 0.007 | 3.17 | 0.084 |
| **Peak frequency** | **1000** | 0.59 | 3.80 | 0.005 | 2.64 | 0.113 |

**Table 2:** Predictors identified by stepwise variable selection, with F-statistics and p-values for the significance of the overall model, and the difference between models when the predictor is included vs excluded.

Using these variables, the linear discriminant analysis was able to predict which condition each speech sample belonged to with 80% accuracy (leave-one-out cross-validation: 67.5%). Figure 2 shows the linear discriminant scores compared to the actual classifications. Four observations (out of 20) from each condition were mis-classified.

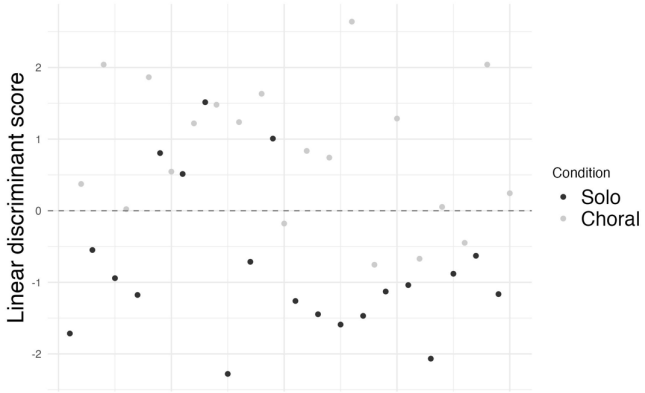


**Figure 3**: Performance of the linear discriminant. Observations above the dotted line were classified as choral speech; those below the line were classified as solo speech. The colour of the data points shows the actual class of the observation.

### 5.3. Association between fluency and rhythm metrics

There was a significant effect of peak frequency on stuttering frequency ($\beta$ = -1.25, SE = 0.35, $\chi^2(1)$ = 12.367 , $p < 0.001$): peak frequency reduced as stuttering frequency increased. However, peak amplitude did not significantly predict stuttering frequency ($\beta$ = -0.01, SE = 0.01, $\chi^2(1)$ = 3.11 , p = 0.08).

The linear combination of six variables identified by the LDA did not significantly predict stuttering frequency ($\chi^2(6)$= 7.93, p = 0.24), nor were any of the individual components significant. Estimates, standard errors, t-statistics and p-values for each of the 6 variables are given in Table 3.

| Metric | Band | β | SE | df | t | p |
|---|---|---|---|---|---|---|
| **Peak frequency** | **125** | -6.16 | 4.55 | 24.9 | -1.36 | 0.188 |
| **Energy 4-10 Hz** | **4000** | -0.01 | 0.20 | 32.8 | -0.03 | 0.978 |
| **Peak frequency** | **2000** | -1.28 | 1.81 | 31.1 | -0.71 | 0.483 |
| **Energy 3-6 Hz** | **Full band** | -23.3 | 64.9 | 28.1 | -0.36 | 0.722 |
| **Energy ratio** | **8000** | 0.01 | 0.10 | 30.8 | 0.05 | 0.958 |
| **Peak frequency** | **1000** | -1.99 | 1.36 | 32.0 | -1.47 | 0.152 |

**Table 3:** Output of linear mixed effects analysis. Degrees of freedom estimated using the Satterthwaite approximation.

## 6. DISCUSSION

In keeping with previous studies, we found that choral speech resulted in changes to speech rhythm and stuttering behaviour, compared to solo reading. Speaking chorally reduced stuttering frequency and duration in our participants, regardless of baseline severity. However, there was no effect on perceived speech naturalness, perhaps because the rhythmic changes caused by choral speech sound as unnatural as stuttering.

Our attempt to characterise the rhythmic profile of choral speech using EMS metrics met with some success. While we found no relationship between peak amplitude and condition or stuttering frequency, choral speech did result in an increase in mean peak frequency compared to solo speech, consistent with [13], and mean peak frequency was inversely correlated with stuttering frequency.

Linear discriminant analysis identified six EMS metrics that, in combination, significantly predicted which condition (solo or choral) an observation belonged to. Despite the association between overall peak frequency and stuttering frequency, this combination of metrics did not significantly predict stuttering severity compared to the null hypothesis.

Overall, it appears that choral speech and induced fluency are both related to increases in the dominant rate of modulation (i.e., peak frequency), but choral speech is further characterised by more subtle changes in speech rhythm that are not correlated with stuttering frequency. As the sample size was relatively small, these results should be interpreted with caution. A further drawback of our study is that we did not test a control group of typical speakers to contrast with PWS. Future work could test the hypothesis that PWS converge on typical speech rhythms during choral speech by comparing the ability of discriminant function analysis to separate PWS and controls when they speak alone with when they speak chorally.

## 7. REFERENCES


[1] E. Yairi and N. Ambrose, 'Epidemiology of stuttering: 21st century advances', *J. Fluen. Disord.*, vol. 38, no. 2, pp. 66–87, Jun. 2013, doi: 10.1016/j.jfludis.2012.11.002.

[2] R. I. Lanyon, 'The Measurement of Stuttering Severity', *J. Speech Hear. Res.*, vol. 10, no. 4, pp. 836–843, Dec. 1967, doi: 10.1044/jshr.1004.836.

[3] O. Bloodstein, 'A Rating Scale Study Of Conditions Under Which Stuttering Is Reduced Or Absent', *J. Speech Hear. Disord.*, vol. 15, no. 1, p. 29, Mar. 1950, doi: 10.1044/jshd.1501.29.

[4] R. Hanna and S. Morris, 'Stuttering, Speech Rate, and the Metronome Effect', *Percept. Mot. Skills*, vol. 44, no. 2, pp. 452–454, Apr. 1977, doi: 10.2466/pms.1977.44.2.452.

[5] J. Kalinowski, J. Armson, A. Stuart, and V. L. Gracco, 'Effects of Alterations in Auditory Feedback and Speech Rate on Stuttering Frequency', *Lang. Speech*, vol. 36, no. 1, pp. 1–16, 1993, doi: 10.1177/002383099303600101.

[6] G. Andrews, P. M. Howie, M. Dozsa, and B. E. Guitar, 'Stuttering: speech pattern characteristics under fluency-inducing conditions.', *J. Speech Hear. Res.*, vol. 25, no. 2, pp. 208–16, Jun. 1982.

[7] W. Johnson and L. Rosen, 'Studies in the psychology of stuttering: Effect of certain changes in speech pattern upon frequency of stuttering', *J. Speech Hear. Res.*, vol. 2, pp. 105–109, 1937.

[8] M. Kiefte and J. Armson, 'Dissecting choral speech: Properties of the accompanist critical to stuttering reduction', *J. Commun. Disord.*, vol. 41, no. 1, pp. 33–48, 2008, doi: 10.1016/j.jcomdis.2007.03.002.

[9] E. R. Brayton and E. G. Conture, 'Effects of Noise and Rhythmic Stimulation on the Speech of Stutterers', *J. Speech Lang. Hear. Res.*, vol. 21, no. 2, p. 285, Jun. 1978, doi: 10.1044/jshr.2102.285.

[10] A. C. Etchell, B. W. Johnson, and P. F. Sowman, 'Behavioral and multimodal neuroimaging evidence for a deficit in brain timing networks in stuttering: a hypothesis and theory', *Front. Hum. Neurosci.*, vol. 8, 2014, Accessed: Jan. 05, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2014.00467

[11] M. E. Wingate, 'Sound and Pattern in "Artificial" Fluency', *J. Speech Hear. Res.*, vol. 12, no. 4, pp. 677–686, Dec. 1969, doi: 10.1044/jshr.1204.677.

[12] J. M. Liss, S. LeGendre, and A. J. Lotto, 'Discriminating Dysarthria Type From Envelope Modulation Spectra', *J. Speech Lang. Hear. Res.*, vol. 53, no. 5, pp. 1246–1255, Oct. 2010, doi: 10.1044/1092-4388(2010/09-0121).

[13] D. Dechamma and S. Maruthy, 'Envelope modulation spectral (EMS) analyses of solo reading and choral reading conditions suggest changes in speech rhythm in adults who stutter', *J. Fluen. Disord.*, vol. 58, pp. 47–60, Dec. 2018, doi: 10.1016/j.jfludis.2018.09.002.

[14] T. Houtgast and H. J. M. Steeneken, 'A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria', *J. Acoust. Soc. Am.*, vol. 77, no. 3, pp. 1069–1077, Mar. 1985, doi: 10.1121/1.392224.

[15] T. Arai and S. Greenberg, 'The temporal properties of spoken Japanese are similar to those of English', in *5th European Conference on Speech Communication and Technology (Eurospeech 1997)*, Sep. 1997, pp. 1011–1014. doi: 10.21437/Eurospeech.1997-355.

[16] G. D. Riley, 'A Stuttering Severity Instrument for Children and Adults', *J. Speech Hear. Disord.*, vol. 37, no. 3, p. 314, 1972, doi: 10.1044/jshd.3703.314.

[17] R. C. Team, 'R: A language and environment for statistical computing (Version 4.0. 5)[Computer software]', *R Found. Stat. Comput.*, 2021.

[18] H. Wickham *et al.*, 'Welcome to the Tidyverse', *J. Open Source Softw.*, vol. 4, no. 43, p. 1686, Nov. 2019, doi: 10.21105/joss.01686.

[19] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S-PLUS*. Springer Science & Business Media, 2013.

[20] M. Kuhn, 'Building Predictive Models in R Using the caret Package', *J. Stat. Softw.*, vol. 28, pp. 1–26, Nov. 2008, doi: 10.18637/jss.v028.i05.

[21] C. Weihs, U. Ligges, K. Luebke, and N. Raabe, 'klaR Analyzing German Business Cycles', in *Data Analysis and Decision Support*, D. Baier, R. Decker, and L. Schmidt-Thieme, Eds. Berlin, Heidelberg: Springer, 2005, pp. 335–343. doi: 10.1007/3-540-28397-8_36.

[22] D. Bates, M. Mächler, B. M. Bolker, and S. C. Walker, 'Fitting linear mixed-effects models using lme4', *J. Stat. Softw.*, vol. 67, no. 1, Oct. 2015, doi: 10.18637/jss.v067.i01.