# An Aggregate Multitaper Test for Polynomial Frequency Modulation

Benjamin Ott
*Dept. of Mathematics and Statistics*
*Queen's University*
Kingston, Canada
16bpo@queensu.ca

Glen Takahara
*Dept. of Mathematics and Statistics*
*Queen's University*
Kingston, Canada
takahara@queensu.ca

Wesley S. Burr
*Dept. of Mathematics*
*Trent University*
Peterborough, Canada
wesleyburr@trentu.ca

***Abstract*—We propose a semiparametric multitaper test for the detection of modulated line components where the modulation is assumed to be created by a polynomial of degree $P$. This proposed test is based on an aggregation of $F$-tests, which are themselves a modification of the specific $F$-test described in [2], where the aggregation is over different multitaper orders. We derive the asymptotic distribution of the $F$-test statistic, and via simulation, we approximate the $F$-test size needed to give a pre-specified level of the aggregated test. This aggregated test is compared to the $F$-test in [2] and to our modified $F$-test via simulation in terms of the probability of detecting a frequency modulated signal.**

***Index Terms*—frequency modulation, $F$-test, aggregated test, time series, multitaper**

## I. Introduction

Detecting periodic signals in noise is an increasingly important problem in many fields. The non-parametric multitaper spectral estimator and the Harmonic $F$-test created in [5] are well known for their ability to detect periodic signals in stationary noise. However, under conditions of the carrier frequency being slowly modulated by a polynomial of degree P, the Harmonic $F$-test is often unable to detect the carrier frequency of the frequency modulated (FM) signal.

We propose a semi-parametric test to detect an FM line component modulated by a polynomial up to a given degree $P$. This test is a combination of a modification of the test in [2], with the modified being denoted by $F_4$ in the following, and an aggregation of the $F_4$ test statistics over a range of multitaper orders, which we denote by $T_a$. The aggregate test $T_a$ is shown to be more effective under broader conditions than the one constructed in [2], which is in turn based on the test given in [6]. The test statistics in [2] and those proposed in this paper are constructed within the multitaper framework. However, whereas the test statistic in [2] uses the Discrete Prolate Spheroidal Sequences (DPSS) [4] as tapers due to their optimal energy concentration properties, we adopt the Sinusoidal tapers [3] as these provide optimal bias properties as well as a (roughly four-fold) increase in computational efficiency due to their closed form. Along with the ability to downweight specific regions of the bandwidth; this is useful for providing a mechanism for controlling Type I error in our new aggregate test.

The remainder of this paper is organized as follows. In Section II, we review the test statistic from [2], which is denoted by $\tilde{F}_3$ in that work, and this. In Section III, we propose and discuss our modification of the $\tilde{F}_3$ test statistic, denoted by $F_4$, and show that the asymptotic distribution of $F_4$ is the same as that of $\tilde{F}_3$. In Section 19, we propose our aggregate test statistic, denoted by $T_a$, which aggregates the $F_4$ test statistic over multitaper orders. In Section V, we compare, via simulation, the performance of the $T_a$, $F_4$ and $\tilde{F}_3$ test statistics in terms of the probability of detecting a frequency modulated signal, at a fixed significance level of $1/N$, where $N$ is the length of the time series. In Section VI, we give concluding remarks and directions for future work.

## II. Derivation of $\tilde{F}_3$

It is necessary to speak about $\tilde{F}_3$ before proceeding to the proposed new test statistic $F_4$. Let $\mathbf{X} = \{X_t\}$, $t = 0, \ldots, N-1$ denote a time series, and assume that $\mathbf{X}$ is of the form

$$X_t = \sum_{m=0}^{M} \mu_m \cos\left(2\pi f_m t + 2\pi \int_0^t \phi_m(\tau) d\tau\right) + Z_t, \quad (1)$$

where $Z_t$ is a stationary noise process, and $\phi_m(\tau) := \sum_{p=0}^{P} a_p \tau^p$ is a polynomial of degree at most $P$ whose range in the time span of the data is assumed to lie in a given bandwidth around 0. We desire to detect the $f_m$, the carrier frequencies, of the frequency modulated signals. Assuming the $M$ carrier frequencies $f_m$ are spaced apart at least the resolution bandwidth, without loss of generality we may assume there is only a single modulated signal. We do so in this paper and adopt the notation $\mu$, $f$ and $\phi(t)$ for the amplitude, carrier frequency, and modulating function for the single modulated signal. The test statistic $\tilde{F}_3$ is created by first constructing the length $N$, complex-valued, standard inverse vector of the time series at a frequency $f$, i.e., $\mathbf{Z} := \mathbf{V}\mathbf{Y} = \mathbf{U} + i\mathbf{W}$ and its derivative $\dot{\mathbf{Z}} := \dot{\mathbf{V}}\mathbf{Y} = \dot{\mathbf{U}} + i\dot{\mathbf{W}}$, where $\mathbf{Y}$, Eqn. (14), is the vector of $K$ (complex) eigencoefficients at frequency $f$, $\mathbf{V}$ is the $N \times K$ matrix of DPSS tapers, $\dot{\mathbf{V}}$ is the time derivative of the DPSS [1], and $K$ is the number of tapers used (we have suppressed the dependence on $f$ in the preceding notation). The instantaneous frequency of the time series $\mathbf{X}$ at each time

This work was supported by National Science and Engineering Council (NSERC) Discovery Grants RGPIN-2019-06751 (Takahara) and RGPIN-2017-04741 (Burr).

$t$ in a prespecified band of width $2W$ around a frequency $f$ is given by (see [2])

$$\psi(t;f) := \frac{U(t;f)\dot{W}(t;f) - \dot{U}(t;f)W(t;f)}{2\pi\left(U^2(t;f) + W^2(t;f)\right)}, \tag{2}$$

where $U(t;f)$ is the $t$th component of $\mathbf{U}$, and similarly for $\dot{U}(t;f)$, $W(t;f)$, and $\dot{W}(t;f)$. The distribution of $\psi(t;f)$ has been shown asymptotically (as $N \to \infty$, $W \to 0$) to be

$$\psi(t;f) \sim \frac{\text{Laplace}(0,1)}{\chi_2^2} \tag{3}$$

up to a proportionality factor, when $\mathbf{X}$ is purely white noise. Additionally $\psi(t;f)$ can be scaled so it has the same distribution at each time $t$, which allows for it to be very well-behaved even though it is a non-linear function of a non-stationary process. Computing the eigencoefficients of this new instantaneous frequency series

$$\mathbf{\Psi}_k(\nu;f) := \sum_{t=0}^{N-1} v_t^{(k)} \psi(t;f) e^{-i2\pi\nu t}, \tag{4}$$

where $v_t^{(k)}$ is the $k$th data taper, we need only observe the case of $\nu = 0$ due to the concentration of the polynomial trends near 0 in the instantaneous frequency series. Thus,

$$\mathbf{\Psi}(f) = \mathbf{V}^T \psi(f), \tag{5}$$

where $\mathbf{\Psi}(f) = (\mathbf{\Psi}_0(0;f), \ldots, \mathbf{\Psi}_{K-1}(0;f))^T$ and $\psi(f) = (\psi(0;f), \ldots, \psi(N-1;f))^T$. By then solving a polynomial regression problem in the frequency domain with $\mathbf{\Psi}$ as our input data in the form

$$\mathbf{\Psi}(f) = H\mathbf{c}(f) + \mathbf{z} \tag{6}$$

where $H$ is a $K \times (P+1)$ orthonormal design matrix associated with the polynomial eigencoefficients at $\nu = 0$, using ordinary least-squares we obtain

$$\hat{\mathbf{c}}(f) = (H^T H)^{-1} H^T \mathbf{\Psi}(f) = H^T \mathbf{\Psi}(f) \tag{7}$$

$$\mathbf{r}(f) = (I_K - HH^T)\mathbf{\Psi}(f). \tag{8}$$

Notice that below we have $\hat{\mathbf{c}}_{P/0}$ which is $\hat{\mathbf{c}}$ without the first component. So, if $P = 1$, then $\hat{\mathbf{c}}_{P/0}$ would be constructed by using the submatrix $H_{P/0}$ made up of only the second column of $H$. Thus,

$$\hat{\mathbf{c}}_{P/0}(f) := (H_{P/0}^T H_{P/0})^{-1} H_{P/0}^T \mathbf{\Psi}(f) = H_{P/0}^T \mathbf{\Psi}(f) \tag{9}$$

$$\mathbf{r}_{P/0}(f) := (I_K - H_{P/0} H_{P/0}^T)\mathbf{\Psi}(\mathbf{f}) \tag{10}$$

where $\hat{\mathbf{c}}_{P/0}$ has dimension $P \times 1$ with elements $\hat{c}_{P/0}$ and $\mathbf{r}_{P/0}$ has dimension $K \times 1$ and is defined for each $p \in \{1, \cdots, P\}$. $\tilde{F}_3$ can then be defined as:

$$\tilde{F}_3(P;f,K) := \frac{\hat{c}_{P/0}^2(f)}{||\mathbf{r}_{P/0}||^2/(K-P)}. \tag{11}$$

As $N \to \infty, W \to 0$, $\tilde{F}_3$ converges in distribution to an $F$ distribution with 1 and $K - P$ degrees of freedom [2].

## III. Weighted $F_4$ Test Statistic

Our proposed statistic is constructed in two steps, the first being a modification of $\tilde{F}_3$ using the same method shown in [2], [6], [7]. To create $F_4$, consider $K$ tapers, enumerated from 0 to $K-1$. The sine tapers are given by:

$$v_t^{(k)} = \sqrt{\frac{2}{N+1}} \sin\left(\frac{\pi(k+1)(t+1)}{N+1}\right). \tag{12}$$

The matrix of these tapers is created by evaluating Eqn. (12) on a discrete mesh over $t \in \{0, \cdots, N-1\}$ and $k \in \{0, \cdots, K-1\}$ resulting in

$$\mathbf{V} := \begin{pmatrix} v_0^{(0)} & v_0^{(1)} & \cdots & v_0^{(K-1)} \\ \vdots & \ddots & & \vdots \\ v_{N-1}^{(0)} & \cdots & \cdots & v_{N-1}^{(K-1)} \end{pmatrix}$$

We can see that Eq. (12) is differentiable at all $t \in \mathbb{R}$, for any $N$ and $K$. The derivative with respect to time of the sine tapers is

$$\dot{v}^{(k)}(t) := \sqrt{\frac{2}{N+1}} \cos\left(\frac{(k+1)\pi(t+1)}{N+1}\right) \frac{(k+1)\pi}{N+1}. \tag{13}$$

Using the same discrete mesh as above, the matrix $\dot{\mathbf{V}}$ of the time derivatives of the sine tapers is obtained. By replacing the DPSS with the sinusoidal tapers we decrease the computational complexity of the test. The DPSS are the solution to an eigenvalue problem that must be computed for a given $K$ and $N$ in each $\tilde{F}_3$ test statistic rather than a direct computation of a sinusoidal function in the case of the sine tapers. The vector of complex-valued eigencoefficients at frequency $f$ is then given by

$$\mathbf{Y}_f := \mathbf{V}^T \mathbf{X}_f \tag{14}$$

where,

$$\mathbf{Y}_f := \begin{pmatrix} \tilde{y}_0(f) \\ \tilde{y}_1(f) \\ \vdots \\ \tilde{y}_{(K-1)}(f) \end{pmatrix}, \quad \mathbf{X}_f := \begin{pmatrix} X_0 e^{-i2\pi f 0} \\ X_1 e^{-i2\pi f 1} \\ \vdots \\ X_{(N-1)} e^{-i2\pi f (N-1)} \end{pmatrix}$$

Notice that this vector is identical to the one defined in Section II other than the type of tapers that are used. Now, let

$$\tilde{\mathbf{Y}}_f := \mathbf{W}\mathbf{Y}_f \tag{15}$$

be a vector of weighted eigencoefficients with non-zero weights and weight matrix

$$\mathbf{W} := \begin{pmatrix} w_0 & 0 & \cdots & 0 \\ 0 & w_1 & \cdots & 0 \\ 0 & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & w_{K-1} \end{pmatrix}$$

Following the same method described in Section II, but replacing $\mathbf{Y}_f$ with $\tilde{\mathbf{Y}}_f$, this new $\psi(t;f)$ can be shown to be still proportional asymptotically (as $N \to \infty$, $W \to 0$) to

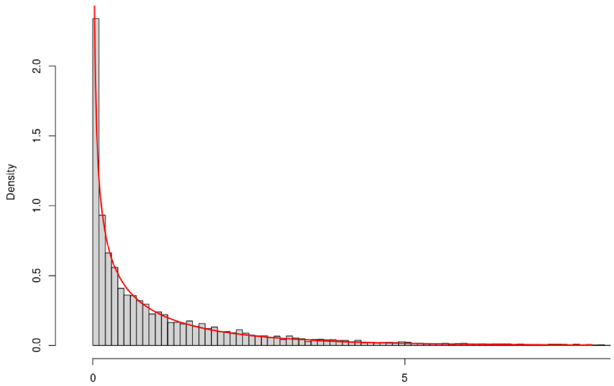


Fig. 1. Theoretical density of the $F_{1,K-P}$ distribution in red overlayed on the simulated density of $F_4$ under the null hypothesis, when the process was Gaussian white noise. The number of simulations was 10,000.

$$\psi(t; f) \sim \frac{\text{Laplace}(0, 1)}{\chi_2^2} \tag{16}$$

when $\mathbf{X}$ is purely white noise. Additionally, $\psi(t; f)$ can still be scaled so it has the same distribution at each time $t$, allowing $\psi(t; f)$ to remain well-behaved. We can then define $F_4$ as:

$$F_4(P; f, K) := \frac{\hat{c}_P^2(f)}{||\mathbf{r}_{P/0}||^2/(K-P)} \tag{17}$$

Asymptotically, $F_4$ still converges in distribution to an $F$ distribution with degrees of freedom 1 and $K-P$. This is illustrated by the simulation shown in Fig. 1, where the theoretical density of the $F_{1,K-P}$ distribution agrees well with a histogram density estimate of the distribution of $F_4(P; f, K)$, under the assumption of the time series being Gaussian white noise.

The weights used in this paper are such that the $i^{th}$ element on the diagonal of $\mathbf{W}$ is given by the $i^{th}$ element of

$$\frac{1}{\text{seq}(\text{from} = 1, \text{to} = \max(1, \text{penalty} * K), \text{length.out} = K)} \tag{18}$$

where the penalty is recommended to be 0.15 (simulation result). To illustrate the reasoning behind the addition of weights to the test statistic, consider Fig. 2. The issue arises from the need to pick $K$ (the number of tapers used in the test) before the test is conducted. We can see that under certain circumstances (see Fig. 2), if the selected $K$ is too large (e.g., $K = 80$), $\tilde{F}_3$ can occasionally miss the carrier frequency altogether. The wider window, $W$, causes the concentration of the peak to be split out into two smaller peaks found near or at $f \pm \frac{(K+1)}{2N}$. This could be impaired further if a lower confidence level was chosen, as one would falsely detect two peaks and not detect the true peak in the same test. In contrast, $F_4$ with the weighting, at $K = 80$ does detect the peak at 0.1 with reduced side spikes around $W$.

Furthermore, using a lower number of tapers, $K = 20$, as shown in Fig. 3 across all the frequencies tested, $F_4$ only detects the carrier at 0.1, whereas $\tilde{F}_3$ falsely detects three other

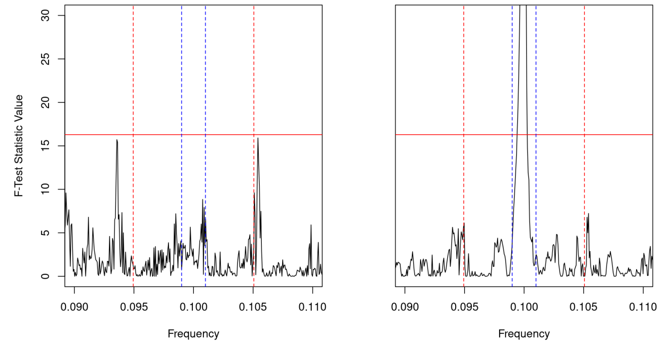


Fig. 2. (Left) $\tilde{F}_3$ applied to FM series (carrier frequency $f = 0.1$) with input parameter $K = 80$. (Right) $F_4$ applied to the same FM series with the same input parameter. Dotted blue lines represent the range in which we should find the signal in the $F$-test, dotted red lines represent $f \pm \frac{(K+1)}{2N}$, and the horizontal red line is the quantile of an $F_{1,79}$ corresponding to a significance level of $1/N$.

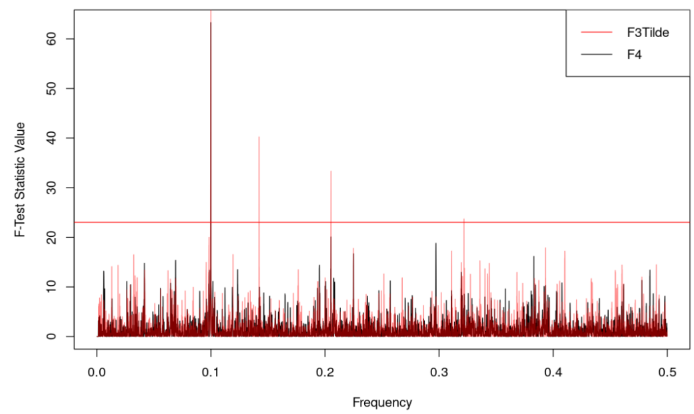


Fig. 3. $\tilde{F}_3$ in red applied to FM series under background Gaussian white noise with input parameter $K = 20$ across all frequencies. $F_4$ in black, applied to the same FM series with the same input parameter. The horizontal red line is the quantile of an $F_{1,79}$ corresponding to a significance level of $1/N$.

frequencies along with the carrier. These weights are applied to mitigate the occurrence of the side spikes around $W$ and widen the number of possible $K$'s that will be effective at detecting the true underlying modulation. Additionally, these weights cannot be effectively applied with the DPSS tapers due to the concentration of the DPSS in the spectral window. This is another reason we chose to use the sine tapers.

## IV. The Aggregate Test

The main downside of $F_4$ is that a specific $K$ still must be chosen before the test can be conducted. We define the aggregate test statistic, denoted by $T_a$, as follows. Let $\mathcal{K} := \{K_1, \cdots, K_{|\mathcal{K}|}\}$, where each $K_j \in \mathcal{K}$ is a specified number of tapers that is used in the $j^{\text{th}}$ $F_4$-test. The test statistic $T_a$ for testing for a modulated signal at frequency $f$ with modulation degree at most $P$ is given by

$$T_a := \sum_{i=1}^{|\mathcal{K}|} \mathbb{1}_{(F_4(P; f, K_i) \geq F_{1, K_i - P, \beta})} \tag{19}$$

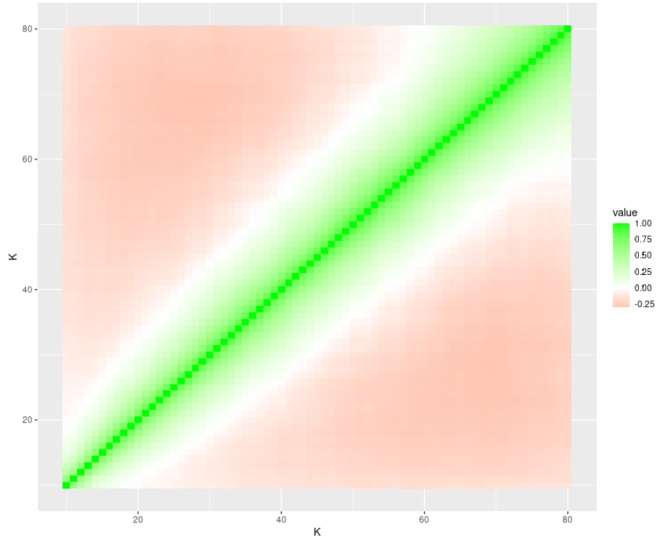


Fig. 4. Estimated correlation between 70 pairs of $(T_a)_i$ ranging from $K = 10$ to $K = 80$. 10,000 simulations were conducted with unique Gaussian white noise as the time series.

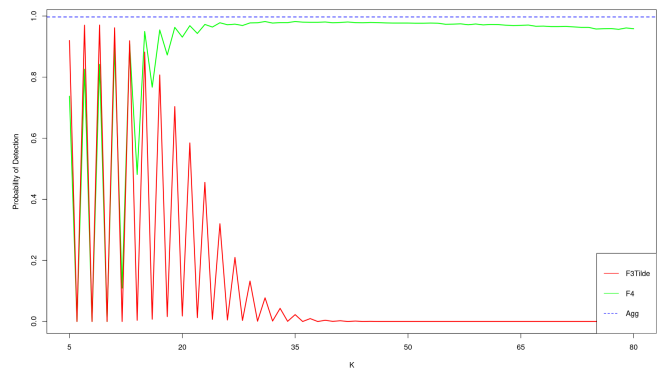


Fig. 5. Probability of detection of $\tilde{F}_3(1; 0.1, K)$ in red, $F_4(1; 0.1, K)$ in green, and the aggregate test with $\mathcal{K} = \{5, \cdots, 80\}$ and R = 5 in blue. 2,000 simulations were conducted with N = 2,000 for each simulation. The signal consisted of linear FM with a small modulation bandwidth at carrier frequency 0.1 that was embedded in Gaussian white noise. It can be given by Eqn. (1) when $\mu = 1.4$, $\phi(\tau) = \frac{0.0005}{\frac{N}{2}}\left(t - \frac{N}{2}\right)$, $f = 0.1$ and $Z_t$ is Gaussian white noise. These tests were conducted at a significance level of $\frac{1}{N} = \frac{1}{2000}$.

where $F_{1,K_i-P,\beta}$ is the $(1-\beta)$-quantile of the $F_{1,K_i-P}$ distribution. The aggregate test rejects the null hypothesis that there is no modulated signal at frequency $f$ if $T_a \geq R$, where $R$ is a specified threshold number of $F_4$ tests that must reject the null hypothesis before the aggregate test does.

If the indicators in $T_a$ were independent then the distribution of $T_a$ would Binomial($|\mathcal{K}|, \beta$). Unfortunately, this is not the case. Letting $(T_a)_i$ denote the $i$th indicator in the sum in $T_a$, each $(T_a)_i \sim$ Bernoulli($\beta$). Notice here that $\beta$ is the Type I error of each $F_4$ test in the sum. As each of the $(T_a)_i$ tests are conducted on the same starting data, only differing in the number of tapers used in the test, there is reason to suppose that the $(T_a)_i$ are correlated with each other. This is further supported by Fig. 4 in which the correlation from 10,000 simulations of 70 unique choices of $K$ was examined under Gaussian white noise. There is a positive correlation for pairs of $K$ close to each other such as $K_i = 49$ and $K_j = 51$, and a negative correlation for points further away such as $K_i = 20$ and $K_j = 80$.

With the correlation $\mathrm{Cor}\left[(T_a)_i, (T_a)_j\right] \neq 0$ for any $i, j$ pair, the distribution of $T_a$ is unknown, as it is the sum of non-independent Bernoulli($\beta$) random variables. In the next section, for a given $R$, we use simulation to determine the Type I error $\beta$ for the $F_4$ tests required to give a specified level $\alpha$ (we use $\alpha = 1/N$) in the aggregate test.

## V. Simulation and Comparison

To directly compare $\tilde{F}_3$, $F_4$, and the aggregate test; we set the Type I error of the tests all to $\frac{1}{N}$. This is trivial for the two $F$-tests, but not as straightforward for the aggregate test due to its unknown distribution. for a given $R$ we need to set $\beta$ such that the resulting $T_a$ test has a Type I error equalling $\frac{1}{N}$. To illustrate this point further, a choice of $R = 5$ or $R = 8$ will lead to choosing $\beta = 6 \times 10^{-5}$ or $\beta = 1.6 \times 10^{-4}$ respectively. Note here that these were found through simulation by choosing $R$ and $\beta$ pairs with the requirement that the resulting Type I error was $\frac{1}{N}$.

Two simulations of size 2,000 were conducted to estimate the probability of detection with $R = 5$. The first simulation was conducted under harder circumstances. We have found that a reduced bandwidth of modulation causes the $\tilde{F}_3$ to have a reduced power at higher $K$, as shown by the red curve in Fig. 5. By comparison, $F_4$ (the green curve) has the power of the test increasing, and it is more beneficial to choose a higher $K$ in practice. The horizontal dashed blue line represents the power of the aggregate test; it has a singular value as it considers all the $K \in \mathcal{K}$ where $\mathcal{K} = \{5, \cdots, 80\}$ simultaneously. However, it is not significantly different from the $F_4$ test for any $K \geq 13$. The second simulation was conducted over a wider modulation bandwidth, with results shown in Fig. 6. The range of highest power for $\tilde{F}_3$ has been shifted to higher values of $K$. Without knowing what the modulation bandwidth is before conducting the experiment, it is unlikely that a "correct" $K$ will be chosen for an arbitrary $\tilde{F}_3$ test. In comparison, $F_4$ and the aggregate test increase their power at the larger modulation bandwidth and a choice of high $K$ for $F_4$ will still yield a high power test. The advantage of the aggregate test is that at no point does $K$ need to be chosen. By choosing $\mathcal{K} = \{5, \cdots, 80\}$, we hope to create an envelope that will form an appropriately conditioned test regardless of the width of modulation contained in the time series.

## VI. Concluding Remarks and Future Work

We have introduced a new $F$-test that is more robust to the choice of $K$ and to the type of modulation contained in the series. This $F$-test is also faster to compute in practice, with a simpler form of tapers. We have presented its asymptotic null distribution and compared its performance against $\tilde{F}_3$ through simulation to show its larger power through a larger range of

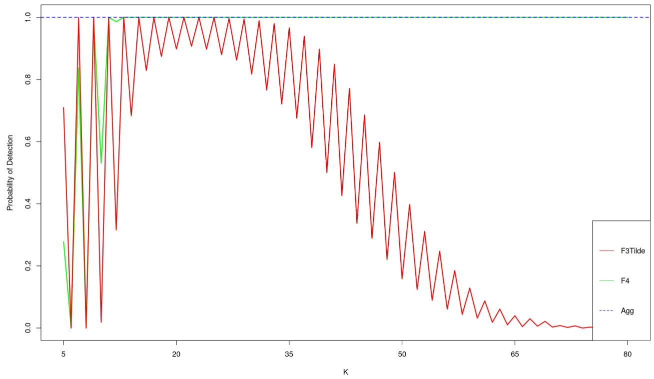


Fig. 6. Probability of detection of $\tilde{F}_3(1; 0.1, K)$ in red, $F_4(1; 0.1, K)$ in green, and the aggregate test with $\mathcal{K} = \{5, \cdots, 80\}$ and R = 5 in blue. 2,000 simulations were conducted with N = 2,000 for each simulation. The signal consisted of linear FM with a larger modulation bandwidth at carrier frequency 0.1 that was embedded in Gaussian white noise. It can be given by Eqn. (1) when, $\mu = 1.4$, $\phi(\tau) = \frac{0.001}{\frac{N}{2}}\left(t - \frac{N}{2}\right)$, $f = 0.1$ and $Z_t$ is Gaussian white noise. These tests were conducted at a significance level of $\frac{1}{N} = \frac{1}{2000}$.

$K$. We also introduced the aggregate test statistic and showed under simulation its consistency under different amounts of modulation, without the disadvantage of the choice of $K$. The issues with the unknown distribution of the aggregate test were discussed, as well as our proposed choices of $R$ and $\beta$ to make a test with Type I error $\frac{1}{N}$.

In future work, we desire to derive an approximate distribution for the sum of correlated Bernoulli random variables allowing for the aggregate test to have a defined distribution. Another interesting problem is choosing an optimal set of weights ($\mathbf{W}$), as we believe that at low $K$, the choice of weighting is causing decreased power in the $F_4$ test. Additionally, there is the interesting result of poor performance choices of $K$ even in $\tilde{F}_3$, as seen by the oscillatory behaviour in Fig. 5 and 6. We suspect this behaviour of $\tilde{F}_3$ (and $F_4$ for smaller $K$) for odd versus even $K$ is related to the modulation bandwidth and to the symmetry of the modulation about the midpoint of the time span.

## REFERENCES


[1] Kian Blanchette. Multitaper statistical tests for the detection of frequency-modulated signals. Master's thesis, Queen's University, 99 University Ave, Kingston, ON K7L 3N6, September 2020.

[2] Kian Blanchette, Wesley Burr, and Glen Takahara. An F-test for polynomial frequency modulation. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2021-June:5010–5014, 2021.

[3] Kurt S. Riedel and Alexander Sidorenko. Minimum bias multiple taper spectral estimation. *IEEE Transactions on Signal Processing*, 43(1):188–195, jan 1995.

[4] D. Slepian. Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty—V: The Discrete Case. *Bell System Technical Journal*, 57(5):1371–1430, 1978.

[5] David J. Thomson. Spectrum Estimation and Harmonic Analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982.

[6] David J Thomson. Polynomial phase demodulation in multitaper analysis; Polynomial phase demodulation in multitaper analysis. *IEEE Workshop on Statistical Signal Processing Proceedings*, 2009.

[7] D.J. Thomson. Inverse-constrained projection filters. *Proceedings SPIE*, 4478:172–183, 2001.