Prosodic Strength Intrinsic to Lexical Items: A Corpus Study of Tone Reduction in Tone4+Tone4 Words in Mandarin Chinese

Abstract

This study explored word-level prosodic strength in Mandarin Chinese reflected by tone reduction on the second syllables in Tone4+Tone4 words, by examining the slope difference between the two consecutive tones as an indicator for tonal reduction. It was found that firstly, the occurrence of tonal reduction is dependent on the internal structure of the word: words formed by apposition, (pseudo-)suffixation and replication were more vulnerable to tonal reduction than verb-object words and loanwords were. Secondly, with regards to language proficiency, higher proficiency speakers do larger amount of reduction than lower proficiency speakers do. Thirdly, the prosodic position has an asymmetrical influence on tone realization for words susceptible or unsusceptible to tone reduction: words susceptible to reduction shows heavier reduction in utterance-final positions than in non-final positions, while words not susceptible to tonal reduction do not show heavier reduction in utterance-final positions, and even show slight strengthening on the second tone in some cases. It is argued that tone reduction in T4+T4 words reflected the existence of fine-grained degrees of prosodic strength intrinsic to lexical items.

Index Terms: prosodic strength, tone reduction, Mandarin

1. Introduction

Prosodic strength at the lexical level refers to the relative prominence of minor prosodic units within a word, i.e. how syllables (or moras in some languages) are "stronger" or "weaker" in relation to each other [1]. Compared to languages with clear lexical stress such as English and Swedish, the encoding and realization of prosodic strength at the lexical level is relatively opaque for Mandarin Chinese.

It is well accepted that there is a contrast of prosodic strength between a full-tone (high, rising, dipping, falling) syllable and a neutral-tone syllable in Mandarin [2-4]. But since neutral tone is mostly associated with grammatical words such as suffixes and auxiliaries, this contrast only applies to a limited portion of words in the lexical inventory. Beyond this contrast, there are often differences of opinion about whether there are other degrees of fine-grained prosodic strength in between. Empirically, however, one commonly-shared observation is that the tone on the second syllable in a disyllabic word often undergoes reduction and behaves like a neutral tone [2, 5-7]. This reduction occurs regardless of the fact that the syllable is an independent morpheme, and is not categorized as a neutral-tone syllable either by a dictionary or by native judgment, either in a word or in isolation.

For example, in Figure 1, it is easy to observe different contrasts in lexical prosodic strength between an tonally idealized Tone4+Tone4 word in the left ("快乐") and a Tone4+Tone0 (neutral tone) word in the right ("快了"). But in fact, what happens to a considerable amount of spoken words in real speech is that their second tone can undergo substantial reduction, as shown in the middle ("快乐" with a reduced second tone). This tone is considerably reduced in pitch contour compared to its counterpart in the panel on the left, but it still retains enough of the tonal target that it differs from the complete neutralization shown in the right word.

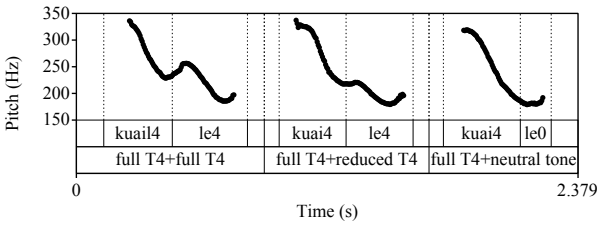

Figure 1: *Different degrees of prosodic strength revealed by a retained tone, a reduced tone and a neutral tone*

This phenomenon has been treated in different ways in previous studies. [2] proposes a distinction between "provisional weakening syllables" that undergo occasional weakening and neutral-tone syllables that undergo "stable weakening". [5] defines the provisionally reduced syllables as "unstressed syllables", and finds them different from neutral-tone syllables in the sense that their acoustic realization varies as the context changes [6]. [7] goes further to divide such provisionally reduced syllables into two categories: syllables "with substantial tone reduction" and syllables "that have lost their lexical tonal specification". However, not all studies treat this reduction as a reflection of word-level prosodic strength, given the difficulty in defining this phenomenon clearly and predicting its occurrence accurately [4].

Beside tone reduction, lexical prosodic strength in Mandarin has also been explored by many other means in previous studies, including other F_0 measurements, duration measurements [8-10], prominence perception [11-13], and prosody modeling [14]. One potential factor relevant to the encoding of lexical prosodic strength in Mandarin is the internal structure of the word. Through acoustic measurements, previous studies find that the second syllable in a disyllabic VO (verb-object) structure is stronger than its counterpart in a disyllabic non-VO structure, in the sense that they have higher F_0 and longer duration than those in non-VO structures [9, 10].

In this paper, we aim to investigate the encoding and realization of lexical prosodic strength in Mandarin by looking into tonal reduction in Tone4+Tone4 (T4T4) words in a corpus of recorded Putonghua Shuiping Ceshi ("普通话水平测试") examinations. The tonal sequence of T4T4 has been reported to be susceptible to tonal reduction in [12]. In particular, we aim to investigate how tone reduction is influenced by the internal structure of the word and how it is realized in different prosodic positions. And given the nature of the corpus we used, we can also examine how tonal reduction varies across speakers of different language proficiency.

2. Experiment

2.1. Corpus

We used a corpus of Putonghua Shuiping Ceshi (PSC) examinations from Beijing Normal University. PSC is the national standard Mandarin proficiency test in China. The test consists of four parts: The first two parts are to read 100 monosyllabic and 50 disyllabic words; the third part is to read an article of 300 characters, randomly selected from a pool of 60 articles; and the last part is to speak freely on a given topic for three minutes. The four parts are graded separately with a numeric score, and the total score (out of 100 points) is converted to a categorical proficiency level. There are six proficiency levels, which are: 一级甲等 (Class 1 Level 1), 一级乙等 (Class 1 Level 2), 二级甲等 (Class 2 Level 1), 二级乙等 (Class 2 Level 2), 三级甲等 (Class 3 Level 1) and 三级乙等 (Class 3 Level 2).

Our corpus consists of recordings of 861 college students at Beijing Normal University who took the PSC test in 2011. Among them, there is no speaker at the highest proficiency level (Class 1 Level 1), and only 24 speakers at the lowest proficiency level (Class 3 Level 2). We thus do not include data of speakers from these two proficiency levels in this study. Table 1 shows the distribution of speakers included in this study with regards to gender and proficiency.

Table 1. *Gender and Proficiency Distribution of Speakers*

Proficiency	Male	Female	Total
Class 1 Level 2	35	171	206
Class 2 Level 1	119	261	380
Class 2 Level 2	109	84	193
Class 3 Level 1	49	9	58

2.2. Forced Alignment and Annotation

For this study we used the read speech from the third part of the test. Forced alignment was conducted to determine segment boundaries with Penn Aligner [15]. All 8111 tokens of T4T4 words contained in the read articles were extracted with Perl. For each token, we annotated the position of the word in the utterance (initial, middle, final) based on punctuations in the text of the articles. We also annotated the internal structure of each word, and the primary categories of word formation in our annotation is shown in Table 2.

Table 2. *Distribution of Tokens in Word Formation*

Word Formation	Number	Examples
Apposition	2430	束缚、梦幻
Modification	1583	胜地、炽爱
Loanwords	729	社会、素质
(Pseudo-)suffixation	666	重量、力量
Verb-Complement	656	灌入、靠近
Verb-Object	584	道歉、踏步
Replication	364	默默、渐渐

Note that the (pseudo-)suffixation words contain both words formed by real suffixes (e.g., "~面") and words formed by "pseudo" suffixes that have content meaning but also are productive (e.g., "~ 量"). Pseudo-suffixation is seen as a kind of modification in some work on Chinese morphology. The verb-complement words contain both complements indicating "tendency" and complements indicating "result". These two subcategories were not separated in this study, but intuitively can make a difference in prosodic strength and should be separated in future studies. The modification words are somewhat complicated, because they are a mixture of multiple structures such as modifier-noun, adverb-verb, adjective-adjective, etc. Thus we did not include modification words this time, but only the other six categories.

2.3. F_0 Contour Modeling

The F_0 values were extracted using esps/get_f0 with a window value of 0.01 seconds [16]. For each speaker, F_0 values above 95th percentile or below 5th percentile of all F_0 values were discarded. The F_0 values were then converted to semitones according to Eq. (1),

$$F_{st} = 12 * log_2(F_0/F_{0\_}base) \quad (1)$$

where $F_{0\_}base$ is the 5th percentile of all F_0 values of a given speaker, and is used as a reference in calculating semitones.

Orthogonal polynomials were used to represent the F_0 contour as a best-fit sum of Legendre polynomials, which resulted in a model for the F_0 contour on each syllable. As has been introduced in [17, 18], the model is specified by a set of coefficients, c_i, that multiply the different Legendre polynomials before they are added together. One can derive coefficients by fitting it with the sum of Legendre polynomials using linear regression:

$$y(x;c) = \sum_x c_i * P_i(x) \quad (2)$$

where c_i are the coefficients that multiply each Legendre polynomial, and $y(x;c)$ is a model for the data. The model (y) is x- (e.g. time) dependent, and also depends on the coefficients, c.

The first few coefficients have straightforward physical interpretations. We used the second coefficient, c_1, as an indicator for the slope of the tone, and calculated the slope difference $\Delta slope$ between the two consecutive tones in a T4T4 word (i.e., the slope of the first tone minus the slope of the second one). Since T4 has a negative slope, a smaller c_1 indicates a sharper slope, and a smaller $\Delta slope$ indicates heavier tonal reduction on the second syllable.

Because the distribution of $\Delta slope$ is unknown, we used a bootstrap method [19] to calculate 95% confidence intervals from 5000 pseudo-samples.

3. Results

Figure 2 demonstrates the mean values and 95% confidence intervals of $\Delta slope$ for words formed by different means.

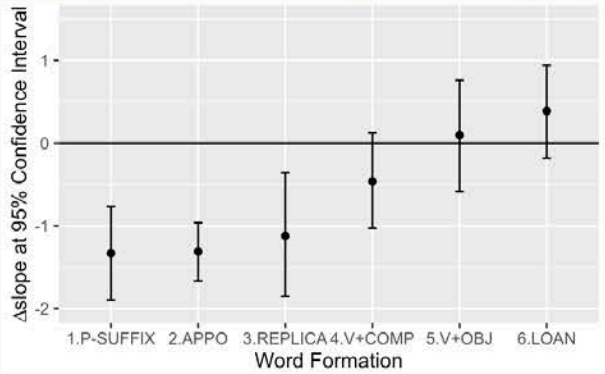

Figure 2: *Means and 95% Confidence Intervals of $\Delta slope$ for words of different types of word formation*

In Fig. 2, we can see that words formed by replication, (pseudo-) suffixation, apposition and verb-complement all have a negative mean value of *Δslope*, and their *Δslope* values at 95% confidence interval are mostly below zero. This suggests that they all undergo some amount of tonal reduction on the second syllable. In contrast, loanwords and verb-object words both show a mean *Δslope* slightly above zero, which suggests that they do not undergo tonal reduction on the second syllable as commonly as the other four types do.

Fig. 3 shows mean values and 95% confidence intervals of *Δslope* for speakers at different levels of language proficiency.

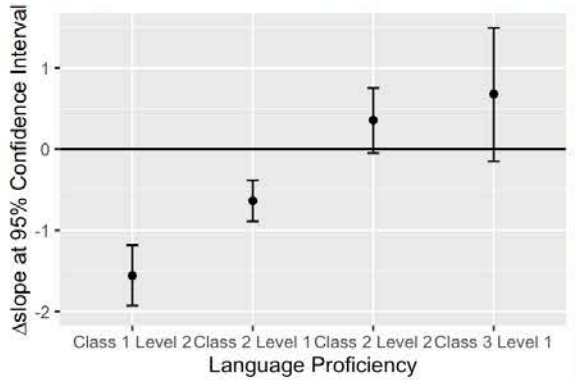

Figure 3: *Δslope at 95% Confidence Interval for Speakers of Different Levels of Language Proficiency*

According to Fig. 3, in general, as speakers' language proficiency becomes lower, the mean value of *Δslope* becomes greater, i.e., speakers of lower language proficiency produce less reduced tone on the second syllable than speakers of higher language proficiency do. Specially, the mean *Δslope* is above zero for speakers at lower proficiency levels (Class 2 Level 2, Class 3 Level 1), while it is below zero for speakers at higher proficiency levels (Class 1 Level 2, Class 2 Level 1). Also, within the high proficiency speakers who tend to reduce the second tone in T4T4 words, language proficiency still affects the amount of tone reduction (mean *Δslope* = -1.6 for Class1 Level 2, and mean *Δslope* = -0.7 for Class 2 Level 1).

Figure 4 shows the mean values and confidence intervals of *Δslope* for words of different formation and speakers of different proficiency.

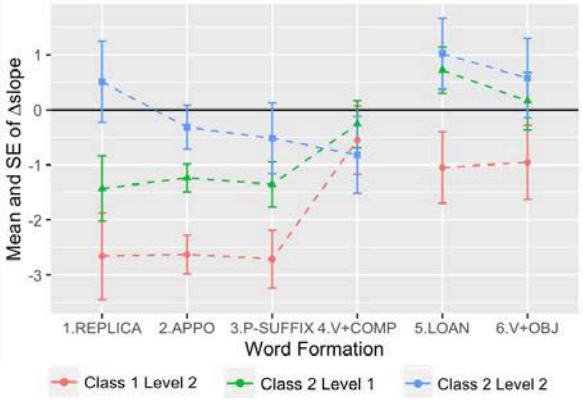

Figure 4: *Mean and Standard Error of Δslope Affected by Word Formation and Language Proficiency*

Interestingly, according to Fig. 4, we find that the above two effects (word formation, language proficiency) can affect tonal reduction in T4T4 words in parallel. On one hand, speakers of higher proficiency have smaller *Δslope* values than speakers of lower proficiency in five types of word formation out of six (except for verb-complement), which means they do heavier tonal reduction than speakers of lower proficiency in almost all types of words. On the other hand, the effect of word formation holds true for speakers of all proficiency levels: loanwords and verb-object words are always less vulnerable to tonal reduction than other types of words for speakers of all language proficiency.

We also examined how tones in T4T4 words are realized in different prosodic positions. Since replication words never occur in utterance-final positions in our corpus, they are not included in this case. Surprisingly, in Fig. 5, we see that words formed in different means can be affected by prosodic positions in asymmetric ways in terms of tonal reduction.

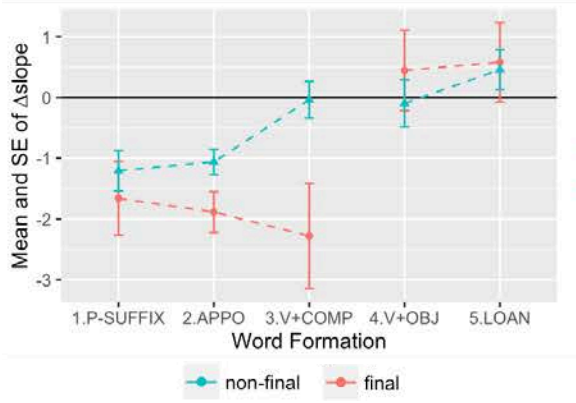

Figure 5: *Mean and 95% confidence interval of Δslope Affected by Word Formation and Prosodic Positions*

As shown in Fig. 5, for the words formed by (pseudo-) suffixation, apposition and verb-complement, whose second syllables are more vulnerable to tonal reduction according to Fig. 2, even larger amount of reduction occurs when they are located at the utterance-final position. In contrast, loanwords and verb-object words, which do not show tonal reduction in Fig. 2, do not show weakening in utterance final positions. Verb-object words in utterance-final positions even shows slight strengthening on their second tone, reflected by larger *Δslope* values in final positions than at non-final positions.

We went ahead to examine whether these two patterns of tonal realization in utterance-final positions associated with certain words, tone reduction and tone retention, are universally adopted by most speakers, or are adopted only by high-proficiency speakers, and the results are shown in Fig. 6

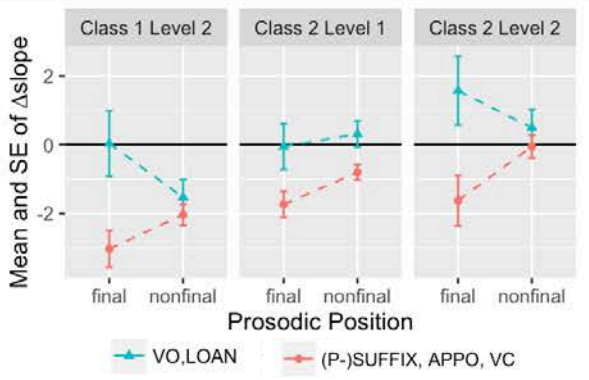

Figure 6: *Mean and 95% confidence interval of Δslope by Word Formation, Prosodic Positions and Proficiency*

Fig. 6 divides words into a reduction group (pseudo-suffixation, verb-complement, apposition) and a retention group (loanwords, verb-object) and evaluates whether the asymmetric influence from prosodic positions on the tone reduction of these two groups of words is consistent across speakers of three proficiency levels. It turns out that the three groups of speakers show fairly similar patterns for tone realization conditioned by prosodic position and word types. For words susceptible to tonal reduction, as represented by the red bars, all three groups of speakers show a heavier tonal reduction on the second syllable at the utterance-final position reflected by a smaller value of *Δslope* than at non-final positions. For words not susceptible to tonal reduction, as represented by the green bars, two groups of speakers out of three (Class 1 Level 2 and Class 2 Level 2) show slight tonal strengthening at the utterance-final position reflected by a larger value of *Δslope.* The other group of speakers (Class 2 Level 1) shows reduction in the utterance-final position for both types of words, but the amount of reduction the verb-object words and loanwords undergo in the utterance-final position is fairly slight compared to the other group. In general, these two patterns of tonal realization in utterance-final positions associated with certain words are fairly universally adopted across speakers, especially for tonal reduction in utterance-final positions for words formed by (pseudo-)suffixation, apposition and verb-complement.

4. Discussion

Compared to tone sandhi, tone reduction is more phonetically gradient and less "structure-preserving", and thus has gained less attention in previous research. However, our study shows that tonal reduction can reflect prosodic strength at the word level that is intrinsic to lexical items. For one thing, we find that tonal reduction shows clear stratification in different types of word formation. This finding is partly echoed by conclusions from previous works, e.g., verb-object words have greater prosodic strength on the second syllable than other types of words [9, 10]. For another thing, the word-intrinsic prosodic strength can yield different patterns of tonal realization in utterance-final positions: words susceptible to reduction undergo even heavier reduction while words insusceptible to reduction still retains their tonal targets in utterance-final positions. This asymmetrical phrasal behavior can be easily explained as a result of exaggerated word-level prosodic contrast, while it is difficult to attribute it to any influence from higher prosodic levels in the hierarchy.

We also find that the occurrence and amount of tonal reduction is by and large word-specific, so what can be said about the stratification is more a matter of gradient tendency than hard-and-fast rules. Much variation still exists within lexical categories, and presumably the variation can also be conditioned by other properties of the lexical items such as word frequency. Variations within word categories add to the difficulty for leaners to acquire the right pattern of tonal reduction for specific words. The likelihood of learner error increases as the language proficiency of the speaker decreases. We see that low proficiency speakers often do less tone reduction than high proficiency speakers do.

Neglecting prosodic strength at word level can cause problems for issues at higher prosodic levels, e.g., the phrasal realization of tone. [8] measured the slope difference for Tone2+Tone2 words and finds bigger slope difference at higher prosodic boundaries, which indicates a shallower tonal shape on the second tone in IP-final positions. Meanwhile, the opposite observation arises in a perception experiment, where listeners tend to perceive the second syllable in T2T2 or T4T4 words to be stronger when the word is located in utterance-final positions [12]. One potential reason for these discrepancies might be that word-level prosodic strength is not taken into consideration, so that the results can be influenced by the specific words used in the study. Teasing apart word-level prosodic strength and higher level prosodic effects will give us a better understanding of their roles and interaction.

5. Conclusion

This paper explores the tonal reduction on the second syllable of Tone4+Tone4 words as a reflection of prosodic strength at the lexical level in Mandarin. We used orthogonal polynomials to model the slope of syllables, and measured the slope difference between two consecutive Tone4 syllables (*Δslope*) as an indicator for tone reduction. Specifically, we examined how tonal reduction is stratified in words of different internal structures, realized in different prosodic positions, and varied across speakers of different language proficiency.

Our findings are as follows. Firstly, with regards to word formation, words formed by apposition, (pseudo-)suffixation and replication are susceptible to tone reduction, while verb-object words and loanwords are insusceptible to tone reduction. Secondly, with regards to language proficiency, higher proficiency speakers do larger amount of tone reduction than lower proficiency speakers do. Moreover, we find that words susceptible to reduction and words insusceptible to reduction are realized asymmetrically in utterance-final positions. Words formed by apposition, (pseudo-)suffixation and verb-complement undergo even larger amount of reduction located in utterance-final positions than they do in non-final positions. In contrast, loanwords and verb-object words do not undergo large reduction on the second tone in utterance-final positions, but instead retain their tonal targets and even show tonal strengthening occasionally. This asymmetry in the phrasal realization of tone is explained as a result of exaggerated word-level prosodic contrast.

6. Acknowledgements

This work was supported by the Beijing Normal University (11-30, 10-12-02), the Fundamental Research Funds for Central University (2010105565004GK), and the State Language Commission 12th Five-Year language research programme (YB125-41).

7. References

[1] M. Liberman and A. Prince. "On stress and linguistic rhythm". *Linguistic inquiry*. vol. 8, no. 2, pp. 249-336, 1997.

[2] Y. Chao. *A grammar of spoken Chinese*. Berkeley: University of California Press. 1968

[3] S. Duanmu. 2004. Left-headed feet and phrasal stres in Chinese. *Cahiers de linguistique-Asie orientale*, *33*(1), pp.65-103. 2004.

[4] Y. Xu. Transmitting tone and intonation simultaneously-the parallel encoding and target approximation (PENTA) model. In *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*. 2004.

[5] Y. Li, J. Tao, K. Hirose, X. Xu, and W. Lai. "Hierarchical stress modeling and generation in mandarin for expressive Text-to-Speech." *Speech Communication* 72, pp. 59-73. 2015.

[6] Y. Li, J. Tao, M. Zhang, S. Pan, and X. Xu. "Text-based unstressed syllable prediction in Mandarin." *The Proceedings of Interspeech*. 2010.

[7] S. Peng, M. Chan, C. Tseng, T. Huang, O. Lee, and Mary E. Beckman. "Towards a Pan-Mandarin system for prosodic transcription." *Prosodic typology: The phonology of intonation and phrasing*, pp. 230-270. 2005.

[8] M. Liu, S. Shi, and J. Zhang. "A preliminary study on acoustic correlates of tone2+ tone2 disyllabic word stress in Mandarin." In *INTERSPEECH*, pp. 179-183. 2014.

[9] C. Lai, Y. Sui, and J. Yuan. "A corpus study of the prosody of polysyllabic words in Mandarin Chinese." *The Proceedings of Speech Prosody*. 2010.

[10] H. Yi. "Phrasal stress in Mandarin disyllabic phrases: an investigation using focus". *The Proceedings of Speech Prosody*. 2016.

[11] Wang, Yunjia. "Tone pattern and word stress in Mandarin." In *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*. 2004.

[12] Deng, Dan, Ming Chen, and Shinan Lu. "Study on stress models of Chinese disyllable." In *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*. 2004.

[13] Y. Li, Y. Lu, X. Xu, J. Tao. "Influence of prosodic hierarchy and tone pattern on Mandarin stress perception in continuous speech". J. Tsinghua Univ. Sci. Technol. 51, pp. 1239-1242.

[14] Kochanski, Greg, Chilin Shih, and Hongyan Jing. "Quantitative measurement of prosodic strength in Mandarin." *Speech Communication* 41, no. 4, pp. 625-645. 2003

[15] J. Yuan, N. Ryant, and M. Liberman, "Automatic phonetic segmentation in Mandarin Chinese: Boundary models, glottal features and tone," Acoustics, Speech and Signal Processing (ICASSP), pp. 2539-2543, 2014.

[16] D. Talkin, D. Lin. Get_f0 online documentation. ESPS/Waves, Entropic Research Laboratory, 1996.

[17] E. Grabe, G. Kochanski, and J. Coleman. "Quantitative modelling of intonational variation." *Proc. of SASRTLM*, pp. 45-57. 2003.

[18] G. Kochanski, E. Grabe, J. Coleman, and B. Rosner. "Loudness predicts prominence: Fundamental frequency lends little." *The Journal of the Acoustical Society of America* 118, no. 2, pp. 1038-1054. 2005.

[19] Efron, Bradley, and Robert J. Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.