Model Identification via Total Frobenius Norm of Multivariate Spectra

Tucker S. McElroy
Anindya Roy [†]

[†] University of Maryland, Baltimore County

Center for Statistical Research and Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Model Identification via Total Frobenius Norm of Multivariate Spectra

Tucker S. McElroy*and Anindya Roy†

Abstract

We study the integral of the Frobenius norm as a measure of the discrepancy between two multivariate spectral densities. Such a measure can be used to fit time series models, and ensures proximity between model and process at all frequencies of the spectral density – this is more demanding than Kullback-Leibler discrepancy, which is instead related to one-step ahead forecasting performance. We develop new asymptotic results for linear and quadratic functionals of the periodogram, and make two applications of the total Frobenius norm: (i) fitting time series models, and (ii) testing whether model residuals are white noise. Model fitting results are further specialized to the case of atomic structural time series models, wherein co-integration rank testing is formally developed. Both applications are studied through simulation studies, as well as illustrations on inflation and construction data.

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

1 Introduction

In the fitting of vector time series models, it is of interest to know whether the model is a good fit to the data. This problem can be addressed through a criterion function that provides a distance measure between probability densities. In the case of univariate time series, Li (2004) gives an overview of classical diagnostic tests of goodness-of-fit (gof), while Paparoditis (2000) and Chen and Deo (2004) discuss frequency domain tests of gof. For multivariate time series, Hosking (1980), Li and McLeod (1981), and Lütkepohl (2007) discuss time-domain gof tests; there has been less literature on frequency-domain model fitting and gof testing, though see Akashi et al. (2017). Our paper proposes a total (i.e., integrated) Frobenius norm of the multivariate spectral density as a

---

*Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, tucker.s.mcelroy@census.gov

†University of Maryland, Baltimore County and U.S. Census Bureau

criterion function for time series, and develops the applications of model fitting and gof testing accordingly.

Conventionally, these applications are handled through the Kullback-Leibler (KL) criterion function (see Taniguchi and Kakizawa (2000) and McElroy and Findley (2015)), which provides a measure of discrepancy between vector processes via assessing their relative Gaussian entropy; this measure is also related to the Gaussian likelihood of a vector time series sample. The KL criterion is related to one-step ahead forecast performance of competing models (at least in the case of separable models, where the innovation covariance is separately parametrized), and therefore presents a nuanced assessment of the proximity of two processes. It is entirely possible for two competing models to forecast equally well, while having spectra that are completely distinct; see McElroy (2016) for the univariate case.

It may be desirable to have a criterion that assesses proximity in a complete way. Instead of requiring that certain functionals of competing model spectra be equal (as in the case of KL discrepancy), we may wish to require that the spectra be identically equal at all frequencies. From a testing standpoint, a non-zero discrepancy indicates some significant difference between the two models' spectra at some non-negligible set of frequencies. Failure to reject the null hypothesis of a discrepancy will indicate that all functionals of the spectra yield identical values, in particular implying that forecast performance is identical for all forecast leads. The total Frobenius norm provides such a holistic criterion, essentially capturing the notion of equivalency of models. (See Vuong (1989) and Rivers and Vuong (2002) for development of the concept of model equivalency for time series.)

To illustrate the potential utility of such a holistic discrepancy paradigm, consider the use of such a criterion to fit vector time series models. Because the model fit is required to be close at all frequencies in the data – as opposed to consideration of one-step ahead forecast performance (c.f., McElroy and Wildi (2013)) – we might expect the resulting model to be more flexible, being suited for a broader range of applications. (Although the estimates arising from a total Frobenius norm criterion – assuming a correct model specification – will be less efficient than the maximum Gaussian likelihood estimators, they need not be inferior if the process is non-Gaussian.)

In the special case of simple multivariate structural time series models, an exact solution to the minimization problem posed by the Frobenius norm criterion is available, resulting in the rapidly computable method-of-moments (MOM) estimators introduced in McElroy (2017). We show these MOM estimators are asymptotically normal with variances that can be easily calculated. We further demonstrate how rank tests for the spectral density matrix can be conducted, thereby providing an assessment of co-integration effects. Such applications, facilitated by the new asymptotic theory developed in Section 3, are an attractive facet of the total Frobenius norm framework when working with moderate dimension (say, greater than three but less than ten) time series.

As a second application, consider the problem of model gof testing. Here the relevant spectra

2

are a fitted model's spectral density and some non-parametric measure of the process' spectra, such as the periodogram or a tapered autocovariance spectral estimator (McElroy and Politis, 2014). Whereas conventional diagnostics (e.g., Ljung-Box (LB) statistics; Ljung and Box (1978)) assess the autocorrelations of the residuals, the total Frobenius norm actually encompasses these by insisting on complete agreement (up to statistical uncertainty) between sample and model autocorrelations at all lags. This is a more stringent criterion, making adequacy of model fit harder to earn – this results in adequate models having a broader range of effective applications, because gof is not restricted to performance at particular autocorrelation lags.

An essentially equivalent formulation of gof testing can be constructed by testing whether the residual spectral density corresponds to white noise, i.e., whether or not the residual spectrum is model equivalent to a white noise spectrum. See Davis and Jones (1968), Drouiche (2007), McElroy and Holan (2009b), and the overview of Kohli and Pourahmadi (2012). This formulation of the gof problem through the total Frobenius norm then yields a criterion resembling the LB statistic, where the trace of squared autocovariances are examined for their discrepancy from zero. However, the advantage of a frequency-domain formulation of the testing problem, is that the null hypothesis corresponds only to white noise, whereas in time-domain formulations (such as LB) any process having zero autocorrelations up to the maximum lag cutoff also satisfies the null (c.f., the portmanteau of Lütkepohl (2007)). Moreover, the asymptotic distribution theory of portmanteau statistics – such as Peña and Rodriguez (2002) or McElroy and Monsell (2014) – require that the process' tri-spectrum is zero; in contrast, our white noise test is valid under quite broad conditions, with a simple asymptotic variance that remarkably is independent of the tri-spectrum.

Other applications could be developed (e.g., model comparisons, discrimination analysis, etc.), but in this paper we focus upon model fitting and model diagnosis (gof). Section 2 provides the basic properties of the new criterion function. For the model diagnosis results, new asymptotic theory for quadratic functionals of the multivariate periodogram is developed and discussed in Section 3. Section 4 develops the applications to model fitting, including a new asymptotic theory for the MOM estimator of structural time series models, and a treatment of rank testing. Model diagnosis via a frequency-domain white noise test is developed in Section 5. Both these applications have simulation studies, validating the asymptotic theory in finite samples. We make two empirical applications in Section 6, applying the MOM estimators, white noise tests, and rank tests to both bivariate inflation data and four-variate housing starts data. Supplementary material includes technical proofs and additional tables pertaining to the empirical applications.

2   Framework

For a complex (possibly non-square) matrix A, the Frobenius norm is defined via

$$\|A\| = \sqrt{\mathrm{tr}(AA^*)},$$

3

where $*$ denotes conjugate transpose. We will abbreviate the trace of a matrix by $[A]$. For a stationary m-dimensional vector time series $\{x_t\}$ with autocovariance function $\Gamma(h) = \text{Cov}(x_{t+h}, x_t)$, the spectral density f is defined via

$$f(\lambda) = \sum_{h \in \mathbb{Z}} \Gamma(h) e^{-ih\lambda}.$$

This is a Hermitian matrix-valued function. Evidently, the Frobenius norm of $f(\lambda)$ depends on λ:

$$\|f(\lambda)\|^2 = \text{tr} f(\lambda)^2 = [f(\lambda)^2].$$

Taking the average over frequencies $\lambda \in [-\pi, \pi]$ yields the square of the total Frobenius spectral norm, which is a type of total variation in the process:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} [f(\lambda)^2] \, d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} \|f(\lambda)\|^2 \, d\lambda = \sum_{h \in \mathbb{Z}} [\Gamma(h)\Gamma(-h)] = \sum_{h \in \mathbb{Z}} \|\Gamma(h)\|^2,$$

which is an expression of the Plancherel identity. The middle equality follows from the identity

$$\Gamma(h) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) e^{i\lambda h} \, d\lambda.$$

This says that $\Gamma(h)$ is the hth Fourier coefficient of f. Henceforth we denote such an integral via $\langle f \rangle_h$. In terms of this notation, the total Frobenius norm is the square root of $[\langle f^2 \rangle_0]$. For short, we use the notation $|f|$, i.e.,

$$|f| := \sqrt{[\langle f^2 \rangle_0]}.$$

Given two m-variate spectra f and g, we define the Frobenius Discrepancy (FD) as the squared total Frobenius spectral norm of their difference:

$$\text{FD}(f, g) = |f - g|^2.$$

Now the Frobenius norm has the property that $\|A\| = 0$ if and only if $A_{jk} = 0$ for all entries $1 \leq j, k \leq m$. Hence $|f| = 0$ if and only if $\|f(\lambda)\|^2 = 0$ for almost every λ (with respect to Lebesgue measure), and hence if and only if $f_{jk}(\lambda) = 0$ for almost every λ and all $1 \leq j, k \leq m$. We can always alter a spectral density on sets of Lebesgue measure zero and retain the same autocovariance sequence, although in such a case we may have $f(\lambda) \neq \sum_{h \in \mathbb{Z}} \Gamma(h) e^{-ih\lambda}$ on a set of measure zero. Two spectra that are equal except on a set of frequencies of Lebesgue measure zero are said to be equal almost everywhere (a.e.). Therefore,

$$\text{FD}(f, g) = 0 \quad \text{if and only if} \quad f \overset{\text{a.e.}}{=} g$$

This property will be referred to as *complete equivalency* of f and g. When f and g pertain to two different fitted models, we say they are *model equivalent*.

4

A related construction is the spectral residual of f with respect to g, which is well-defined so long as g is invertible a.e. This spectral residual is given by $f g^{-1}$, and is denoted f/g for short. Such a quantity appears in multivariate time series analysis as the basis for fitting models via KL discrepancy, where f is the multivariate periodogram and g is the model spectral density; then

$$\mathrm{KL}(f,g) = \langle [f/g] \rangle_0 + \langle \log \det g \rangle_0,$$

which resembles $\mathrm{FD}(f,g)$ in some ways. In the context of signal extraction, the spectral residual of the signal with respect to the process yields the frequency response function of the optimal Wiener-Komogorov filter (McElroy and Trimbur, 2015). Note that such an object can be defined even in cases where the residual itself is difficult to directly compute.

The key idea in model diagnosis in time series analysis is to ensure that all pertinent information has been extracted from the data by the model – where the definition of *pertinent* is contingent on the exact application, be it one-step ahead forecasting or the detection of cyclical turning points. Formulating this paradigm in terms of entropy leads to the KL discrepancy; more generally, the model fitting project can be described as an attempt to whiten the data, i.e., determine a g such that the spectral residual of the data spectrum f (which might be the multivariate periodogram, or some other nonparametric estimate) corresponds to white noise. Mathematically, we can express such a situation via

$$f/g \equiv \langle f/g \rangle_0.$$

This is equivalent a.e. to the formulation in terms of the Frobenius norm:

$$|f/g - \langle f/g \rangle_0| = 0.$$

Although the spectral residual f/g is not a spectral density (it is not Hermitian, in general), by direct calculation

$$|f/g - \langle f/g \rangle_0|^2 = [\langle (f/g)^2 \rangle_0] - [\langle f/g \rangle_0^2] = \sum_{h \neq 0} \|\langle f/g \rangle_h\|^2.$$

Hence, this quantity equals zero if and only if all the residual autocovariances – i.e., the quantities $\langle f/g \rangle_h$ – have Frobenius norm zero, for $h \neq 0$. Of course, this resembles the LB criterion of a sum of squared residual autocorrelations. These results show that whiteness of the spectral residual (and hence, adequacy of model fit) is equivalent to $|f/g - \langle f/g \rangle_0| = 0$, and hence model diagnosis can proceed by the statistical testing of this null hypothesis.

Often a time series model is specified through a class of spectra $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$, leaving the marginal structure (or the higher-order polyspectra) unspecified; while this is sufficient to describe a Gaussian process, such an approach is frequently used to to model non-Gaussian processes as well. Denote the spectral density of the data process by $\widetilde{f}$. If $\widetilde{f} \in \mathcal{F}$, then the model is correctly specified, and there exists some true $\widetilde{\theta}$ such that $\widetilde{f} = f_{\widetilde{\theta}}$. The model $\mathcal{F}$ is fitted to the data via

some criterion function, such as KL discrepancy, and if the model is correctly specified then the minimizer is $\widetilde{\theta}$. If the model is mis-specified, then $\widetilde{f} \notin \mathcal{F}$, but we still obtain a minimizer $\widetilde{\theta}$, which is called the pseudo-true value (PTV). The PTV yields the element of $\mathcal{F}$ that is closest to $\widetilde{f}$, according to the distance metric corresponding to the fitting criterion.

From this discussion, it is apparent that $\mathrm{FD}(f_\theta, \widetilde{f})$ could be used as a model fitting criterion, with the value zero attained at the minimizer $\widetilde{\theta}$ corresponding to complete agreement, i.e., $f_{\widetilde{\theta}} \overset{\text{a.e.}}{=} \widetilde{f}$. In contrast, the KL criterion involves computing the trace variance of the spectral residual. The application to model fitting is developed in Section 4. In practice, $\widetilde{f}$ is unknown and will be replaced by some nonparametric estimate $\widehat{f}$, such as the periodogram or a tapered autocovariance spectral estimator, and we obtain empirical estimates $\widehat{\theta}$ by minimizing the corresponding criterion.

For model diagnosis, we could determine the spectral residual $\widehat{f}/f_{\widehat{\theta}}$ (assuming an invertible model) and proceed to check for whiteness. Alternatively, it may be simpler to compute the multivariate periodogram of the model residuals, and use this as a proxy for the spectral residual. (This approach has the drawback that parameter uncertainty is not accounted for, and hence there is no protection against overfitting.) This application is further developed in Section 5.

A third application is given by model comparison testing. Suppose that a second model is present, denoted by $\mathcal{G} = \{f_\xi : \xi \in \Xi\}$, and is fitted (perhaps by the same criterion) to the data process, yielding PTV $\widetilde{\xi}$. Potentially both models are mis-specified. These two models can be either nested (which means that the intersection of $\mathcal{F}$ and $\mathcal{G}$ is equal to the nested model) or non-nested (both models have some spectra not contained in the intersection). The null hypothesis is model equivalency, i.e., $\mathrm{FD}(f_{\widetilde{\theta}}, f_{\widetilde{\xi}}) = 0$, and this would be tested by fitting both models and computing $\mathrm{FD}(f_{\widehat{\theta}}, f_{\widehat{\xi}})$. We shall not formally pursue this application any further here.

3 Asymptotic Theory of Linear and Quadratic Functionals of the Periodogram

For subsequent analysis of the asymptotic properties of the total Frobenius norm, we need an understanding of the first and second moments of linear and quadratic functionals of the periodogram. The foundation for these results is found in Brillinger (2001), although several novel extensions are needed, as we discuss below. We consider a sample of size T, denoted $\{x_1, x_2, \ldots, x_T\}$, of the strictly stationary m-variate time series $\{x_t\}$. The jth component of x_t is denoted $x_{t,j}$. The vector sample mean is denoted by $\overline{x}$. Let the sample autocovariance be defined as

$$\widehat{\Gamma}(h) = T^{-1} \sum_{t=1}^{T-h} (x_{t+h} - \overline{x})(x_t - \overline{x})'$$

for $h \geq 0$, and $\widehat{\Gamma}(h) = \widehat{\Gamma}(-h)'$ for $h < 0$. We next define the Discrete Fourier Transform (DFT). For any $\lambda \in [-\pi, \pi]$, let

$$d(\lambda) = \sum_{t=1}^{T} (x_t - \overline{x}) \, e^{-i\lambda t}.$$

Note that this definition (as in the treatment of Brillinger (2001)) does not normalize by $T^{-1/2}$, which makes the asymptotic analysis easier to parse. We refer to components as $d_k(\lambda)$. The periodogram I is defined such that $I = T^{-1} d \, d^*$ (suppressing λ), and it follows that $\langle I \rangle_h = \widehat{\Gamma}(h)$.

Next, for continuous real matrix-valued functions φ_1, φ_2, φ_3 of frequency λ we define quadratic and linear functionals via

$$Q_{\varphi_1, \varphi_2}(f, g) = \langle [\varphi_1 \, f \, \varphi_2 \, g] \rangle_0 \qquad L_{\varphi_3}(f) = \langle [\varphi_3 \, f] \rangle_0.$$

Note that the trace and integration operators are interchangeable by linearity. Plug-in estimators for the quadratic and linear functionals $Q_{\varphi_1, \varphi_2}(f, f)$ and $L_{\varphi_3}(f)$ are $Q_{\varphi_1, \varphi_2}(I, I)$ and $L_{\varphi_3}(I)$; unfortunately, the quadratic estimator is biased because integration over frequencies is not equivalent to averaging over Fourier frequencies when considering nonlinear functions of the periodogram; see Chen and Deo (2000) and discussion in McElroy and Holan (2009a). As a result, it is important to construct estimators based on discretizing the functionals' integrals, restricting to the Riemann mesh of Fourier frequencies. In particular, let $\lambda_j = 2\pi(j-1)/T - \pi$, which for $j = 1, 2, \ldots, T$ is in the set $[-\pi, \pi]$, and define

$$\widehat{Q}_{\varphi_1, \varphi_2}(f, g) = T^{-1} \sum_{j=1}^{T} [\varphi_1 \, f \, \varphi_2 \, g](\lambda_j) \qquad \widehat{L}_{\varphi_3}(f) = T^{-1} \sum_{j=1}^{T} [\varphi_3 \, f](\lambda_j).$$

Then consistent estimators of $Q_{\varphi_1, \varphi_2}(f, f)$ and $L_{\varphi_3}(f)$ are given by $\widehat{Q}_{\varphi_1, \varphi_2}(I, I)$ and $\widehat{L}_{\varphi_3}(I)$. (Actually, if we integrate over all frequencies, the estimators $Q_{\varphi_1, \varphi_2}(I, I)$ and $L_{\varphi_3}(I)$ are also consistent, but the variability in the former estimator will differ from what is given in Proposition 2 below.) This section provides asymptotic theory for $\widehat{L}_{\varphi_3}(I)$ and $\widehat{Q}_{\varphi_1, \varphi_2}(I, I)$, based on assumptions involving summability conditions on the cumulants of the $\{x_t\}$ process. Supposing that all moments exist and the cumulant functions are defined via

$$\gamma_{a_1, \ldots, a_k}(t_1, \ldots, t_{k-1}) = \text{cum}\{x_{t_1, a_1}, x_{t_2, a_2}, \ldots, x_{t_{k-1}, a_{k-1}}, x_{0, a_k}\}$$

for $a_1, \ldots, a_k \in \{1, \ldots, m\}$. Clearly, $\gamma_{a_1, a_2}(t_1)$ equals the (a_1, a_2) entry of $\Gamma(t_1)$. We will assume that Assumption B1 holds.

Assumption (B1): for all $k \geq 2$ and each $j = 1, \ldots, k-1$ and any k-tuple $a_1, \ldots, a_k \in \{1, \ldots, m\}$, we have

$$\sum_{t_1, \ldots, t_{k-1} \in \mathbb{Z}} (1 + |t_j|) \, |\gamma_{a_1, \ldots, a_k}(t_1, \ldots, t_{k-1})| < \infty.$$

Under this condition the asymptotic behavior of the first moments of the linear and quadratic functionals can be established.

Proposition 1. *(Convergence of First Moments) Assume that $\{x_t\}$ is strictly stationary with spectral density $\widetilde{f}$, and satisfies condition (B1). Then as $T \to \infty$,*

$$\mathbb{E}\widehat{L}_{\varphi_3}(I) = L_{\varphi_3}(\widetilde{f}) + O(T^{-1})$$
$$\mathbb{E}\widehat{Q}_{\varphi_1,\varphi_2}(I,I) = Q_{\varphi_1,\varphi_2}(\widetilde{f},\widetilde{f}) + \langle [\varphi_1\,\widetilde{f}]\,[\varphi_2\,\widetilde{f}]\rangle_0 + O(T^{-1}).$$

Remark 1. It is already known (Lemma 3.1.1 of Taniguchi and Kakizawa (2000)) that

$$\widehat{L}_{\varphi_2}(I) \xrightarrow{P} L_{\varphi_2}(\widetilde{f}).$$

This result, as well as consistency of the quadratic functional, follows from Proposition 2 below. Hence, we introduce the notation $\overline{L}_{\varphi_2}(I)$ and $\overline{Q}_{\varphi_1,\varphi_2}(I,I)$ for the estimated functionals centered by their respective means.

In the next result we utilize the notation $[h] = [\varphi_h\,\widetilde{f}]$, $[hi] = [\varphi_h\,\widetilde{f}\,\varphi_i\,\widetilde{f}]$, $[hij] = [\varphi_h\,\widetilde{f}\,\varphi_i\,\widetilde{f}\,\varphi_j\,\widetilde{f}]$, and $[hijk] = [\varphi_h\,\widetilde{f}\,\varphi_i\,\widetilde{f}\,\varphi_j\,\widetilde{f}\,\varphi_k\,\widetilde{f}]$ for $h,i,j,k \in \{1,2,3,4\}$. It can happen that a function φ appears in such a term with its argument reflected and the matrix transposed, i.e, $\varphi(-\lambda)'$, in which case we place an underscore under the index. The tri-spectral density is denoted by $\widetilde{f}$ with four subindices and three frequency arguments, and is defined via

$$\widetilde{f}_{\ell ksr}(\lambda,\theta,\omega) = \sum_{h_1,h_2,h_3 \in \mathbb{Z}} \gamma_{\ell ksr}(h_1,h_2,h_3)\,e^{-i(\lambda h_1 + \theta h_2 + \omega h_3)}.$$

We also use the notation

$$[[A\,\widetilde{f}(\lambda,-\lambda,\omega)\,B]] = \sum_{\ell,k,r,s} A_{k\ell}\,\widetilde{f}_{\ell ksr}(\lambda,-\lambda,\omega)\,B_{rs},$$

which can be visualized as taking the trace of A times the trispectrum with respect to the first two indices, and also taking the trace of B times the trispectrum with respect to the latter two indices.

Proposition 2. *(Convergence of Second Moments) Assume that $\{x_t\}$ is strictly stationary with spectral density $\widetilde{f}$ and satisfies condition (B1). Then as $T \to \infty$,*

$$Cov\left(\sqrt{T}\,\widehat{L}_{\varphi_1}(I), \sqrt{T}\,\widehat{L}_{\varphi_2}(I)\right) \to \langle [12]\rangle_0 + \langle [1\underline{2}]\rangle_0$$
$$+ (2\pi)^{-2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} [[\varphi_1(\lambda)\,\widetilde{f}(\lambda,-\lambda,\omega)\,\varphi_2(\omega)]]\,d\lambda\,d\omega$$

and

$$Cov\left(\sqrt{T}\,\widehat{Q}_{\varphi_1,\varphi_2}(I,I), \sqrt{T}\,\widehat{L}_{\varphi_3}(I)\right) \to \langle [1]\,[23]\rangle_0 + \langle [132]\rangle_0 + \langle [123]\rangle_0 + \langle [2]\,[13]\rangle_0$$
$$+ \langle [1]\,[2\underline{3}]\rangle_0 + \langle [1\underline{3}2]\rangle_0 + \langle [12\underline{3}]\rangle_0 + \langle [2]\,[1\underline{3}]\rangle_0$$
$$+ (2\pi)^{-2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g_{123}(\lambda,\omega)\,d\lambda\,d\omega$$

8

and

$$Cov\left(\sqrt{T}\,\widehat{Q}_{\varphi_1,\varphi_2}(I,I),\sqrt{T}\,\widehat{Q}_{\varphi_3,\varphi_4}(I,I)\right) \to (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g_{1234}(\lambda,\omega)\,d\lambda\,d\omega$$

$$+ \langle[1][4][32]\rangle_0 + \langle[1][243]\rangle_0 + \langle[1][234]\rangle_0 + \langle[1][3][24]\rangle_0$$
$$+ \langle[4][132]\rangle_0 + \langle[1342]\rangle_0 + \langle[3][142]\rangle_0 + \langle[1432]\rangle_0$$
$$+ \langle[2][4][31]\rangle_0 + \langle[31][42]\rangle_0 + \langle[1423]\rangle_0 + \langle[2][143]\rangle_0$$
$$+ \langle[4][123]\rangle_0 + \langle[1243]\rangle_0 + \langle[2][134]\rangle_0 + \langle[1324]\rangle_0$$
$$+ \langle[14][23]\rangle_0 + \langle[3][2][14]\rangle_0 + \langle[1234]\rangle_0 + \langle[3][124]\rangle_0$$
$$+ \langle[1][2\underline{43}]\rangle_0 + \langle[\underline{4}][1][23]\rangle_0 + \langle[1][\underline{3}][2\underline{4}]\rangle_0 + \langle[1][2\underline{34}]\rangle_0$$
$$+ \langle[1\underline{342}]\rangle_0 + \langle[1\underline{32}][\underline{4}]\rangle_0 + \langle[1\underline{42}][\underline{3}]\rangle_0 + \langle[1\underline{432}]\rangle_0$$
$$+ \langle[\underline{4}][2][1\underline{3}]\rangle_0 + \langle[1\underline{3}][2\underline{4}]\rangle_0 + \langle[12\underline{43}]\rangle_0 + \langle[12\underline{3}][\underline{4}]\rangle_0$$
$$+ \langle[1\underline{423}]\rangle_0 + \langle[1\underline{43}][2]\rangle_0 + \langle[12\underline{4}][\underline{3}]\rangle_0 + \langle[12\underline{34}]\rangle_0$$
$$+ \langle[1\underline{34}]\rangle_0 + \langle[1\underline{324}]\rangle_0 + \langle[1\underline{4}][2][\underline{3}]\rangle_0 + \langle[1\underline{4}][2\underline{3}]\rangle_0,$$

where the functions $g_{123}(\lambda,\omega)$ *and* $g_{1234}(\lambda,\omega)$ *depend upon the tri-spectral density, and are given by (A.5) and (A.4) of the proof.*

We here introduce a notation for the limiting covariances: write $V_{\varphi_1|\varphi_2}$, $V_{\varphi_1,\varphi_2|\varphi_3}$, and $V_{\varphi_1,\varphi_2|\varphi_3,\varphi_4}$, for the cases of linear-linear, quadratic-linear, and quadratic-quadratic, respectively. These quantities depend on the matrix-valued functions φ_i, as well as the spectral density $\widetilde{f}(\lambda)$ and the tri-spectral density $\widetilde{f}(\lambda,-\lambda,\omega)$. Based on the moment convergence, we can formulate a central limit theorem for vectors of linear and quadratic functionals.

Theorem 1. *(CLT) Assume that* $\{x_t\}$ *is strictly stationary and satisfies condition (B1). Then the vector of* r *linear and* s *quadratic functionals are jointly asymptotically normal:*

$$\sqrt{T}\left[\overline{L}_{\vartheta_1}(I),\ldots,\overline{L}_{\vartheta_r}(I),\overline{Q}_{\varphi_1,\psi_1}(I),\ldots,\overline{Q}_{\varphi_s,\psi_s}(I)\right] \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0,V\right),$$

$$V = \begin{bmatrix}
V_{\vartheta_1|\vartheta_1} & \cdots & V_{\vartheta_1|\vartheta_r} & V_{\vartheta_1|\varphi_1,\psi_1} & \cdots & V_{\vartheta_1|\varphi_s,\psi_s} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
V_{\vartheta_r|\vartheta_1} & \ddots & V_{\vartheta_r|\vartheta_r} & V_{\vartheta_r|\varphi_1,\psi_1} & \ddots & V_{\vartheta_r|\varphi_s,\psi_s} \\
V_{\varphi_1,\psi_1|\vartheta_1} & \ddots & V_{\varphi_1,\psi_1|\vartheta_r} & V_{\varphi_1,\psi_1|\varphi_1,\psi_1} & \ddots & V_{\varphi_1,\psi_1|\varphi_s,\psi_s} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
V_{\varphi_s,\psi_s|\vartheta_1} & \ddots & V_{\varphi_s,\psi_s|\vartheta_r} & V_{\varphi_s,\psi_s|\varphi_1,\psi_1} & \ddots & V_{\varphi_s,\psi_s|\varphi_s,\psi_s}
\end{bmatrix}.$$

This theory is somewhat more general than needed for our particular applications below, but nonetheless provides a complete framework for understanding how second and fourth cumulant functions impact the covariances of linear and quadratic functionals of the periodogram.

4   Model Fitting

4.1   General Theory

Here we consider the fitting of model $\mathcal{F}$ via the integrated Frobenius norm, on the basis of the periodogram. The discrepancy of the model $\mathcal{F}$ from truth is given by the measure $\mathrm{FD}(f_\theta, \widetilde{f}) = |\widetilde{f} - f_\theta|^2$, which equals

$$
Q_{\mathrm{id,id}}(\widetilde{f}, \widetilde{f}) - 2\, Q_{\mathrm{id,id}}(\widetilde{f}, f_\theta) + Q_{\mathrm{id,id}}(f_\theta, f_\theta)
$$
$$
= Q_{\mathrm{id,id}}(\widetilde{f}, \widetilde{f}) - 2\, L_{f_\theta}(\widetilde{f}) + \langle [f_\theta^2] \rangle_0.
$$

Replacing the quadratic functionals by their empirical versions yields the criterion function

$$
\widehat{\mathrm{FD}}(f_\theta, I) = \widehat{Q}_{\mathrm{id,id}}(I, I) - 2\, L_{f_\theta}(I) + \langle [f_\theta^2] \rangle_0.
$$

We define the estimator $\widehat{\theta}$ obtained by minimizing this criterion. First we show that $\widehat{\mathrm{FD}}(f_\theta, I)$ converges in probability to $\mathrm{FD}(f_\theta, \widetilde{f})$, plus a positive bias term.

Proposition 3. *Assume that $\{x_t\}$ is strictly stationary with spectral density $\widetilde{f}$, and satisfies condition (B1). Then for all $\theta \in \Theta$*

$$
\widehat{FD}(f_\theta, I) \xrightarrow{P} |\widetilde{f} - f_\theta|^2 + \langle [\widetilde{f}]^2 \rangle_0.
$$

So long as this criterion is continuous with respect to θ, then $\widehat{\theta}$ is consistent for the PTV, which is the minimizer of $|\widetilde{f} - f_\theta|$. (Note that the minimizer of $|\widetilde{f} - f_\theta|$ is equal to the minimizer of $|\widetilde{f} - f_\theta|^2 + \langle [\widetilde{f}]^2 \rangle_0$ as well, because the second term is free of θ.) The PTV is denoted $\widetilde{\theta}$. We also define the matrix M_θ with jkth entry given by $[\langle \partial_{\theta_j} f_\theta \, \partial_{\theta_k} f_\theta \rangle_0]$.

Theorem 2. *Assume that $\{x_t\}$ is strictly stationary with spectral density $\widetilde{f}$ and satisfies condition (B1). Also suppose that $\widehat{FD}(f_\theta, I)$ is a twice continuously differentiable function of θ, that the PTV $\widetilde{\theta}$ exists and is unique, and is in the interior of the parameter space. If $M_{\widetilde{\theta}}$ is invertible, then as $T \to \infty$*

$$
\sqrt{T}\,(\widehat{\theta} - \widetilde{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, M_{\widetilde{\theta}}^{-1} V_{\widetilde{\theta}} M_{\widetilde{\theta}}^{-1}),
$$

where V_θ is a matrix with jkth entry given by $2\,[\langle \widetilde{f} \, \partial_{\theta_j} f_\theta \, \widetilde{f} \, \partial_{\theta_k} f_\theta \rangle_0]$.

Remark 2. For a short-hand, we will write $M_\theta = [\langle \nabla f_\theta \, \nabla' f_\theta \rangle_0]$, it being understood that the trace operator does not act upon the gradients. Likewise, we write $V_\theta = 2\,[\langle \widetilde{f} \, \nabla f_\theta \, \widetilde{f} \, \nabla' f_\theta \rangle_0]$.

4.2   Structural Models

We consider structural models (Harvey, 1989) for a time series $\{x_t\}$ that has been differenced to stationarity. As shown in McElroy (2017), the model spectrum takes the form

$$
f_\theta(\lambda) = \sum_{k=1}^{K} g_k(\lambda) \Theta_k, \tag{1}
$$

10

for scalar parameter-free real-valued even functions g_k, and positive definite (pd) parameter matrices Θ_k. Writing $\Theta = [\Theta_1, \Theta_2, \ldots, \Theta_K]$, we have the parameter $\theta = \text{vec}(\Theta)$ by definition. (There is redundancy in θ, but it is more convenient to avoid using the vech operator.) Let G be the matrix given by

$$G_{ik} = \langle g_i g_k \rangle_0.$$

Also let g denote the vector of scalar functions g_k. Then we define the MOM estimator (cf. McElroy (2017)) to be the sample analogue of the minimizer with respect to θ of $|\widetilde{f} - f_\theta|$, where we allow the components of θ to be real-valued. That is, we do not enforce the p.d. constraints on θ.

Proposition 4. *The MOM estimator for the structural model (1) has formula*

$$\widehat{\Theta} = \langle g' \otimes I \rangle_0 \cdot \left[G^{-1} \otimes 1_m \right] = \langle g' G^{-1} \otimes I \rangle_0, \tag{2}$$

where 1_m denotes the identity matrix of dimension m.

Note that the MOM estimator in (2) is defined using the integral over all frequencies, and not the average over Fourier frequencies, as the former is more convenient for computation – and the same asymptotic theory applies, because for linear functionals of the periodogram there is an asymptotic equivalency between integration and averaging over Fourier frequencies. For computation, we have

$$\widehat{\Theta}_k = \langle g' G^{-1} e_k \otimes I \rangle_0,$$

where e_k is the kth unit vector. These estimators are very easy to calculate, amounting to just fixed linear combinations of sample autocovariances, but are not guaranteed to be p.d. However, the estimators are symmetric, which is proved using the property that I is Hermitian and that each g_k is an even function of λ. The whole vector of estimates is $\widehat{\theta} = \text{vec}(\widehat{\Theta})$; the corresponding true parameter is denoted $\widetilde{\theta}$, and does indeed correspond to p.d. $\widetilde{\Theta}_k$. Replacing I by the true spectral density in Proposition 4, it is immediate that the PTV exists whenever $\widetilde{\Theta}_k$ is p.d., and in such a case the PTV is unique. When the model is correct, the secondary conditions of Theorem 2 are satisfied, and the asymptotic theory for $\widehat{\theta}$ is fairly straightforward.

Theorem 3. *Assume that $\{x_t\}$ is strictly stationary with spectral density $\widetilde{f}$ and satisfies condition (B1). Then $\widehat{\theta}$ is consistent for θ and*

$$\sqrt{T} \left(\widehat{\theta} - \widetilde{\theta} \right) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, M^{-1} V M^{-1}),$$

where $M = G \otimes 1_{m^2}$ and $V = 2 \langle gg' \otimes f_{\widetilde{\theta}} \otimes f'_{\widetilde{\theta}} \rangle_0$.

Remark 3. From equation (1) it follows that

$$f_\theta \otimes f'_\theta = \sum_{\ell, k=1}^{K} g_\ell g_k \, \Theta_\ell \otimes \Theta_k,$$

and hence the ijth block of V, of dimension $m^2 \times m^2$, is

$$V_{ij} = 2 \sum_{\ell,k=1}^{K} \langle g_i g_j g_\ell g_k \rangle \, \Theta_\ell \otimes \Theta_k.$$

Substituting the estimates $\widehat{\Theta}$ then yields an estimator $\widehat{V}$, which can be swiftly calculated once the four-array $\langle g_i g_j g_\ell g_k \rangle$ has been determined. Moreover $\widehat{V} \xrightarrow{P} V$ follows from the consistency of $\widehat{\theta}$ in Theorem 3.

4.3 Simulation results

To evaluate the finite sample performance of the MOM estimator (2) we used a structural model and tabulated the mean squared error of the estimators. Specifically, we used a model with a trend $\{\mu_t\}$, seasonal $\{\xi_t\}$ (of period s), and irregular $\{\iota_t\}$ components, which are related to the observed process $\{x_t\}$ via

$$x_t = \mu_t + \xi_t + \iota_t. \tag{3}$$

Here $(1 - B)\mu_t = \eta_t$, $(1 + B + B^2 + \ldots + B^{s-1})\xi_t = \zeta_t$, and $\{\eta_t\}$, $\{\zeta_t\}$, and $\{\iota_t\}$ are each independent Gaussian vector white noise processes, with mean zero and variance matrices Σ_μ, Σ_ξ and Σ_ι, respectively. For simulation the variance matrices are chosen as

$$\Sigma_\mu = \begin{bmatrix} \sigma_{1,\mu}^2 & \rho_\mu \sigma_{1,\mu} \sigma_{2,\mu} \\ \rho_\mu \sigma_{1,\mu} \sigma_{2,\mu} & \sigma_{2,\mu}^2 \end{bmatrix}, \quad \Sigma_\xi = \begin{bmatrix} \sigma_{1,\xi}^2 & \rho_\xi \sigma_{1,\xi} \sigma_{2,\xi} \\ \rho_\xi \sigma_{1,\xi} \sigma_{2,\xi} & \sigma_{2,\xi}^2 \end{bmatrix}, \quad \Sigma_\iota = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

In model (3) the differencing functions are $g_1 = (1 - \cos(s\lambda))(1 - \cos(\lambda))^{-1}$, $g_2(\lambda) = 2(1 - \cos(\lambda))$, and $g_3(\lambda) = 2(1 - \cos(s\lambda))$. In terms of the notation of Section 4.2, the variance component matrices are $\Theta_1 = \Sigma_\mu$, $\Theta_2 = \Sigma_\xi$ and $\Theta_3 = \Sigma_\iota$. The terms $\langle g_i g_j g_\ell g_k \rangle$ used in the variance expression as in Theorem 3 were computed using the integrate function in R. The variance parameters were set to $\sigma_{1,\mu} = \sigma_{1,\xi} = 1$, $\sigma_{2,\mu} = 0.8$, $\sigma_{2,\xi} = 0.6$, $\rho_\mu \in \{0, 0.6, 0.9, 0.95, 1\}$, and $\rho_\xi \in \{0, 0.4, 0.8, 0.9, 1\}$. We generated $n = 5000$ Monte Carlo replications, and recorded the efficiency of the MOM estimators as measured by their Monte Carlo root mean squared error (RMSE). The Monte Carlo RMSE for a MOM estimator $\widehat{\Theta}_k$ for a variance component Θ_k is defined as

$$\text{RMSE}(\widehat{\Theta}_k) = n^{-1} \sum_{i=1}^{n} \|\widehat{\Theta}_k - \Theta_k\|^2.$$

The sample sizes were varied between $T \in \{200, 500\}$. Table 1 shows the efficiency of the MOM estimators for model (3). The RMSE values decline when sample size increases from $T = 200$ to $T = 500$ across all parameters by a factor close to $\sqrt{500/200}$, corroborating the asymptotic $\sqrt{T}$ rate of convergence. The RMSE for the trend component is generally the smallest while that for the irregular component is the largest.

12

	ρ_ξ	0.00	0.40	0.80	0.90	1.00	0.00	0.40	0.80	0.90	1.00
	ρ_μ			$T = 200$					$T = 500$		
Σ_μ	0.00	0.195	0.193	0.196	0.198	0.198	0.125	0.126	0.128	0.124	0.124
	0.60	0.208	0.206	0.210	0.211	0.205	0.134	0.129	0.130	0.135	0.132
	0.90	0.221	0.211	0.211	0.214	0.222	0.141	0.136	0.141	0.136	0.136
	0.90	0.226	0.210	0.221	0.226	0.217	0.140	0.141	0.137	0.134	0.137
	1.00	0.220	0.222	0.227	0.220	0.221	0.139	0.137	0.137	0.137	0.136
Σ_ξ	0.00	0.282	0.272	0.270	0.268	0.271	0.186	0.185	0.175	0.172	0.173
	0.60	0.290	0.290	0.289	0.296	0.295	0.190	0.192	0.185	0.187	0.192
	0.90	0.285	0.297	0.302	0.300	0.308	0.194	0.194	0.199	0.196	0.195
	0.95	0.288	0.298	0.302	0.311	0.303	0.194	0.197	0.198	0.205	0.196
	1.00	0.280	0.284	0.307	0.306	0.306	0.195	0.196	0.196	0.202	0.201
Σ_ι	0.00	0.469	0.467	0.463	0.470	0.461	0.312	0.315	0.315	0.314	0.309
	0.60	0.490	0.458	0.478	0.467	0.482	0.319	0.314	0.309	0.320	0.321
	0.90	0.482	0.480	0.488	0.482	0.492	0.323	0.325	0.334	0.317	0.318
	0.95	0.500	0.481	0.485	0.499	0.509	0.332	0.321	0.314	0.334	0.324
	1.00	0.475	0.471	0.475	0.488	0.476	0.325	0.320	0.321	0.321	0.325

Table 1: Efficiency (root mean squared error) for the different variance component matrices using MOM estimation for model (3)

4.4 Rank testing

An interesting application of the asymptotic normal theory developed in Theorem 3 is that of testing for reduced rank structures of the variance matrices. A reduced rank variance matrix for any of the structural components would indicate a corresponding co-integration effect (McElroy, 2017). Following McElroy and Jach (2017), testing whether variance matrices have reduced rank can proceed via adopting the null hypothesis that one of the diagonal entries in the generalized Cholesky Decomposition (Golub and Van Loan, 1996) is zero. These diagonal entries correspond to sequential Schur complements, computed with respect to each upper left sub-matrix (McElroy, 2018), and therefore can be expressed as a ratio of determinants. Hence these Schur complements are smooth functions of the entries of a matrix, and we can use the delta method in conjunction with Theorem 3 to perform a reduced rank test.

For the present application we consider bivariate time series, and rank reduction is equivalent to the determinant being zero (see McElroy and Jach (2017) for further discussion). We test whether Σ_μ or Σ_ξ have rank one by checking if the determinant of the matrix is zero. Specifically, we test

$$H_0 : \det(\Theta_k) = 0 \quad vs \quad H_1 : \det(\Theta_k) \neq 0$$

for variance component Θ_k, for $k = 1, 2$. Because $N = 2$ the variance components have the form

$$\Theta_k = \begin{bmatrix} \Theta_{k,11} & \Theta_{k,12} \\ \Theta_{k,21} & \Theta_{k,22} \end{bmatrix},$$

and the estimated asymptotic variance of $\det \widehat{\Theta}_k$ will be $v_k = b'\, W_{kk}\, b$, where $b = (\widehat{\Theta}_{k,22}, -2\widehat{\Theta}_{k,21}, \widehat{\Theta}_{k,11})'$ and W_{kk} is the diagonal block of the variance matrix $M^{-1}VM^{-1}$ in Theorem 3 associated with the entries of Θ_k, with parameters replaced by their MOM estimates.

For simulation we used model (3) with the parameter specifications given in the previous subsection. Note that for either Σ_μ or Σ_ξ, when the correlation parameter is equal to one, the matrices are of reduced rank (i.e., rank one). Otherwise, for all other values the matrices have full rank (i.e., rank two). Thus the type I error rate will be reached when the correlation is one, whereas when the correlation moves away from the unity the power of the test will increase. Table 2 shows the size and power of the test for reduced rank component for both the trend and the seasonal components. The values corresponding to null values are identified in bold. The tests are generally conservative, particularly at smaller sample sizes, with the test for the seasonal component being more conservative than that for the trend component. The power for the trend component test rises more rapidly than that of the seasonal component as the parameter value moves away from the null.

		$H_0 : \det(\Sigma_\mu) = 0$					$H_0 : \det(\Sigma_\xi) = 0$				
		ρ_ξ					ρ_ξ				
		0.00	0.40	0.80	0.90	1.00	0.00	0.40	0.80	0.90	1.00
		$T = 200$					$T = 200$				
	0.00	1.000	1.000	0.999	0.998	0.998	0.091	0.050	0.015	0.011	**0.012**
	0.60	0.976	0.982	0.993	0.989	0.993	0.115	0.056	0.024	0.014	**0.010**
ρ_μ	0.90	0.367	0.421	0.468	0.479	0.484	0.122	0.091	0.027	0.019	**0.005**
	0.95	0.125	0.146	0.171	0.167	0.180	0.105	0.093	0.025	0.014	**0.009**
	1.00	**0.020**	**0.014**	**0.019**	**0.015**	**0.007**	0.136	0.097	0.021	0.014	**0.012**
		$T = 500$					$T = 500$				
	0.00	1.000	1.000	1.000	1.000	1.000	0.506	0.213	0.052	0.046	**0.019**
	0.60	1.000	1.000	1.000	1.000	1.000	0.462	0.323	0.064	0.037	**0.016**
ρ_μ	0.90	0.870	0.910	0.940	0.939	0.944	0.453	0.422	0.076	0.067	**0.029**
	0.95	0.436	0.514	0.548	0.555	0.575	0.437	0.435	0.097	0.057	**0.030**
	1.00	**0.039**	**0.025**	**0.032**	**0.036**	**0.027**	0.397	0.451	0.081	0.047	**0.023**

Table 2: Power of the test of reduced rank structure based on MOM estimates. Type I errors at the null values are given in bold.

5   Model Evaluation

5.1   Methodology

It is possible to consider the gof test statistic $\widehat{\mathrm{FD}}(f_\theta, I)$, where the parameter estimates are those described above (or are maximum likelihood estimators, or some other $\sqrt{T}$-consistent estimators). However, the asymptotic bias in $\widehat{\mathrm{FD}}(f_\theta, I)$ of $\langle[\widetilde{f}]^2\rangle_0$ means that the distribution theory becomes fairly complex; instead we consider model evaluation via assessment of the spectral residual. Instead of analyzing $I/f_{\widehat{\theta}}$ directly, we consider the periodogram J of the residual process. This may be preferable in scenarios where the fitted model involves transformations, fixed regressors, and stochastic effects, such that f_θ is only a partial description of the model.

Let J denote the periodogram of the estimated residuals, with $\widetilde{g}$ denoting the spectral density of the true residual process, and consider the testing problem $|\widetilde{g}|^2 - \|\langle\widetilde{g}\rangle_0\|^2 = 0$; this quantity is always non-negative, but equals zero if and only if $\widetilde{g}$ corresponds to white noise. Rewriting this functional as

$$|\widetilde{g}|^2 - \|\langle\widetilde{g}\rangle_0\|^2 = Q_{\mathrm{id,id}}(\widetilde{g}, \widetilde{g}) - [\langle\widetilde{g}\rangle_0^2]$$

it can be seen that substituting J for $\widetilde{g}$ yields a statistic that converges to $|\widetilde{g}|^2 + \langle[\widetilde{g}]^2\rangle_0 - [\langle\widetilde{g}\rangle_0^2]$ (this is shown in the proof of Proposition 5, below). In the case that $\widetilde{g} \equiv \Sigma$, corresponding to a white noise process, the limit reduces to $[\Sigma]^2$, which is nonzero except in trivial cases. In order to obtain a statistic that converges to zero (the null hypothesis value of $|\widetilde{g}|^2 - \|\langle\widetilde{g}\rangle_0\|^2$), we propose to subtract the quantity $[\langle J\rangle_0]^2$, which converges to $[\Sigma]^2$ under the null. These arguments suggest defining a model evaluation functional $\mathrm{Eval}(\widetilde{g})$ and estimator $\widehat{\mathrm{Eval}}(J)$ as follows:

$$\mathrm{Eval}(\widetilde{g}) = |\widetilde{g}|^2 - \|\langle\widetilde{g}\rangle_0\|^2 + \langle[\widetilde{g}]^2\rangle_0 - [\langle\widetilde{g}\rangle_0]^2 = Q_{\mathrm{id,id}}(\widetilde{g}, \widetilde{g}) + \langle[\widetilde{g}]^2\rangle_0 - [\langle\widetilde{g}\rangle_0^2] - [\langle\widetilde{g}\rangle_0]^2$$
$$\widehat{\mathrm{Eval}}(J) = |J|^2 - \|\langle J\rangle_0\|^2 - [\langle J\rangle_0]^2 = \widehat{Q}_{\mathrm{id,id}}(J, J) - [\langle J\rangle_0^2] - [\langle J\rangle_0]^2.$$

The estimator $\widehat{\mathrm{Eval}}(J)$ can be alternatively expressed with the final two terms, $[\langle J\rangle_0^2]$ and $[\langle J\rangle_0]^2$, having their integral replaced by a sum over Fourier frequencies; this can make calculation easier. However, because these two terms involve integrals/sums over a linear function of the periodogram, the asymptotic theory is the same, whether we use integrals or sums (c.f., Chen and Deo (2000)). The evaluation estimator converges in probability to the correct quantity for the testing problem, as shown below.

Proposition 5. *Assume that $\{x_t\}$ is strictly stationary with spectral density $\widetilde{f}$, and satisfies condition (B1). Then as $T \to \infty$,*

$$\widehat{\mathrm{Eval}}(J) \xrightarrow{P} \mathrm{Eval}(\widetilde{g}).$$

Also, $\mathbb{E}\widehat{\mathrm{Eval}}(J)$ converges to the same limit.

Remark 4. Proposition 5 does not assume that the null hypothesis is true. But if the residual is white noise, then $\widetilde{g} \equiv \Sigma$, and

$$\text{Eval}(\widetilde{g}) = |\Sigma|^2 - \|\langle\Sigma\rangle_0\|^2 + \langle[\Sigma]^2\rangle_0 - [\langle\Sigma\rangle_0]^2 = 0.$$

Evidently, $\widehat{\text{Eval}}(J)$ can be negative, because of the presence of the bias-correction term $[\langle J\rangle_0]^2$.

In order to test the null hypothesis, we need a limit theory for the model evaluation estimator. Unlike the case of the LB statistic, based upon a finite number of sample autocovariances, in this case the limit distribution is normal – this is essentially due to the inclusion of the bias-correction term $[\langle J\rangle_0]^2$.

Theorem 4. *Assume that $\{x_t\}$ is strictly stationary with spectral density $\widetilde{g}$ and satisfies condition (B1). Then as $T \to \infty$*

$$\sqrt{T}\left(\widehat{\text{Eval}}(J) - \text{Eval}(\widetilde{g})\right) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, v'\, W\, v)$$

$$W = \begin{bmatrix} V_{\langle\widetilde{g}\rangle_0|\langle\widetilde{g}\rangle_0} & V_{\langle\widetilde{g}\rangle_0|id} & V_{\langle\widetilde{g}\rangle_0|id,id} \\ V_{id|\langle\widetilde{g}\rangle_0} & V_{id|id} & V_{id|id,id} \\ V_{id,id|\langle\widetilde{g}\rangle_0} & V_{id,id|id} & V_{id,id|id,id} \end{bmatrix},$$

where $v' = \{-2, -2[\langle\widetilde{g}\rangle_0], 1\}$ and the block entries of W (which is symmetric) are given by

$$V_{\langle\widetilde{g}\rangle_0|\langle\widetilde{g}\rangle_0} = 2\,\langle[\langle\widetilde{g}\rangle_0\,\widetilde{g}\,\langle\widetilde{g}\rangle_0\,\widetilde{g}]\rangle_0 + (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}[[\langle\widetilde{g}\rangle_0\,\widetilde{g}(\lambda, -\lambda, \omega)\,\langle\widetilde{g}\rangle_0]]\,d\lambda\,d\omega,$$

$$V_{id|\langle\widetilde{g}\rangle_0} = 2\,\langle[\widetilde{g}\,\langle\widetilde{g}\rangle_0\,\widetilde{g}]\rangle_0 + (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}[[\widetilde{g}(\lambda, -\lambda, \omega)\,\langle\widetilde{g}\rangle_0]]\,d\lambda\,d\omega,$$

$$V_{id|id} = 2\,\langle[\widetilde{g}^2]\rangle_0 + (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}[[\widetilde{g}(\lambda, -\lambda, \omega)]]\,d\lambda\,d\omega,$$

$$V_{\langle\widetilde{g}\rangle_0|id,id} = 4\,\langle[\widetilde{g}]\,[\widetilde{g}\,\langle\widetilde{g}\rangle_0\,\widetilde{g}]\rangle_0 + 4\,\langle[\widetilde{g}^2\,\langle\widetilde{g}\rangle_0\,\widetilde{g}]\rangle_0 + (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}g_{123,\langle\widetilde{g}\rangle_0}(\lambda, \omega)\,d\lambda\,d\omega,$$

$$g_{123,\langle\widetilde{g}\rangle_0}(\lambda, \omega) = 2\,[[\widetilde{g}(\lambda, -\lambda, \omega)\,\langle\widetilde{g}\rangle_0]]\,[\widetilde{g}(\lambda)] + 2\,[[\widetilde{g}(\lambda)\,\widetilde{g}(\lambda, -\lambda, \omega)\,\langle\widetilde{g}\rangle_0]],$$

$$V_{id|id,id} = 4\,\langle[\widetilde{g}]\,[\widetilde{g}^2]\rangle_0 + 4\,\langle[\widetilde{g}^3]\rangle_0 + (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}g_{123,id}(\lambda, \omega)\,d\lambda\,d\omega,$$

$$g_{123,id}(\lambda, \omega) = 2\,[[\widetilde{g}(\lambda, -\lambda, \omega)]]\,[\widetilde{g}(\lambda)] + 2\,[[\widetilde{g}(\lambda)\,\widetilde{g}(\lambda, -\lambda, \omega)]],$$

$$V_{id,id|id,id} = 8\,\langle[\widetilde{g}]^2\,[\widetilde{g}^2]\rangle_0 + 16\,\langle[\widetilde{g}]\,[\widetilde{g}^3]\rangle_0 + 12\,\langle[\widetilde{g}]^4\rangle_0 + 4\,\langle[\widetilde{g}^2]^2\rangle_0 + (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}g_{1234}(\lambda, \omega)\,d\lambda\,d\omega,$$

$$g_{1234}(\lambda, \omega) = 4\,[[\widetilde{g}(\lambda, -\lambda, \omega)]]\,[\widetilde{g}(\lambda)]\,[\widetilde{g}(\omega)] + 4\,[[\widetilde{g}(\lambda, -\lambda, \omega)\,\widetilde{g}(\omega)]]\,[\widetilde{g}(\lambda)]$$
$$+ 4\,[[\widetilde{g}(\lambda)\,\widetilde{g}(\lambda, -\lambda, \omega)]]\,[\widetilde{g}(\omega)] + 4\,[[\widetilde{g}(\lambda)\,\widetilde{g}(\lambda, -\lambda, \omega)\,\widetilde{g}(\omega)]].$$

In general, calculation of W is daunting due to the complexity of the covariances appearing in Theorem 4, but under the null hypothesis of white noise there is a remarkable simplification to the limiting variance – all the dependence on the tri-spectrum vanishes.

Corollary 1. *Assume that $\{x_t\}$ is strictly stationary with white noise spectral density $\widetilde{g} = \Sigma$, and satisfies condition (B1). Then as $T \to \infty$*

$$\sqrt{T}\,\widehat{Eval}(J) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, 4\,[\Sigma^4] + 4\,[\Sigma^2]^2).$$

Corollary 1 can be applied by substituting the sample variance matrix of the residual process, i.e., $\widehat{\Sigma} = \widehat{\Gamma}(0)$, in the expression for the asymptotic variance.

5.2 Simulation results

To test the utility of Corollary 1 in finite sample, we use the evaluation criterion to check model order specification in a two dimensional VAR(2) process. Specifically, the true model, chosen for data generation, was

$$x_t = \begin{bmatrix} 0.3 & -0.3 \\ 0 & 0.4 \end{bmatrix} x_{t-1} + \begin{bmatrix} -0.01 & -0.1 \\ -0.1 & 0.25 \end{bmatrix} x_{t-2} + z_t, \tag{4}$$

where $\{z_t\}$ were Gaussian white noise with mean zero and variance I_2, the two dimensional identity matrix. The roots of the VAR polynomial are 0.812, -0.338, and $0.113 \pm 0.121i$. The sample sizes explored were $T = 200$, 500, and 1000. The number of Monte Carlo replications was 5000. After generating the data from the VAR(2) specification, we repeatedly fit the data using a VAR(p) model for $p \in \{1, 2, \ldots, 8\}$. Thus, for $p = 1$ the model will be mis-specified (and rejections pertain to empirical power), whereas for $p \geq 2$ the models are correctly specified. However, as p grows the properties of the test in over-specified models are affected by finite sample terms, and hence the nominal size level may be violated. Table 3 provides the proportion of empirical rejections for two sided tests of correct model order, using the limiting distribution of $\widehat{Eval}(J)$ given in Corollary 1, where the error variance Σ is estimated from the residuals. Of course in the mis-specified models the estimate of Σ will be also affected by the mis-specification.

The variance expression in Corollary 1 holds even when the innovations have a non-Gaussian distribution. To check the power properties of the test, we repeated the simulation exercise for model (4) when the innovations for the true model are generated from a multivariate Student's t distribution with 4 degrees of freedom and identity as the scale matrix. The results are given in Table 4; the proportions are remarkably close to the corresponding values for the Gaussian case, indicating the robustness of the proposed test against distributional assumptions.

6 Empirical Analyses

6.1 Bivariate Inflation

We examine the bivariate Personal Consumption Expenditures (PCE) inflation data discussed in McElroy and Trimbur (2015). The first series measures core inflation (excluding food and energy

p	$T = 200$	$T = 500$	$T = 1000$
1	0.089	0.217	0.697
2	0.024	0.043	0.045
3	0.027	0.050	0.055
4	0.038	0.059	0.059
5	0.062	0.073	0.064
6	0.078	0.089	0.064
7	0.096	0.108	0.066
8	0.103	0.132	0.082

Table 3: Size and power of the model evaluation test, computed based on residuals from a VAR(p) to the VAR(2) model (4), for different model orders and different sample sizes. Values of $p \geq 2$ correspond to size, and should be close to the nominal level of 5%, whereas values in the row for $p = 1$ correspond to power. The innovations for the true model are generated from a Gaussian distribution.

p	$T = 200$	$T = 500$	$T = 1000$
1	0.062	0.183	0.376
2	0.029	0.048	0.051
3	0.050	0.066	0.059
4	0.062	0.080	0.077
5	0.085	0.100	0.084
6	0.100	0.127	0.098
7	0.132	0.139	0.130
8	0.146	0.176	0.139

Table 4: Size and power of the model evaluation test, computed based on residuals from a VAR(p) to the VAR(2) model (4), for different model orders and different sample sizes. Values of $p \geq 2$ correspond to size, and should be close to the nominal level of 5%, whereas values in the row for $p = 1$ correspond to power. The innovations for the true model are generated from a multivariate Student's t with 4 degrees of freedom and identity as the scale matrix.

items) whereas the second series is total inflation; the data were obtained from the Bureau of Economic Analysis, and covers 1986Q1 through 2010Q4. Changes in level to both series over the sample period indicate the possible presence of a stochastic trend; as in McElroy and Trimbur (2015), we proceed by fitting a structural model (3) with only random walk trend and irregular components, because the seasonal component $\{\xi_t\}$ is not needed.

Because the dimension and sample size are small, direct MLE is feasible. We computed MLE results for comparison with the MOM estimators, examining both an unrestricted model (i.e., the basic model) and the common trends restriction, whereby Σ_μ is enforced to have rank one. These results are summarized in Appendix B of the Supplementary Material. While parameter

estimates do indeed differ between the MOM and MLE, the former approach does effect a whitening transformation on the data, and at the same time is considerably faster – less than a second (on our machine) is required to compute the MOM estimators, whereas over a minute is required for the MLE optimization routine. In order to assess the model gof, we applied the white noise test to the PCE data, fitted according to the MOM method: $\sqrt{T}\,\widehat{\text{Eval}}(J) = -1.893$, with the variance estimated at 475.458. Hence the model cannot be rejected (normalized test statistic is -0.0868, with p-value .465).

Both the MOM and MLE estimates of Σ_μ indicate the possibility of common trends, as the former estimate has a cross-correlation of .784, whereas the latter is .999906. The MOM correlation estimate is not particularly close to unity, though if we account for variability it is possible that the rank one hypothesis cannot be rejected. For the test of common trend the results were 741.3569 for the $\det(\Theta_k)$, with an estimated variance of $500,699.6$, and a normalized test statistic of 1.047704. Hence the null hypothesis of common trends cannot be rejected.

6.2 Four-variate Housing Starts

Our second empirical illustration involves housing starts, which are published by the U.S. Census Bureau on a monthly basis, for the regions corresponding to South, West, Northeast (NE), and Midwest (MW). As in McElroy (2017), we study "New Residential Construction 1964–2012, Housing Units Started, Single Family Units" from the Survey of Construction of the U.S. Census Bureau, available at `http://www.census.gov/construction/nrc/how_the_data_are_collected/soc.html`. Because of the presence of both a highly dynamic trend and seasonal, a structural model (3) with second order stochastic trend ($d = 2$) is used along with a seasonal $\{\xi_t\}$ that is additively composed of six atomic seasonal processes, one for each principal monthly seasonal frequency. The model involves eight latent components, each of which is specified by a 4×4-dimensional covariance matrix, yielding a total of 80 parameters.

With a restricted sample size corresponding to the latest 9 years, we were able to run MLE for an unrestricted model, and make comparisons to the MOM estimates. These results are summarized in Appendix B of the Supplementary Material. Again, parameter estimates differ between the MOM and MLE, but the former approach does effect a whitening transformation on the data. The difference in speed is significant: 78 minutes for the MLE optimization, versus less than one second for MOM calculation. We also fitted the model using MOM applied to the entire data span of 49 years; in this case, the MOM calculation is still less than one second, whereas a single Gaussian likelihood evaluation (with an efficient Durbin-Levinson implementation) takes more than a second. In order to assess the model gof, we applied the white noise test, obtaining $\sqrt{T}\,\widehat{\text{Eval}}(J) = -7.728$, with the variance estimated at 48.042. Hence the model cannot be rejected.

References

[1] Akashi, F., Odashima, H., Taniguchi, M. and Monti, A.C. (2017) A new look at portmanteau tests. *Sankhya A*, 1–17.

[2] Brillinger, D.R. (2001) *Time Series: Data Analysis and Theory*. Philadelphia: Siam.

[3] Chen, W. and Deo, R. (2004) A generalized portmanteau goodness-of-fit test for time series models. *Econometric Theory* **20**, 382–416.

[4] Chen, W. and Deo, R. (2000) On the integral of the squared periodogram. *Stochastic Processes and Their Applications* **85**, 159–176.

[5] Davis, H.T. and Jones, R.H. (1968) Estimation of the innovation variance of a stationary time series. *Journal of the American Statistical Association* **63**, 141–149.

[6] Drouiche, K. (2007) A test for spectrum flatness. *Journal of Time Series Analysis* **28**(6), 793–806.

[7] Golub G. and Van Loan, C. (1996) *Matrix Computations*. Baltimore: Johns Hopkins University Press.

[8] Harvey, A. (1989) *Forecasting, Structural Time Series Models, and the Kalman Filter*. Cambridge: Cambridge University Press.

[9] Hosking, J. R. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association* **75**, 602–608.

[10] Kohli, P. and Pourahmadi, M. (2012) Nonparametric estimation of the innovation variance and judging the fit of ARMA models. In W. Bell, S. Holan, and T. McElroy (Eds.), *Economic Time Series: Modeling and Seasonality*. New York: Chapman and Hall.

[11] Li, W.K. (2004) *Diagnostic Checks in Time Series*. CRC Press.

[12] Li, W. K. and McLeod, A. I. (1981) Distribution of the residual autocorrelations in multivariate ARMA time series models. *Journal of the Royal Statistical Society, Series B* **43**, 231–239.

[13] Ljung, G. and Box, G. (1978) On a measure of lack of fit in time series models. *Biometrika* **65**, 297-303.

[14] McElroy, T.S. (2016) Non-nested model comparisons for time series. *Biometrika* **103**, 905–914.

[15] McElroy, T.S. (2017) Multivariate seasonal adjustment, economic identities, and seasonal taxonomy. *Journal of Business and Economics Statistics* **35**(4), 511-525.

[16] McElroy, T.S. (2018) Recursive computation for block nested covariance matrices. Published online, *Journal of Time Series Analysis*.

[17] McElroy, T.S. and Findley, D.F. (2015) Fitting constrained vector autoregression models. In J. Beran, Y. Feng, and H. Hebbel (Eds.), *Empirical Economic and Financial Research – Theory, Methods, and Practice*, pp.451–470. New York: Springer.

[18] McElroy, T.S. and Holan, S. (2009a) A local spectral approach for assessing time series model misspecification. *Journal of Multivariate Analysis* **100**, 604–621.

[19] McElroy, T. and Holan, S. (2009b) Spectral domain diagnostics for testing model proximity and disparity in time series data. *Statistical Methodology* **6**, 1–20.

[20] McElroy, T. and Jach, A. (2017) Testing collinearity of vector time series. Mimeo.

[21] McElroy, T. and Monsell, B. (2014) The multiple testing problem for Box-Pierce statistics. *Electronic Journal of Statistics* **8**, 497–522.

[22] McElroy, T.S. and Politis, D. (2014) Spectral density and spectral distribution inference for long memory time series via fixed-b asymptotics. *Journal of Econometrics* **182**, 211–225.

[23] McElroy, T.S. and Trimbur, T. (2015) Signal extraction for nonstationary multivariate time series with illustrations for trend inflation. *Journal of Time Series Analysis* **36**, 209–227. Also in "Finance and Economics Discussion Series," Federal Reserve Board. 2012-45 http://www.federalreserve.gov/pubs/feds/2012/201245/201245abs.html

[24] McElroy, T. and Wildi, M. (2013) Multi-step ahead estimation of time series models. *International Journal of Forecasting* **29**, 378–394.

[25] Paparoditis, E. (2000) Spectral density based goodness-of-fit tests for time series models. *Scandinavian Journal of Statistics* **27**, 143–176.

[26] Peña, D. and Rodriguez, J. (2002) A powerful portmanteau test of lack of fit for time series. *Journal of the American Statistical Association* **97**, 601–610.

[27] Rivers, D. and Vuong, Q. (2002) Model selection tests for nonlinear dynamic models. *Econometrics Journal* **5**, 1–39.

[28] Taniguchi, M., and Kakizawa, Y. (2000) *Asymptotic Theory of Statistical Inference for Time Series.* Springer-Verlag, New York.

[29] Vuong, Q. (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**, 307–333.

<div align="center">

Supplement to

Model Identification via Total Frobenius Norm of Multivariate Spectra

</div>

<div align="center">

Tucker S. McElroy[*]and Anindya Roy[†]

</div>

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

Appendix A   Proofs

Proof of Proposition 1. Without loss of generality, we replace the sample mean in the DFT by the true mean, because the error in doing so is of lower order. The result for linear functionals is known (Lemma 3.1.1 of Taniguchi and Kakizawa (2000)). For the quadratic case, we note that

$$Q_{\varphi_1,\varphi_2}(I,I) = \langle [\varphi_1 \, I \, \varphi_2 \, I] \rangle_0 = T^{-2} \, \langle (d^* \, \varphi_1 \, d) \, (d^* \, \varphi_2 \, d) \rangle_0,$$

which by Lemma P5.1 of Brillinger (2001) approximates the estimator $\widehat{Q}_{\varphi_1,\varphi_2}(I,I)$, with error of order T^{-1}. Based on such an approximation, one could work with the estimator $Q_{\varphi_1,\varphi_2}(I,I)$ instead of $\widehat{Q}_{\varphi_1,\varphi_2}(I,I)$, but as mentioned in the text it is simpler to focus upon the latter. To that end, we write

$$\widehat{Q}_{\varphi_1,\varphi_2}(I,I) = T^{-3} \sum_{j=1}^{T} (d^* \, \varphi_1 \, d)(\lambda_j) \, (d^* \, \varphi_2 \, d)(\lambda_j). \tag{A.1}$$

Here, λ_j is a Fourier frequency defined as $\lambda_j = 2\pi j/T - \pi$. (The subtraction by π ensures that we stay in the interval $[-\pi, \pi]$.) Denote entry r, s of φ via $\varphi(r, s; \lambda)$; each of these components are a function of λ. Then

$$\mathbb{E}\widehat{Q}_{\varphi_1,\varphi_2}(I,I) = T^{-3} \sum_{j=1}^{T} \sum_{\ell,k,r,s} \varphi_1(\ell, k; \lambda_j) \, \varphi_2(r, s; \lambda_j) \, \mathbb{E}[d_\ell^*(\lambda_j) \, d_k(\lambda_j) \, d_r^*(\lambda_j) \, d_s(\lambda_j)],$$

and we can apply Theorem 2.3.2 of Brillinger (2001) to the inner expectation. Equation (4.3.15) of Brillinger (2001) yields

$$\mathrm{cum}(d_\ell^*(\lambda_j), d_k(\lambda_i)) = O(1) + \Delta^{(T)}(\lambda_j - \lambda_i) \, \widetilde{f}_{\ell k}^*(\lambda_j), \tag{A.2}$$

where $\Delta^{(T)}(\omega)$ equals T if $\omega = 0$, but equals zero otherwise. The inner expectation of the four DFTs is broken into a sum over all indecomposable partitions; because the mean of a DFT is zero, we only need to

---

[*]Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, tucker.s.mcelroy@census.gov

[†]University of Maryland, Baltimore County and U.S. Census Bureau

consider three partitions that each involve two pairs. For two of these partitions, we would obtain $\Delta^{(T)} = T$, but for the third partition we obtain $\Delta^{(T)} = 0$; writing the table as $\{\ell\,k\,r\,s\}$, the substantive partitions are $\{(\ell\,k)\,(r\,s)\}$ and $\{(\ell\,s)(k\,r)\}$. Therefore

$$
\mathbb{E}\widehat{Q}_{\varphi_1,\varphi_2}(I,I) = T^{-3} \sum_{j=1}^{T} \sum_{\ell,k,r,s} \varphi_1(\ell,k;\lambda_j)\,\varphi_2(r,s;\lambda_j)\cdot
$$
$$
\left\{ \left(O(1) + T\,\widetilde{f}_{\ell k}^{*}(\lambda_j)\right)\left(O(1) + T\,\widetilde{f}_{rs}^{*}(\lambda_j)\right) \right.
$$
$$
\left. + \left(O(1) + T\,\widetilde{f}_{\ell s}^{*}(\lambda_j)\right)\left(O(1) + T\,\widetilde{f}_{kr}^{*}(\lambda_j)\right)\right\}
$$
$$
\rightarrow \langle [\varphi_1\,\widetilde{f}]\,[\varphi_2\,\widetilde{f}]\rangle_0 + \langle [\varphi_1\,\widetilde{f}\,\varphi_2\,\widetilde{f}]\rangle_0
$$

as $T \rightarrow \infty$, where the final line uses the fact that $\widetilde{f}$ is Hermitian. $\qquad\square$

Proof of Proposition 2. We provide the proof for the hardest case (the third), noting that similar techniques yield the easier two cases. Applying (A.1) twice, we obtain

$$
\mathrm{Cov}\left(\sqrt{T}\,\widehat{Q}_{\varphi_1,\varphi_2}(I,I), \sqrt{T}\,\widehat{Q}_{\varphi_3,\varphi_4}(I,I)\right)
$$
$$
= T^{-5} \sum_{j_1,j_2=1}^{T} \sum_{\ell_1,\ell_2} \sum_{k_1,k_2} \sum_{r_1,r_2} \sum_{s_1,s_2} \varphi_1(\ell_1,k_1;\lambda_{j_1})\,\varphi_2(r_1,s_1;\lambda_{j_1})\,\varphi_3(\ell_2,k_2;\lambda_{j_2})\,\varphi_4(r_2,s_2;\lambda_{j_2})\cdot
$$
$$
\mathrm{cum}\left(d_{\ell_1}^{*}(\lambda_{j_1})\,d_{k_1}(\lambda_{j_1})\,d_{r_1}^{*}(\lambda_{j_1})\,d_{s_1}(\lambda_{j_1}), d_{\ell_2}^{*}(\lambda_{j_2})\,d_{k_2}(\lambda_{j_2})\,d_{r_2}^{*}(\lambda_{j_2})\,d_{s_2}(\lambda_{j_2})\right).
$$

To compute the cumulant we utilize Theorem 2.3.2 of Brillinger (2001), which indicates that we proceed by summing over all indecomposable partitions of the *table* with two rows and four columns, multiplying the cumulants for sets of random variables (DFTs) corresponding to each set of a given partition. Hence, any partitions involving a 1-element set contribute zero, because the cumulant of a single DFT is its mean, which is zero. Which partitions are relevant depends on whether the sum over frequencies collapses to a single summation: if $\lambda_{j_1} = \pm\lambda_{j_2}$, the sum over frequencies collapses to a single summation, and the only partitions we need consider are those involving four sets of size 2 (proved below); otherwise, if $\lambda_{j_1} \neq \pm\lambda_{j_2}$ there is a double summation and the relevant partitions involve one set of size 4 and two sets of size 2 (proved below). In determining which partitions are relevant, we can focus on those indecomposable partitions of the table that yield the highest order in T, all other partitions of lesser order being asymptotically negligible.

Diagonal Case: First suppose that $\lambda_{j_1} = \pm\lambda_{j_2}$. Because no 1-element sets need be considered, the maximal number of sets in a partition is four (which must be four 2-element sets) – and we show that some of these partitions will yield a cumulant $O(T^4)$. Any other type of partition would have fewer than three sets, so that the cumulant would be at most $O(T^3)$, and thus can be ignored. First setting $\lambda_{j_1} = \lambda_{j_2}$, we write the table

$$
d_{\ell_1}^{*}(\lambda_{j_1}) \quad d_{k_1}(\lambda_{j_1}) \quad d_{r_1}^{*}(\lambda_{j_1}) \quad d_{s_1}(\lambda_{j_1})
$$
$$
d_{\ell_2}^{*}(\lambda_{j_1}) \quad d_{k_2}(\lambda_{j_1}) \quad d_{r_2}^{*}(\lambda_{j_1}) \quad d_{s_2}(\lambda_{j_1}).
$$

A four 2-element set partition that is indecomposable must have at least one 2-element set with an element in both of the two rows. There are many of these, but only 20 of them are $O(T^4)$: using (A.2), we only need consider 2-element sets where the sum of the corresponding frequencies is zero, i.e., sets where one element

2

corresponds to a DFT and the other element to a conjugate DFT. We denote these 20 partitions with the following notation: the symbols $\sharp$, $\flat$, $\natural$, and $\star$ will denote membership in a particular 2-element set:

$$
\begin{bmatrix} \sharp & \sharp & \flat & \natural \\ \natural & \flat & \star & \star \end{bmatrix}
\begin{bmatrix} \sharp & \sharp & \flat & \natural \\ \star & \flat & \natural & \star \end{bmatrix}
\begin{bmatrix} \sharp & \sharp & \flat & \natural \\ \natural & \star & \star & \flat \end{bmatrix}
\begin{bmatrix} \sharp & \sharp & \flat & \natural \\ \star & \star & \natural & \flat \end{bmatrix}
$$

$$
\begin{bmatrix} \sharp & \flat & \natural & \sharp \\ \flat & \natural & \star & \star \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \natural & \sharp \\ \flat & \star & \star & \natural \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \natural & \sharp \\ \star & \star & \flat & \natural \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \natural & \sharp \\ \star & \natural & \flat & \star \end{bmatrix}
$$

$$
\begin{bmatrix} \sharp & \flat & \natural & \natural \\ \flat & \sharp & \star & \star \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \natural & \star \\ \flat & \sharp & \star & \natural \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \natural & \star \\ \star & \sharp & \flat & \natural \end{bmatrix}
$$

$$
\begin{bmatrix} \sharp & \flat & \natural & \natural \\ \star & \sharp & \flat & \star \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \flat & \natural \\ \natural & \sharp & \star & \star \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \flat & \natural \\ \star & \sharp & \natural & \star \end{bmatrix}
$$

$$
\begin{bmatrix} \sharp & \flat & \natural & \natural \\ \flat & \star & \star & \sharp \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \natural & \star \\ \flat & \natural & \star & \sharp \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \natural & \star \\ \star & \natural & \flat & \sharp \end{bmatrix}
$$

$$
\begin{bmatrix} \sharp & \flat & \natural & \natural \\ \star & \star & \flat & \sharp \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \flat & \natural \\ \natural & \star & \star & \sharp \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \flat & \natural \\ \star & \star & \natural & \sharp \end{bmatrix}
$$

As a result, the covariance of the quadratic forms has an asymptotic contribution from the diagonal case (with $\lambda_{j_1} = \lambda_{j_2}$) of

$$
T^{-1} \sum_{j=1}^{T} \sum_{\ell_1,\ell_2} \sum_{k_1,k_2} \sum_{r_1,r_2} \sum_{s_1,s_2} \varphi_1(\ell_1,k_1;\lambda_j)\,\varphi_2(r_1,s_1;\lambda_j)\,\varphi_3(\ell_2,k_2;\lambda_j)\,\varphi_4(r_2,s_2;\lambda_j) \cdot
$$

$$
\Big\{ \widetilde{f}^*_{\ell_1 k_1}(\lambda_j)\,\widetilde{f}^*_{r_1 k_2}(\lambda_j)\,\widetilde{f}_{s_1 \ell_2}(\lambda_j)\,\widetilde{f}^*_{r_2 s_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 k_1}(\lambda_j)\,\widetilde{f}^*_{r_1 k_2}(\lambda_j)\,\widetilde{f}_{s_1 r_2}(\lambda_j)\,\widetilde{f}^*_{\ell_2 s_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 k_1}(\lambda_j)\,\widetilde{f}^*_{r_1 s_2}(\lambda_j)\,\widetilde{f}_{s_1 \ell_2}(\lambda_j)\,\widetilde{f}_{k_2 r_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 k_1}(\lambda_j)\,\widetilde{f}^*_{r_1 s_2}(\lambda_j)\,\widetilde{f}_{s_1 r_2}(\lambda_j)\,\widetilde{f}^*_{\ell_2 k_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 s_1}(\lambda_j)\,\widetilde{f}_{k_1 \ell_2}(\lambda_j)\,\widetilde{f}^*_{r_1 k_2}(\lambda_j)\,\widetilde{f}^*_{r_2 s_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 s_1}(\lambda_j)\,\widetilde{f}_{k_1 \ell_2}(\lambda_j)\,\widetilde{f}^*_{r_1 s_2}(\lambda_j)\,\widetilde{f}_{k_2 r_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 s_1}(\lambda_j)\,\widetilde{f}_{k_1 r_2}(\lambda_j)\,\widetilde{f}^*_{r_1 s_2}(\lambda_j)\,\widetilde{f}^*_{\ell_2 k_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 s_1}(\lambda_j)\,\widetilde{f}_{k_1 r_2}(\lambda_j)\,\widetilde{f}^*_{r_1 k_2}(\lambda_j)\,\widetilde{f}^*_{\ell_2 s_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 k_2}(\lambda_j)\,\widetilde{f}_{k_1 \ell_2}(\lambda_j)\,\widetilde{f}^*_{r_1 s_1}(\lambda_j)\,\widetilde{f}^*_{r_2 s_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 k_2}(\lambda_j)\,\widetilde{f}_{k_1 \ell_2}(\lambda_j)\,\widetilde{f}^*_{r_1 s_2}(\lambda_j)\,\widetilde{f}_{s_1 r_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 k_2}(\lambda_j)\,\widetilde{f}_{k_1 r_2}(\lambda_j)\,\widetilde{f}^*_{r_1 s_2}(\lambda_j)\,\widetilde{f}_{s_1 \ell_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 k_2}(\lambda_j)\,\widetilde{f}_{k_1 r_2}(\lambda_j)\,\widetilde{f}^*_{r_1 s_1}(\lambda_j)\,\widetilde{f}^*_{\ell_2 s_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 k_2}(\lambda_j)\,\widetilde{f}_{k_1 r_1}(\lambda_j)\,\widetilde{f}_{s_1 \ell_2}(\lambda_j)\,\widetilde{f}^*_{r_2 s_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 k_2}(\lambda_j)\,\widetilde{f}_{k_1 r_1}(\lambda_j)\,\widetilde{f}_{s_1 r_2}(\lambda_j)\,\widetilde{f}^*_{\ell_2 s_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 s_2}(\lambda_j)\,\widetilde{f}_{k_1 \ell_2}(\lambda_j)\,\widetilde{f}^*_{r_1 s_1}(\lambda_j)\,\widetilde{f}_{k_2 r_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 s_2}(\lambda_j)\,\widetilde{f}_{k_1 \ell_2}(\lambda_j)\,\widetilde{f}^*_{r_1 k_2}(\lambda_j)\,\widetilde{f}_{s_1 r_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 s_2}(\lambda_j)\,\widetilde{f}_{k_1 r_2}(\lambda_j)\,\widetilde{f}^*_{r_1 k_2}(\lambda_j)\,\widetilde{f}_{s_1 \ell_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 s_2}(\lambda_j)\,\widetilde{f}_{k_1 r_2}(\lambda_j)\,\widetilde{f}^*_{r_1 s_1}(\lambda_j)\,\widetilde{f}_{\ell_2 k_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 s_2}(\lambda_j)\,\widetilde{f}_{k_1 r_1}(\lambda_j)\,\widetilde{f}_{s_1 \ell_2}(\lambda_j)\,\widetilde{f}_{k_2 r_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 s_2}(\lambda_j)\,\widetilde{f}_{k_1 r_1}(\lambda_j)\,\widetilde{f}_{s_1 r_2}(\lambda_j)\,\widetilde{f}^*_{\ell_2 k_2}(\lambda_j) \Big\}
$$

This converges to

$$
\langle [1][4][32] \rangle_0 + \langle [1][243] \rangle_0 + \langle [1][234] \rangle_0 + \langle [1][3][24] \rangle_0
$$

$$
+ \langle [4][132] \rangle_0 + \langle [1342] \rangle_0 + \langle [3][142] \rangle_0 + \langle [1432] \rangle_0
$$

$$
+ \langle [2][4][31] \rangle_0 + \langle [31][42] \rangle_0 + \langle [1423] \rangle_0 + \langle [2][143] \rangle_0
$$

$$
+ \langle [4][123] \rangle_0 + \langle [1243] \rangle_0 + \langle [2][134] \rangle_0 + \langle [1324] \rangle_0
$$

$$
+ \langle [14][23] \rangle_0 + \langle [3][2][14] \rangle_0 + \langle [1234] \rangle_0 + \langle [3][124] \rangle_0
$$

3

Next, setting $\lambda_{j_1} = -\lambda_{j_2}$ we write the table

$$d^*_{\ell_1}(\lambda_{j_1}) \quad d_{k_1}(\lambda_{j_1}) \quad d^*_{r_1}(\lambda_{j_1}) \quad d_{s_1}(\lambda_{j_1})$$

$$d_{\ell_2}(\lambda_{j_1}) \quad d^*_{k_2}(\lambda_{j_1}) \quad d_{r_2}(\lambda_{j_1}) \quad d^*_{s_2}(\lambda_{j_1}).$$

Again, there are 20 relevant partitions:

$$
\begin{bmatrix} \sharp & \sharp & \flat & \star \\ \flat & \natural & \natural & \star \end{bmatrix}
\begin{bmatrix} \sharp & \sharp & \flat & \natural \\ \flat & \natural & \star & \star \end{bmatrix}
\begin{bmatrix} \sharp & \sharp & \flat & \natural \\ \star & \star & \flat & \natural \end{bmatrix}
\begin{bmatrix} \sharp & \sharp & \flat & \natural \\ \star & \natural & \flat & \star \end{bmatrix}
$$

$$
\begin{bmatrix} \sharp & \flat & \natural & \sharp \\ \star & \flat & \natural & \star \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \natural & \sharp \\ \natural & \flat & \star & \star \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \natural & \sharp \\ \star & \star & \natural & \flat \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \natural & \sharp \\ \natural & \star & \star & \flat \end{bmatrix}
$$

$$
\begin{bmatrix} \sharp & \flat & \natural & \natural \\ \sharp & \flat & \star & \star \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \natural & \star \\ \sharp & \flat & \natural & \star \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \flat & \star \\ \sharp & \natural & \natural & \star \end{bmatrix}
$$

$$
\begin{bmatrix} \sharp & \flat & \flat & \natural \\ \sharp & \natural & \star & \star \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \star & \natural \\ \sharp & \natural & \star & \flat \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \star & \star \\ \sharp & \natural & \natural & \flat \end{bmatrix}
$$

$$
\begin{bmatrix} \sharp & \flat & \flat & \star \\ \natural & \natural & \sharp & \star \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \flat & \star \\ \natural & \star & \sharp & \natural \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \star & \star \\ \natural & \flat & \sharp & \natural \end{bmatrix}
$$

$$
\begin{bmatrix} \sharp & \flat & \natural & \star \\ \natural & \flat & \sharp & \star \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \star & \star \\ \natural & \natural & \sharp & \flat \end{bmatrix}
\begin{bmatrix} \sharp & \flat & \natural & \star \\ \natural & \star & \sharp & \flat \end{bmatrix}
$$

Hence, the covariance of the quadratic forms has an asymptotic contribution from the diagonal case (with $-\lambda_{j_1} = \lambda_{j_2}$) of

$$
T^{-1} \sum_{j=1}^{T} \sum_{\ell_1,\ell_2} \sum_{k_1,k_2} \sum_{r_1,r_2} \sum_{s_1,s_2} \varphi_1(\ell_1, k_1; \lambda_j)\, \varphi_2(r_1, s_1; \lambda_j)\, \varphi_3(\ell_2, k_2; -\lambda_j)\, \varphi_4(r_2, s_2; -\lambda_j) \cdot
$$

$$
\Big\{ \widetilde{f}^*_{\ell_1 k_1}(\lambda_j)\, \widetilde{f}^*_{r_1 \ell_2}(\lambda_j)\, \widetilde{f}^*_{k_2 r_2}(\lambda_j)\, \widetilde{f}_{s_1 s_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 k_1}(\lambda_j)\, \widetilde{f}^*_{r_1 \ell_2}(\lambda_j)\, \widetilde{f}_{s_1 k_2}(\lambda_j)\, \widetilde{f}_{r_2 s_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 k_1}(\lambda_j)\, \widetilde{f}^*_{r_1 r_2}(\lambda_j)\, \widetilde{f}_{s_1 s_2}(\lambda_j)\, \widetilde{f}_{\ell_2 k_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 k_1}(\lambda_j)\, \widetilde{f}^*_{r_1 r_2}(\lambda_j)\, \widetilde{f}_{s_1 k_2}(\lambda_j)\, \widetilde{f}_{\ell_2 s_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 s_1}(\lambda_j)\, \widetilde{f}_{k_1 k_2}(\lambda_j)\, \widetilde{f}^*_{r_1 r_2}(\lambda_j)\, \widetilde{f}_{\ell_2 s_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 s_1}(\lambda_j)\, \widetilde{f}_{k_1 k_2}(\lambda_j)\, \widetilde{f}^*_{r_1 \ell_2}(\lambda_j)\, \widetilde{f}_{r_2 s_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 s_1}(\lambda_j)\, \widetilde{f}_{k_1 s_2}(\lambda_j)\, \widetilde{f}^*_{r_1 r_2}(\lambda_j)\, \widetilde{f}_{\ell_2 k_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 s_1}(\lambda_j)\, \widetilde{f}_{k_1 s_2}(\lambda_j)\, \widetilde{f}^*_{r_1 \ell_2}(\lambda_j)\, \widetilde{f}^*_{k_2 r_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 \ell_2}(\lambda_j)\, \widetilde{f}_{k_1 k_2}(\lambda_j)\, \widetilde{f}^*_{r_1 s_1}(\lambda_j)\, \widetilde{f}_{r_2 s_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 \ell_2}(\lambda_j)\, \widetilde{f}_{k_1 k_2}(\lambda_j)\, \widetilde{f}^*_{r_1 r_2}(\lambda_j)\, \widetilde{f}_{s_1 s_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 \ell_2}(\lambda_j)\, \widetilde{f}_{k_1 r_1}(\lambda_j)\, \widetilde{f}^*_{k_2 r_2}(\lambda_j)\, \widetilde{f}_{s_1 s_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 \ell_2}(\lambda_j)\, \widetilde{f}_{k_1 r_1}(\lambda_j)\, \widetilde{f}_{s_1 k_2}(\lambda_j)\, \widetilde{f}_{r_2 s_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 \ell_2}(\lambda_j)\, \widetilde{f}_{k_1 s_2}(\lambda_j)\, \widetilde{f}_{s_1 k_2}(\lambda_j)\, \widetilde{f}^*_{r_1 r_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 \ell_2}(\lambda_j)\, \widetilde{f}_{k_1 s_2}(\lambda_j)\, \widetilde{f}^*_{k_2 r_2}(\lambda_j)\, \widetilde{f}^*_{r_1 s_1}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 r_2}(\lambda_j)\, \widetilde{f}_{k_1 r_1}(\lambda_j)\, \widetilde{f}_{\ell_2 k_2}(\lambda_j)\, \widetilde{f}_{s_1 s_2}(\lambda_j) + \widetilde{f}^*_{\ell_1 r_2}(\lambda_j)\, \widetilde{f}_{k_1 r_1}(\lambda_j)\, \widetilde{f}_{\ell_2 s_2}(\lambda_j)\, \widetilde{f}_{s_1 k_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 r_2}(\lambda_j)\, \widetilde{f}_{k_1 k_2}(\lambda_j)\, \widetilde{f}_{\ell_2 s_2}(\lambda_j)\, \widetilde{f}^*_{r_1 s_1}(\lambda_j) + \widetilde{f}^*_{\ell_1 r_2}(\lambda_j)\, \widetilde{f}_{k_1 k_2}(\lambda_j)\, \widetilde{f}^*_{r_1 \ell_2}(\lambda_j)\, \widetilde{f}_{s_1 s_2}(\lambda_j)
$$

$$
+ \widetilde{f}^*_{\ell_1 r_2}(\lambda_j)\, \widetilde{f}_{k_1 s_2}(\lambda_j)\, \widetilde{f}_{\ell_2 k_2}(\lambda_j)\, \widetilde{f}^*_{r_1 s_1}(\lambda_j) + \widetilde{f}^*_{\ell_1 r_2}(\lambda_j)\, \widetilde{f}_{k_1 s_2}(\lambda_j)\, \widetilde{f}^*_{r_1 \ell_2}(\lambda_j)\, \widetilde{f}_{s_1 k_2}(\lambda_j) \Big\}
$$

4

Noting that the argument of φ_3 and φ_4 is $-\lambda_j$, and using the underscore notation, the above converges to

$$\langle[1][2\underline{43}]\rangle_0 + \langle[\underline{4}][1][2\underline{3}]\rangle_0 + \langle[1][\underline{3}][2\underline{4}]\rangle_0 + \langle[1][2\underline{34}]\rangle_0$$
$$+ \langle[1\underline{342}]\rangle_0 + \langle[1\underline{32}][\underline{4}]\rangle_0 + \langle[1\underline{42}][\underline{3}]\rangle_0 + \langle[1\underline{432}]\rangle_0$$
$$+ \langle[\underline{4}][2][13]\rangle_0 + \langle[13][2\underline{4}]\rangle_0 + \langle[12\underline{43}]\rangle_0 + \langle[12\underline{3}][\underline{4}]\rangle_0$$
$$+ \langle[1\underline{423}]\rangle_0 + \langle[1\underline{43}][2]\rangle_0 + \langle[12\underline{4}][\underline{3}]\rangle_0 + \langle[12\underline{34}]\rangle_0$$
$$+ \langle[13\underline{4}]\rangle_0 + \langle[1\underline{324}]\rangle_0 + \langle[1\underline{4}][2][\underline{3}]\rangle_0 + \langle[1\underline{4}][2\underline{3}]\rangle_0.$$

This accounts for the entire contribution from the diagonal case.

Off-diagonal Case: Now we suppose that $\lambda_{j_1} \neq \lambda_{j_2}$, and hence the double sum does not collapse to a single summation. In this case, of the partitions with four 2-element sets they are all either decomposable or only contribute terms of order $O(T^2)$. This is because such a partition that is indecomposable must have two sets with an element in each row – see examples in the prior case. (The definition of indecomposable requires one set to have an element in each row, but as this would only leave three free slots in each row, we must have at least one other set with this property.) But because $\lambda_{j_1} \neq \lambda_{j_2}$, (A.2) ensures that the contribution to the cumulant from such sets is $O(1)$, indicating that the largest possible order from such a partition is $O(T^2)$. Such terms can be ignored, because there exist indecomposable partitions yielding $O(T^3)$ terms in the cumulants: these partitions involve one 4-element set and two 2-element sets. Moreover, no other partitions need be considered: the only other partitions (that don't involve 1-element sets) with three sets would have two 3-element sets and one 2-element set, but the cumulant of a 3-element set will never be $O(T)$, according to equation (4.3.15) of Brillinger (2001). This expression states that the m-fold cumulant of m DFTs can be $O(T)$ so long as the sum of the frequency arguments is zero; there is no way this can happen when the frequencies take the form $\lambda_{j_1}, -\lambda_{j_1}, \lambda_{j_2}, -\lambda_{j_2}$. The expression of equation (4.3.15) of Brillinger (2001) in the case $m = 4$ is

$$\mathrm{cum}(d_\ell(\lambda_{j_1}), d_k(\lambda_{j_2}), d_r(\lambda_{j_3}), d_s(\lambda_{j_4})) = O(1) + \Delta^{(T)}\left(\sum_{i=1}^{4} \lambda_{j_i}\right) \widetilde{f}_{\ell krs}(\lambda_{j_1}, \lambda_{j_2}, \lambda_{j_3}). \qquad (A.3)$$

Using the table

$$d_{\ell_1}^*(\lambda_{j_1}) \quad d_{k_1}(\lambda_{j_1}) \quad d_{r_1}^*(\lambda_{j_1}) \quad d_{s_1}(\lambda_{j_1})$$
$$d_{\ell_2}^*(\lambda_{j_2}) \quad d_{k_2}(\lambda_{j_2}) \quad d_{r_2}^*(\lambda_{j_2}) \quad d_{s_2}(\lambda_{j_2}),$$

we find that there are 16 indecomposable partitions consisting of one 4-element and two 2-element sets, such that the contribution is $O(T^3)$. The four elements of the 4-element set must be allocated with two elements in each row, because if three belong to a single row it is impossible for the sum of all frequencies to equal zero, as required in (A.3). (Also, if all four elements belonged to a single row, the partition would be decomposable.) Also, once the 4-element set is allocated with two members in each row, each of the remaining 2-element sets must be contained in a single row (otherwise the sum of frequencies cannot equal

5

zero, and the contribution will be less than $O(T^3)$). There are 16 such partitions, which we list below:

$$\begin{bmatrix} \sharp & \sharp & \flat & \flat \\ \sharp & \sharp & \natural & \natural \end{bmatrix} \quad \begin{bmatrix} \sharp & \sharp & \flat & \flat \\ \sharp & \natural & \natural & \sharp \end{bmatrix} \quad \begin{bmatrix} \sharp & \sharp & \flat & \flat \\ \natural & \sharp & \sharp & \natural \end{bmatrix} \quad \begin{bmatrix} \sharp & \sharp & \flat & \flat \\ \natural & \natural & \sharp & \sharp \end{bmatrix}$$

$$\begin{bmatrix} \sharp & \flat & \flat & \sharp \\ \sharp & \sharp & \natural & \natural \end{bmatrix} \quad \begin{bmatrix} \sharp & \flat & \flat & \sharp \\ \sharp & \natural & \natural & \sharp \end{bmatrix} \quad \begin{bmatrix} \sharp & \flat & \flat & \sharp \\ \natural & \sharp & \sharp & \natural \end{bmatrix} \quad \begin{bmatrix} \sharp & \flat & \flat & \sharp \\ \natural & \natural & \sharp & \sharp \end{bmatrix}$$

$$\begin{bmatrix} \flat & \sharp & \sharp & \flat \\ \sharp & \sharp & \natural & \natural \end{bmatrix} \quad \begin{bmatrix} \flat & \sharp & \sharp & \flat \\ \sharp & \natural & \natural & \sharp \end{bmatrix} \quad \begin{bmatrix} \flat & \sharp & \sharp & \flat \\ \natural & \sharp & \sharp & \natural \end{bmatrix} \quad \begin{bmatrix} \flat & \sharp & \sharp & \flat \\ \natural & \natural & \sharp & \sharp \end{bmatrix}$$

$$\begin{bmatrix} \flat & \flat & \sharp & \sharp \\ \sharp & \sharp & \natural & \natural \end{bmatrix} \quad \begin{bmatrix} \flat & \flat & \sharp & \sharp \\ \sharp & \natural & \natural & \sharp \end{bmatrix} \quad \begin{bmatrix} \flat & \flat & \sharp & \sharp \\ \natural & \sharp & \sharp & \natural \end{bmatrix} \quad \begin{bmatrix} \flat & \flat & \sharp & \sharp \\ \natural & \natural & \sharp & \sharp \end{bmatrix}$$

The contribution to the covariance is therefore

$$T^{-2} \sum_{j_1 \neq j_2 = 1}^{T} \sum_{\ell_1, \ell_2} \sum_{k_1, k_2} \sum_{r_1, r_2} \sum_{s_1, s_2} \varphi_1(\ell_1, k_1; \lambda_{j_1}) \, \varphi_2(r_1, s_1; \lambda_{j_1}) \, \varphi_3(\ell_2, k_2; \lambda_{j_2}) \, \varphi_4(r_2, s_2; \lambda_{j_2}) \cdot$$

$$\left\{ \widetilde{f}_{\ell_1 k_1 \ell_2 k_2}(-\lambda_{j_1}, \lambda_{j_1}, -\lambda_{j_2}) \, \widetilde{f}^*_{r_1 s_1}(\lambda_{j_1}) \, \widetilde{f}^*_{r_2 s_2}(\lambda_{j_2}) + \widetilde{f}_{\ell_1 k_1 \ell_2 s_2}(-\lambda_{j_1}, \lambda_{j_1}, -\lambda_{j_2}) \, \widetilde{f}^*_{r_1 s_1}(\lambda_{j_1}) \, \widetilde{f}^*_{k_2 r_2}(\lambda_{j_2}) \right.$$

$$+ \widetilde{f}_{\ell_1 k_1 k_2 r_2}(-\lambda_{j_1}, \lambda_{j_1}, \lambda_{j_2}) \, \widetilde{f}^*_{r_1 s_1}(\lambda_{j_1}) \, \widetilde{f}^*_{\ell_2 s_2}(\lambda_{j_2}) + \widetilde{f}_{\ell_1 k_1 r_2 s_2}(-\lambda_{j_1}, \lambda_{j_1}, -\lambda_{j_2}) \, \widetilde{f}^*_{r_1 s_1}(\lambda_{j_1}) \, \widetilde{f}^*_{\ell_2 k_2}(\lambda_{j_2})$$

$$+ \widetilde{f}_{\ell_1 s_1 \ell_2 k_2}(-\lambda_{j_1}, \lambda_{j_1}, -\lambda_{j_2}) \, \widetilde{f}_{k_1 r_1}(\lambda_{j_1}) \, \widetilde{f}^*_{r_2 s_2}(\lambda_{j_2}) + \widetilde{f}_{\ell_1 s_1 \ell_2 s_2}(-\lambda_{j_1}, \lambda_{j_1}, -\lambda_{j_2}) \, \widetilde{f}_{k_1 r_1}(\lambda_{j_1}) \, \widetilde{f}_{k_2 r_2}(\lambda_{j_2})$$

$$+ \widetilde{f}_{\ell_1 s_1 k_2 r_2}(-\lambda_{j_1}, \lambda_{j_1}, \lambda_{j_2}) \, \widetilde{f}_{k_1 r_1}(\lambda_{j_1}) \, \widetilde{f}^*_{\ell_2 s_2}(\lambda_{j_2}) + \widetilde{f}_{\ell_1 s_1 r_2 s_2}(-\lambda_{j_1}, \lambda_{j_1}, -\lambda_{j_2}) \, \widetilde{f}_{k_1 r_1}(\lambda_{j_1}) \, \widetilde{f}^*_{\ell_2 k_2}(\lambda_{j_2})$$

$$+ \widetilde{f}_{k_1 r_1 \ell_2 k_2}(\lambda_{j_1}, -\lambda_{j_1}, -\lambda_{j_2}) \, \widetilde{f}^*_{\ell_1 s_1}(\lambda_{j_1}) \, \widetilde{f}^*_{r_2 s_2}(\lambda_{j_2}) + \widetilde{f}_{k_1 r_1 \ell_2 s_2}(\lambda_{j_1}, -\lambda_{j_1}, -\lambda_{j_2}) \, \widetilde{f}^*_{\ell_1 s_1}(\lambda_{j_1}) \, \widetilde{f}_{k_2 r_2}(\lambda_{j_2})$$

$$+ \widetilde{f}_{k_1 r_1 k_2 r_2}(\lambda_{j_1}, -\lambda_{j_1}, \lambda_{j_2}) \, \widetilde{f}^*_{\ell_1 s_1}(\lambda_{j_1}) \, \widetilde{f}^*_{\ell_2 s_2}(\lambda_{j_2}) + \widetilde{f}_{k_1 r_1 r_2 s_2}(\lambda_{j_1}, -\lambda_{j_1}, -\lambda_{j_2}) \, \widetilde{f}^*_{\ell_1 s_1}(\lambda_{j_1}) \, \widetilde{f}^*_{\ell_2 k_2}(\lambda_{j_2})$$

$$+ \widetilde{f}_{r_1 s_1 \ell_2 k_2}(-\lambda_{j_1}, \lambda_{j_1}, -\lambda_{j_2}) \, \widetilde{f}^*_{\ell_1 k_1}(\lambda_{j_1}) \, \widetilde{f}^*_{r_2 s_2}(\lambda_{j_2}) + \widetilde{f}_{r_1 s_1 \ell_2 s_2}(-\lambda_{j_1}, \lambda_{j_1}, -\lambda_{j_2}) \, \widetilde{f}^*_{\ell_1 k_1}(\lambda_{j_1}) \, \widetilde{f}_{k_2 r_2}(\lambda_{j_2})$$

$$\left. + \widetilde{f}_{r_1 s_1 k_2 r_2}(-\lambda_{j_1}, \lambda_{j_1}, \lambda_{j_2}) \, \widetilde{f}^*_{\ell_1 k_1}(\lambda_{j_1}) \, \widetilde{f}^*_{\ell_2 s_2}(\lambda_{j_2}) + \widetilde{f}_{r_1 s_1 r_2 s_2}(-\lambda_{j_1}, \lambda_{j_1}, -\lambda_{j_2}) \, \widetilde{f}^*_{\ell_1 k_1}(\lambda_{j_1}) \, \widetilde{f}^*_{\ell_2 k_2}(\lambda_{j_2}) \right\}.$$

The tri-spectral density has the properties that $\widetilde{f}_{\ell ksr}(\lambda, -\lambda, \omega) = \widetilde{f}_{\ell krs}(\lambda, -\lambda, -\omega)$, $\widetilde{f}_{\ell ksr}(\lambda, -\lambda, \omega) = \widetilde{f}_{k\ell sr}(-\lambda, \lambda, \omega)$, and $\widetilde{f}_{\ell ksr}(\lambda, -\lambda, \omega) = \widetilde{f}_{sr\ell k}(\omega, -\omega, \lambda)$. These are verified as follows:

$$\widetilde{f}_{\ell ksr}(\lambda, -\lambda, \omega) = \sum_{h_1, h_2, h_3} \gamma_{\ell ksr}(h_1, h_2, h_3) \, e^{-i\lambda(h_1 - h_2) - i\omega h_3}$$

$$= \sum_{h_1, h_2, h_3} \gamma_{\ell krs}(h_1 - h_3, h_2 - h_3, -h_3) \, e^{-i\lambda(h_1 - h_2) - i\omega h_3}$$

$$= \sum_{k_1, k_2, k_3} \gamma_{\ell krs}(k_1, k_2, k_3) \, e^{-i\lambda(k_1 - k_2) + i\omega k_3} = \widetilde{f}_{\ell krs}(\lambda, -\lambda, -\omega)$$

$$\widetilde{f}_{\ell ksr}(\lambda, -\lambda, \omega) = \sum_{h_1, h_2, h_3} \gamma_{k\ell sr}(h_2, h_1, h_3) \, e^{-i\lambda(h_1 - h_2) - i\omega h_3}$$

$$= \sum_{k_1, k_2, k_3} \gamma_{k\ell sr}(k_1, k_2, k_3) \, e^{i\lambda(k_1 - k_2) - i\omega k_3} = \widetilde{f}_{k\ell sr}(-\lambda, \lambda, \omega)$$

$$\widetilde{f}_{\ell ksr}(\lambda, -\lambda, \omega) = \sum_{h_1, h_2, h_3} \gamma_{\ell ksr}(h_1, h_2, h_3) \, e^{-i\lambda(h_1 - h_2) - i\omega h_3}$$

$$= \sum_{h_1, h_2, h_3} \gamma_{sr\ell k}(h_3 - h_2, -h_2, h_1 - h_2) \, e^{-i\lambda(h_1 - h_2) - i\omega h_3}$$

$$= \sum_{k_1, k_2, k_3} \gamma_{sr\ell k}(k_1, k_2, k_3) \, e^{i\lambda k_3 - i\omega(k_1 - k_2)} = \widetilde{f}_{sr\ell k}(\omega, -\omega, \lambda).$$

6

In the first calculation, this uses

$$\gamma_{\ell k s r}(h_1, h_2, h_3) = \text{cum}\{x_{h_1,\ell}, x_{h_2,k}, x_{h_3,s}, x_{0,r}\}$$

$$= \text{cum}\{x_{h_1-h_3,\ell}, x_{h_2-h_3,k}, x_{-h_3,r}, x_{0,s}\} = \gamma_{\ell k r s}(h_1 - h_3, h_2 - h_3, -h_3)$$

and the change of variable $k_1 = h_1 - h_3$, $k_2 = h_2 - h_3$, and $k_3 = -h_3$. The second calculation uses the fact that cumulant arguments can be permuted, and the change of variable $k_1 = h_2$, $k_2 = h_1$, and $k_3 = h_3$. The third calculation uses

$$\gamma_{\ell k s r}(h_1, h_2, h_3) = \text{cum}\{x_{h_1,\ell}, x_{h_2,k}, x_{h_3,s}, x_{0,r}\}$$

$$= \text{cum}\{x_{h_3-h_2,s}, x_{-h_2,r}, x_{h_1-h_2,\ell}, x_{0,k}\} = \gamma_{s r \ell k}(h_3 - h_2, -h_2, h_1 - h_2)$$

and the change of variable $k_1 = h_3 - h_2$, $k_2 = -h_2$, and $k_3 = h_1 - h_2$. We can use these properties to express every occurence of the tri-spectral density in terms of the frequency arguments $(\lambda_{j_1}, -\lambda_{j_1}, \lambda_{j_2})$ by rearranging the subscript indices. Then it is possible to greatly simplify the summations, letting $T \to \infty$. (The double sum over frequencies will tend to a double integral, not-withstanding the omission of the diagonal portion, which has measure zero.) Then the limiting contribution to the covariance (where curly braces denote a matrix argument to the double bracket) simplifies to the double integral (weighted by $(2\pi)^{-2}$) over λ and ω of

$$
\begin{aligned}
g_{1234}(\lambda, \omega) = &\; [[\varphi_1(\lambda)\,\widetilde{f}(\lambda, -\lambda, \omega)\,\varphi_3(\omega)]]\,[\varphi_2(\lambda)\,\widetilde{f}(\lambda)]\,[\varphi_4(\omega)\,\widetilde{f}(\omega)] \qquad\qquad\text{(A.4)}\\
&+ [[\varphi_1(\lambda)\,\widetilde{f}(\lambda, -\lambda, \omega)\,\{\varphi_3(\omega)\,\widetilde{f}(\omega)\,\varphi_4(\omega)\}]]\,[\varphi_2(\lambda)\,\widetilde{f}(\lambda)]\\
&+ [[\varphi_1(\lambda)\,\widetilde{f}(\lambda, -\lambda, \omega)\,\{\varphi_4(\omega)\,\widetilde{f}(\omega)\,\varphi_3(\omega)\}]]\,[\varphi_2(\lambda)\,\widetilde{f}(\lambda)]\\
&+ [[\varphi_1(\lambda)\,\widetilde{f}(\lambda, -\lambda, \omega)\,\varphi_4(\omega)]]\,[\varphi_2(\lambda)\,\widetilde{f}(\lambda)]\,[\varphi_3(\omega)\,\widetilde{f}(\omega)]\\
&+ [[\{\varphi_1(\lambda)\,\widetilde{f}(\lambda)\,\varphi_2(\lambda)\}\,\widetilde{f}(\lambda, -\lambda, \omega)\,\varphi_3(\omega)]]\,[\varphi_4(\omega)\,\widetilde{f}(\omega)]\\
&+ [[\{\varphi_1(\lambda)\,\widetilde{f}(\lambda)\,\varphi_2(\lambda)\}\,\widetilde{f}(\lambda, -\lambda, \omega)\,\{\varphi_3(\omega)\,\widetilde{f}(\omega)\,\varphi_4(\omega)\}]]\\
&+ [[\{\varphi_1(\lambda)\,\widetilde{f}(\lambda)\,\varphi_2(\lambda)\}\,\widetilde{f}(\lambda, -\lambda, \omega)\,\{\varphi_4(\omega)\,\widetilde{f}(\omega)\,\varphi_3(\omega)\}]]\\
&+ [[\{\varphi_1(\lambda)\,\widetilde{f}(\lambda)\,\varphi_2(\lambda)\}\,\widetilde{f}(\lambda, -\lambda, \omega)\,\varphi_4(\omega)]]\,[\varphi_3(\omega)\,\widetilde{f}(\omega)]\\
&+ [[\{\varphi_2(\lambda)\,\widetilde{f}(\lambda)\,\varphi_1(\lambda)\}\,\widetilde{f}(\lambda, -\lambda, \omega)\,\varphi_3(\omega)]]\,[\varphi_4(\omega)\,\widetilde{f}(\omega)]\\
&+ [[\{\varphi_2(\lambda)\,\widetilde{f}(\lambda)\,\varphi_1(\lambda)\}\,\widetilde{f}(\lambda, -\lambda, \omega)\,\{\varphi_3(\omega)\,\widetilde{f}(\omega)\,\varphi_4(\omega)\}]]\\
&+ [[\{\varphi_2(\lambda)\,\widetilde{f}(\lambda)\,\varphi_1(\lambda)\}\,\widetilde{f}(\lambda, -\lambda, \omega)\,\{\varphi_4(\omega)\,\widetilde{f}(\omega)\,\varphi_3(\omega)\}]]\\
&+ [[\{\varphi_2(\lambda)\,\widetilde{f}(\lambda)\,\varphi_1(\lambda)\}\,\widetilde{f}(\lambda, -\lambda, \omega)\,\varphi_4(\omega)]]\,[\varphi_3(\omega)\,\widetilde{f}(\omega)]\\
&+ [[\varphi_2(\lambda)\,\widetilde{f}(\lambda, -\lambda, \omega)\,\varphi_3(\omega)]]\,[\varphi_1(\lambda)\,\widetilde{f}(\lambda)]\,[\varphi_4(\omega)\,\widetilde{f}(\omega)]\\
&+ [[\varphi_2(\lambda)\,\widetilde{f}(\lambda, -\lambda, \omega)\,\{\varphi_3(\omega)\,\widetilde{f}(\omega)\,\varphi_4(\omega)\}]]\,[\varphi_1(\lambda)\,\widetilde{f}(\lambda)]\\
&+ [[\varphi_2(\lambda)\,\widetilde{f}(\lambda, -\lambda, \omega)\,\{\varphi_4(\omega)\,\widetilde{f}(\omega)\,\varphi_3(\omega)\}]]\,[\varphi_1(\lambda)\,\widetilde{f}(\lambda)]\\
&+ [[\varphi_2(\lambda)\,\widetilde{f}(\lambda, -\lambda, \omega)\,\varphi_4(\omega)]]\,[\varphi_1(\lambda)\,\widetilde{f}(\lambda)]\,[\varphi_3(\omega)\,\widetilde{f}(\omega)].
\end{aligned}
$$

Similar calculations for the linear and mixed quadratic-linear cases yield terms involving the tri-spectrum as

well, and

$$
\begin{aligned}
g_{123}(\lambda, \omega) = &\ [[\varphi_1(\lambda)\, \widetilde{f}(\lambda, -\lambda, \omega)\, \varphi_3(\omega)]]\, [\varphi_2(\lambda)\, \widetilde{f}(\lambda)] && \text{(A.5)} \\
&+ [[\{\varphi_1(\lambda)\, \widetilde{f}(\lambda)\, \varphi_2(\lambda)\}\, \widetilde{f}(\lambda, -\lambda, \omega)\, \varphi_3(\omega)]] \\
&+ [[\{\varphi_2(\lambda)\, \widetilde{f}(\lambda)\, \varphi_1(\lambda)\}\, \widetilde{f}(\lambda, -\lambda, \omega)\, \varphi_3(\omega)]] \\
&+ [[\varphi_2(\lambda)\, \widetilde{f}(\lambda, -\lambda, \omega)\, \varphi_3(\omega)]]\, [\varphi_1(\lambda)\, \widetilde{f}(\lambda)].
\end{aligned}
$$

$\square$

Proof of Theorem 1. We apply the method of cumulants to the functionals. The key result that is needed is that higher-order cumulants of the linear and quadratic functionals tend to zero. We demonstrate this through quadratic functionals only, the other cases being similar. Suppose we have an m-fold cumulant of normalized quadratic functionals, each of which takes the form $\sqrt{T}\, \widehat{Q}_{\varphi,\psi}(I, I)$. Generalizing the arguments of the proof of Proposition 2, we obtain a factor of order of T^{-2m} from the $2m$ periodograms, T^{-m} from the discretization of the m integrals, and $T^{m/2}$ from the normalizations, for an overall $T^{3m-m/2}$ in the denominator. For the cumulant, we must now consider a table with m rows and 4 columns. We claim that the largest possible order of the sums (over all frequencies) of such cumulants is $O(T^{2m+1})$. In the diagonal case, where the m sums really collapse to a single sum – in the manner discussed in the proof of Proposition 2 – one obtains the highest order possible for the cumulant by taking an indecomposable partition with $2m$ sets of size 2. By (A.2), this would yield $O(T^{2m})$, which together with the single summation gives an overall $O(T^{2m+1})$. However, of the m summations we might allow some pairs to collapse to a single summation, and others may not.

Suppose we consider a pair of summations to be distinct, but all the others collapse to a single summation. Now, the partition involving $2m$ sets of size 2 will be of lower order, as any 2-element sets straddling distinct rows corresponding to the two frequencies of the paired summation will no longer satisfy $\lambda_{j_1} + \lambda_{j_2} = 0$; there is at least one such 2-element set (because the partition is indecomposable), so the cumulant order drops to T^{2m-1}. Moreover, by combining two 2-element sets into a 4-element set, we can obtain a factor of T if (A.3) is satisfied, although there will be at most $2m - 1$ sets in such a partition – ultimately yielding $O(T^{2m-1})$. Because the total number of summations is two, we would obtain an overall $O(T^{2m+1})$ for this case.

Proceeding by the same argument, distinct summations over frequencies add an overall order of T but also limit the types of partitions that will yield cumulant terms of order T; we can always merge two 2-element sets into a 4-element set when determining the relevant partitions, in moving from a collapsed summation to a double summation – but this will decrease the size of the partition by one. This is compensated by having an additional summation – so the largest possible order is T^{2m+1}. Pairing this with the denominator $T^{3m-m/2}$, the m-fold cumulant is $O(T^{1-m/2})$, which tends to zero when $m > 2$.

Clearly, if we are talking about the m-fold cumulants of the same quadratic functional, this establishes that it is asymptotically normal. But because the discussion pertains to m-fold cumulants of any collection of quadratic functionals, joint asymptotic normality also follows from the Cramer-Wold device. Extending these arguments to joint relations with linear functionals completes the proof. $\square$

Proof of Proposition 3. The convergence in probability follows from Propositions 1 and 2 (cf. Remark 1), applied to $|I - f_\theta|^2$:

$$
|I - f_\theta|^2 = \langle [I^2] \rangle_0 - 2\, \langle [I\, f_\theta] \rangle_0 + |f_\theta|^2 \xrightarrow{P} \langle [\widetilde{f}^2] \rangle_0 + \langle [\widetilde{f}]^2 \rangle_0 - 2\, \langle [\widetilde{f}\, f_\theta] \rangle_0 + |f_\theta|^2.
$$

8

$\square$

Theorem 2. Write $\widehat{F}(\theta)$ for $\mathrm{FD}(f_\theta, I)$, and $\widetilde{F}(\theta)$ for $\mathrm{FD}(f_\theta, \widetilde{f})$. First,

$$0 = \nabla\widehat{F}(\widehat{\theta}) = \nabla\widehat{F}(\widetilde{\theta}) + \nabla\nabla'\widehat{F}(\widetilde{\theta})\,(\widehat{\theta} - \widetilde{\theta}) + R_T$$

by a Taylor series expansion, where R_T depends on $\widehat{\theta} - \widetilde{\theta}$ quadratically. We proceed to compute the gradient and Hessian of $\widehat{F}(\theta)$:

$$\partial\theta_j\widehat{F}(\theta) = -2\left[\langle\partial\theta_j f_\theta\,(I - f_\theta)\rangle_0\right]$$
$$\partial\theta_j\partial\theta_k\widehat{F}(\theta) = -2\left[\langle\partial\theta_j\partial\theta_k f_\theta\,(I - f_\theta)\rangle_0\right] + 2\left[\langle\partial\theta_j f_\theta\,\partial\theta_k f_\theta\rangle_0\right],$$

and the first term has asymptotic mean

$$-2\left[\langle\partial\theta_j f_\theta\,(\widetilde{f} - f_\theta)\rangle_0\right].$$

This we recognize as the derivative of $\widetilde{F}(\theta)$, which is zero at $\theta = \widetilde{\theta}$. Therefore

$$\nabla\widehat{F}(\widetilde{\theta}) = -2\left[\langle\nabla f_\theta\,(I - \widetilde{f})\rangle_0\right],$$

where the trace operator does not act on the gradient. An application of Theorem 1, in conjunction with our other assumptions yields

$$\sqrt{T}\,\nabla\widehat{F}(\widetilde{\theta}) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, 4\,V_{\widetilde{\theta}}). \tag{A.6}$$

Next, the Hessian of $\widehat{F}(\theta)$ converges in probability to $2\,M_\theta$, because the first term is actually $O_P(T^{-1/2})$ (by Lemma 3.1.1 of Taniguchi and Kakizawa (2000)). If this matrix is invertible at $\widetilde{\theta}$, we conclude that $\widehat{\theta} - \widetilde{\theta} = O_P(T^{-1/2})$ and further that

$$\widehat{\theta} - \widetilde{\theta} = .5\,M_{\widetilde{\theta}}^{-1}\,\langle[\nabla f_\theta\,(I - \widetilde{f})]\rangle_0 + o_P(T^{-1/2}).$$

So using (A.6), the theorem is proved. $\square$

Proof of Proposition 4. We can write the criterion function as

$$|I - f_\theta|^2 = \sum_{i,k=1}^{K} G_{ik}[\Theta_i\,\Theta_k'] - 2\sum_{i=1}^{K}[\Theta_i\,\langle Ig_i\rangle_0] + \langle[I^2]\rangle_0$$
$$= [\Theta\,(G \otimes 1_m)\,\Theta'] - 2[\Theta\,\langle g \otimes I\rangle_0] + \langle[I^2]\rangle_0.$$

Computing the gradient, we now see that the stated formula (2) for the MOM estimator is a critical point, and a minimizer, of this criterion. The same proof, with $\widetilde{f}$ in place of I, shows that the formula for the PTV is obtained by replacing $\widetilde{f}$ for I in (2). $\square$

Lemma 1. *For a possibly non-square matrix A,*

$$\sqrt{T}\,\langle A\,(vec(I - f_{\widetilde{\theta}}))\rangle_0 \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, 2\,\langle A\,(f_{\widetilde{\theta}} \otimes f_{\widetilde{\theta}}')\,A'\rangle_0\right).$$

Proof of Lemma 1. To prove the Lemma, for each ℓ let α_ℓ be the matrix such that $\mathrm{vec}(\alpha'_\ell) = \{e_\ell A\}'$. Then for any f

$$A \, \mathrm{vec}(f) = \left\{ \begin{array}{c} e'_1 A \, \mathrm{vec}(f) \\ e'_2 A \, \mathrm{vec}(f) \\ \vdots \end{array} \right\} = \left\{ \begin{array}{c} \mathrm{vec}(\alpha'_1)' \mathrm{vec}(f) \\ \mathrm{vec}(\alpha'_2)' \mathrm{vec}(f) \\ \vdots \end{array} \right\} = \left\{ \begin{array}{c} [\alpha_1 f] \\ [\alpha_2 f] \\ \vdots \end{array} \right\}.$$

Now by Lemma 3.1.1 of Taniguchi and Kakizawa (2000), we have the joint CLT

$$\sqrt{T} \, \langle A \, (\mathrm{vec}(I - f_{\widehat{\theta}})) \rangle_0 = \sqrt{T} \left\{ \begin{array}{c} \left[\alpha_1 |I - f_{\widehat{\theta}}|^2\right] \\ \left[\alpha_2 |I - f_{\widehat{\theta}}|^2\right] \\ \vdots \end{array} \right\} \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, 2\{\langle [\alpha_j f_{\widehat{\theta}} \alpha'_k f_{\widehat{\theta}}]\rangle_0\}_{j,k}\right).$$

It can be shown using algebraic identities that

$$[\alpha_j f_{\widehat{\theta}} \alpha'_k f_{\widehat{\theta}}] = \mathrm{vec}(\alpha'_j)' \, \mathrm{vec}(f_{\widehat{\theta}} \alpha'_k f_{\widehat{\theta}}) = e'_j A \, f_{\widehat{\theta}} \otimes f'_{\widehat{\theta}} \mathrm{vec}(\alpha'_k) = e'_j A \, f_{\widehat{\theta}} \otimes f'_{\widehat{\theta}} A' e_k,$$

and hence that $\langle [\alpha_j f_{\widehat{\theta}} \alpha'_k f_{\widehat{\theta}}]\rangle_0$ is the jkth entry of $A \, \langle f_{\widehat{\theta}} \otimes f'_{\widehat{\theta}} \rangle_0 A'$. The result follows. Note that the transpose on the second appearance of $f_{\widehat{\theta}}$ guarantees that the limiting covariance matrix is symmetric. $\square$

Proof of Theorem 3. The result follows from Lemma 1 upon writing

$$\theta = \mathrm{vec} \, \Theta = \left\{ \begin{array}{c} \langle (g'G^{-1}e_1) \, \mathrm{vec}(f) \rangle_0 \\ \langle (g'G^{-1}e_2) \, \mathrm{vec}(f) \rangle_0 \\ \vdots \end{array} \right\} = [\langle A \, \mathrm{vec}(f) \rangle_0\},$$

where $A = G^{-1}g \otimes 1_{m^2}$. $\square$

Proof of Proposition 5. We claim that

$$\mathbb{E}|J|^2 = O(T^{-1}) + |\widetilde{g}|^2 + \langle [\widetilde{g}]^2 \rangle_0, \tag{A.7}$$

by Proposition 1, and also

$$|J|^2 = o_P(1) + |\widetilde{g}|^2 + \langle [\widetilde{g}]^2 \rangle_0$$

follows from both Propositions 1 and 2. Next, we determine the mean of $[\langle J \rangle_0^2]$, using the techniques of the proof of Proposition 1:

$$\mathbb{E}[\langle J \rangle_0^2] = \mathbb{E}[\langle T^{-1}d \, d^* \rangle_0^2] = T^{-4} \sum_{j_1, j_2} \sum_{\ell, r} \mathbb{E}[d^*_\ell(\lambda_{j_1}) \, d_\ell(\lambda_{j_2}) \, d^*_r(\lambda_{j_2}) \, d_r(\lambda_{j_1})],$$

and the outer summations are broken into terms where $j_1 = j_2$, $j_1 = -j_2$, or $j_1 \neq \pm j_2$. Utilizing the arguments of Proposition 2, the first two such terms contribute $O(T^{-1})$, and only the off-diagonal term is dominant, yielding

$$\mathbb{E}[\langle J \rangle_0^2] = O(T^{-1}) + T^{-4} \sum_{j_1 \neq \pm j_2} \sum_{\ell, r} (O(1) + T \, f^*_{\ell r}(\lambda_{j_1})) \, (O(1) + T \, f_{\ell r}(\lambda_{j_2})) = O(T^{-1}) + [\langle \widetilde{g} \rangle_0^2].$$

Furthermore, $\mathrm{vec}\langle J \rangle_0 \overset{P}{\longrightarrow} \mathrm{vec}\langle \widetilde{g} \rangle_0$ by Lemma 3.1.1 of Taniguchi and Kakizawa (2000), and hence $[\langle J \rangle_0^2] \overset{P}{\longrightarrow} [\langle \widetilde{g} \rangle_0^2]$, using

$$[\langle J \rangle_0^2] = \mathrm{vec}\langle J' \rangle_0' \, \mathrm{vec}\langle J \rangle_0.$$

Next, we determine the second moment of $[\langle J \rangle_0]$:

$$\mathbb{E}[\langle J \rangle_0]^2 = \mathbb{E}\langle T^{-1} d^* d \rangle_0^2 = T^{-4} \sum_{j_1, j_2} \sum_{\ell, r} \mathbb{E}[d_\ell^*(\lambda_{j_1}) \, d_\ell(\lambda_{j_1}) \, d_r^*(\lambda_{j_2}) \, d_r(\lambda_{j_2})],$$

and the outer summations are broken into terms where $j_1 = j_2$, $j_1 = -j_2$, or $j_1 \neq \pm j_2$. Again, the first two such terms contribute $O(T^{-1})$, and only the off-diagonal term is dominant, yielding

$$\mathbb{E}[\langle J \rangle_0]^2 = O(T^{-1}) + T^{-4} \sum_{j_1 \neq \pm j_2} \sum_{\ell, r} (O(1) + T \, f_{\ell\ell}^*(\lambda_{j_1})) \, (O(1) + T \, f_{rr}^*(\lambda_{j_2})) = O(T^{-1}) + [\langle \widetilde{g} \rangle_0]^2.$$

Moreover, $[\langle J \rangle_0] = L_{\mathrm{id}}(J)$, so its convergence in probability to $[\langle \widetilde{g} \rangle_0]$ follows from Propositions 1 and 2. Assembling these results, we find that

$$\widehat{\mathrm{Eval}}(J) = |J|^2 - [\langle J \rangle_0^2] - [\langle J \rangle_0]^2 \xrightarrow{P} |\widetilde{g}|^2 + \langle [\widetilde{g}]^2 \rangle_0 - [\langle \widetilde{g} \rangle_0^2] - [\langle \widetilde{g} \rangle_0]^2,$$

which equals $\mathrm{Eval}(\widetilde{g})$. Also,

$$\mathbb{E}\,\widehat{\mathrm{Eval}}(J) = \mathbb{E}|J|^2 - \mathbb{E}[\langle J \rangle_0^2] - \mathbb{E}[\langle J \rangle_0]^2 = O(T^{-1}) + \mathrm{Eval}(\widetilde{g}).$$

$\square$

Proof of Theorem 4. We begin by expanding $\widehat{\mathrm{Eval}}(J) - \mathrm{Eval}(\widetilde{g})$ into four terms:

$$
\begin{aligned}
\widehat{\mathrm{Eval}}(J) - \mathrm{Eval}(\widetilde{g}) &= |J|^2 - |\widetilde{g}|^2 - \langle [\widetilde{g}]^2 \rangle_0 \\
&\quad - [\langle J - \widetilde{g} \rangle_0 \, \langle J + \widetilde{g} \rangle_0] - \langle [J - \widetilde{g}] \rangle_0 \, \langle [J + \widetilde{g}] \rangle_0 \\
&= \left(|J|^2 - \mathbb{E}|J|^2 \right) + \left(\mathbb{E}|J|^2 - |\widetilde{g}|^2 - \langle [\widetilde{g}]^2 \rangle_0 \right) \\
&\quad - \mathrm{vec}\langle J' - \widetilde{g}' \rangle_0{}' \, \mathrm{vec}\langle J + \widetilde{g} \rangle_0 - \langle [J - \widetilde{g}] \rangle_0 \, \langle [J + \widetilde{g}] \rangle_0.
\end{aligned}
$$

The second term is $O(T^{-1})$ by (A.7). For the third term, we know that $\mathrm{vec}\langle J + \widetilde{g} \rangle_0 = \mathrm{vec}\langle 2\widetilde{g} \rangle_0 + O_P(T^{-1/2})$, whereas

$$\langle J - \widetilde{g} \rangle_0 = \langle J - \mathbb{E}J \rangle_0 + \langle \mathbb{E}J - \widetilde{g} \rangle_0 = O_P(T^{-1/2}) + O(T^{-1}).$$

As a result, the third term equals

$$\mathrm{vec}\langle J' - \mathbb{E}J' \rangle_0{}' \, \mathrm{vec}\langle 2\widetilde{g} \rangle_0 + O_P(T^{-1}) + O(T^{-1}) = 2\left[\langle \widetilde{g} \rangle_0 \, \langle J - \mathbb{E}J \rangle_0 \right] + O_P(T^{-1}) + O(T^{-1}).$$

Similarly, for the fourth term we have $\langle [J + \widetilde{g}] \rangle_0 = \langle [2\widetilde{g}] \rangle_0 + O_P(T^{-1/2})$, so that overall we obtain

$$\widehat{\mathrm{Eval}}(J) - \mathrm{Eval}(\widetilde{g}) = \left(|J|^2 - \mathbb{E}|J|^2 \right) - 2[\langle \widetilde{g} \rangle_0 \, \langle J - \mathbb{E}J \rangle_0] - 2\,[\langle \widetilde{g} \rangle_0] \cdot [\langle J - \mathbb{E}J \rangle_0] + O(T^{-1}) + O_P(T^{-1}).$$

In the notation of Theorem 1, we are studying a linear combination of two linear and one quadratic functional, each centered by its expectation. Note that by passing $\langle \widetilde{g} \rangle_0$ into the inner integral, we obtain

$$[\langle \widetilde{g} \rangle_0 \, \langle J \rangle_0] = L_{\langle \widetilde{g} \rangle_0}(J).$$

Therefore

$$\sqrt{T}\left(\widehat{\mathrm{Eval}}(J) - \mathrm{Eval}(\widetilde{g})\right)$$

$$= \sqrt{T}\left(-2\left(L_{\langle\widetilde{g}\rangle_0}(J) - \mathbb{E}L_{\langle\widetilde{g}\rangle_0}(J)\right) - 2[\langle\widetilde{g}\rangle_0]\left(L_{\mathrm{id}}(J) - \mathbb{E}L_{\mathrm{id}}(J)\right)\right.$$

$$\left. + \left(Q_{\mathrm{id},\mathrm{id}}(J,J) - \mathbb{E}Q_{\mathrm{id},\mathrm{id}}(J,J)\right)\right) + o_P(1)$$

$$= \{-2, -2[\langle\widetilde{g}\rangle_0], 1\} \cdot \sqrt{T}\begin{bmatrix} L_{\langle\widetilde{g}\rangle_0}(J) - \mathbb{E}L_{\langle\widetilde{g}\rangle_0}(J) \\ L_{\mathrm{id}}(J) - \mathbb{E}L_{\mathrm{id}}(J) \\ Q_{\mathrm{id},\mathrm{id}}(J,J) - \mathbb{E}Q_{\mathrm{id},\mathrm{id}}(J,J) \end{bmatrix} + o_P(1).$$

Applying Theorem 1, the trivariate random vector is asymptotically normal with variance matrix W as given in the statement of the theorem, and with individual entries computed according to Proposition 2. The stated result now follows. $\qquad\square$

Proof of Corollary 1. Taking $\widetilde{g} \equiv \Sigma = \langle\widetilde{g}\rangle_0$ in the variance expressions of Theorem 4 yields

$$V_{\langle\widetilde{g}\rangle_0|\langle\widetilde{g}\rangle_0} = 2\left[\Sigma^4\right] + (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\left[\Sigma\,\widetilde{g}(\lambda, -\lambda, \omega)\,\Sigma\right]\right]d\lambda\,d\omega,$$

$$V_{\mathrm{id}|\langle\widetilde{g}\rangle_0} = 2\left[\Sigma^3\right] + (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\left[\widetilde{g}(\lambda, -\lambda, \omega)\,\Sigma\right]\right]d\lambda\,d\omega,$$

$$V_{\mathrm{id}|\mathrm{id}} = 2\left[\Sigma^2\right] + (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\left[\widetilde{g}(\lambda, -\lambda, \omega)\right]\right]d\lambda\,d\omega,$$

$$V_{\langle\widetilde{g}\rangle_0|\mathrm{id},\mathrm{id}} = 4\left[\Sigma\right]\left[\Sigma^3\right] + 4\left[\Sigma^4\right] + (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}g_{123,\langle\widetilde{g}\rangle_0}(\lambda, \omega)\,d\lambda\,d\omega,$$

$$g_{123,\langle\widetilde{g}\rangle_0}(\lambda, \omega) = 2\left[\left[\widetilde{g}(\lambda, -\lambda, \omega)\,\Sigma\right]\right]\left[\Sigma\right] + 2\left[\left[\Sigma\,\widetilde{g}(\lambda, -\lambda, \omega)\,\Sigma\right]\right],$$

$$V_{\mathrm{id}|\mathrm{id},\mathrm{id}} = 4\left[\Sigma\right]\left[\Sigma^2\right] + 4\left[\Sigma^3\right] + (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}g_{123,\mathrm{id}}(\lambda, \omega)\,d\lambda\,d\omega,$$

$$g_{123,\mathrm{id}}(\lambda, \omega) = 2\left[\left[\widetilde{g}(\lambda, -\lambda, \omega)\right]\right]\left[\Sigma\right] + 2\left[\left[\Sigma\,\widetilde{g}(\lambda, -\lambda, \omega)\right]\right],$$

$$V_{\mathrm{id},\mathrm{id}|\mathrm{id},\mathrm{id}} = 8\left[\Sigma\right]^2\left[\Sigma^2\right] + 16\left[\Sigma\right]\left[\Sigma^3\right] + 12\left[\Sigma^4\right] + 4\left[\Sigma^2\right]^2 + (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}g_{1234}(\lambda, \omega)\,d\lambda\,d\omega,$$

$$g_{1234}(\lambda, \omega) = 4\left[\Sigma\right]^2\left[\left[\widetilde{g}(\lambda, -\lambda, \omega)\right]\right] + 4\left[\Sigma\right]\left[\left[\widetilde{g}(\lambda, -\lambda, \omega)\,\Sigma\right]\right]$$

$$+ 4\left[\left[\Sigma\,\widetilde{g}(\lambda, -\lambda, \omega)\right]\right]\left[\Sigma\right] + 4\left[\left[\Sigma\,\widetilde{g}(\lambda, -\lambda, \omega)\,\Sigma\right]\right].$$

Now with $v' = \{-2, -2[\Sigma], 1\}$, we find the limiting variance is (after some cancellations)

$$4\,V_{\langle\widetilde{g}\rangle_0|\langle\widetilde{g}\rangle_0} + 8\left[\Sigma\right]V_{\mathrm{id}|\langle\widetilde{g}\rangle_0} + 4\left[\Sigma\right]^2 V_{\mathrm{id}|\mathrm{id}} - 4\,V_{\langle\widetilde{g}\rangle_0|\mathrm{id},\mathrm{id}} - 4\left[\Sigma\right]V_{\mathrm{id}|\mathrm{id},\mathrm{id}} + V_{\mathrm{id},\mathrm{id}|\mathrm{id},\mathrm{id}}$$

$$= 4\left[\Sigma^4\right] + 4\left[\Sigma^2\right]^2 + (2\pi)^{-2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}h_{1234}(\lambda, \omega)\,d\lambda\,d\omega,$$

$$h_{1234}(\lambda, \omega) = 4\left[\Sigma\right]\left[\left[\widetilde{g}(\lambda, -\lambda, \omega)\,\Sigma\right]\right] - 4\left[\Sigma\right]\left[\left[\Sigma\,\widetilde{g}(\lambda, -\lambda, \omega)\right]\right].$$

Next, we claim that the double integral of $[[A\,f(\lambda, -\lambda, \omega)\,B]]$ is equal to that of $[[B\,f(\omega, -\omega, \lambda)\,A]]$, for any tri-spectrum f and matrices A and B. This is proved from the definition of the double bracket, and utilizing the third property of the tri-spectrum derived in the proof of Proposition 2, namely that $f_{\ell ksr}(\lambda, -\lambda, \omega) = f_{sr\ell k}(\omega, -\omega, \lambda)$; two changes of variable in the sums allow us to swap pairs of indices, and thereby interchange the positions of A and B. As a result, the integral of h_{1234} is zero, and the result is proved. $\qquad\square$

12

References

[1] Brillinger, D.R. (2001) *Time Series: Data Analysis and Theory.* Philadelphia: Siam.

[2] Taniguchi, M., and Kakizawa, Y. (2000) *Asymptotic Theory of Statistical Inference for Time Series.* Springer-Verlag, New York.

Appendix B Supplementary Tables

A.1 Bivariate Inflation

We computed the Gaussian divergence (-2 times the log Gaussian likelihood) for MOM, unrestricted MLE, univariate MLE (i.e., the models restricted such that all cross-correlations are zero, resulting in univariate structural models), and common trend MLE (i.e., the trend covariance matrix is enforced to have rank one), resulting in -1680.292, -1708.125, -1665.786, and -1708.125 respectively. Since lower values are better, we see that the MOM fit appears to be better than the univariate MLE fit, but substantially worse than unrestricted MLE. The common trend MLE gives approximately the same divergence as the unrestricted case, because the trend correlation is so close to unity when it is freely estimated.

We can also directly compare the resulting covariance matrix estimates, expressed here in units of millions. A substantial discrepancy between the MOM and MLE results is observed, although both fits to the data ultimately provide an adequate whitening transformation.

	MOM		unrestricted MLE		constrained MLE	
	Core	Total	Core	Total	Core	Total
Core	15.660	34.396	4.588	4.977	4.589	4.979
Total	34.396	122.889	4.977	5.400	4.979	5.402

Table B.1: Estimates of Σ_μ, the trend covariance matrix (units of millions) for bivariate inflation data, based on MOM, unrestricted MLE, and common trends MLE.

	MOM		unrestricted MLE		constrained MLE	
	Core	Total	Core	Total	Core	Total
Core	10.781	5.256	18.916	27.757	18.914	27.749
Total	5.256	81.805	27.757	174.140	27.749	174.124

Table B.2: Estimates of Σ_ι, the irregular covariance matrix (units of millions) for bivariate inflation data, based on MOM, unrestricted MLE, and common trends MLE.

A.2 Four-variate Housing Starts

For the reduced span of the last nine years, we fitted both MOM and MLE (unrestricted), with divergences 959.806 and 913.573 respectively. From the standpoint of likelihood, the MOM estimates are inferior to MLE, although both adequately whiten the data. As for the covariance estimates, there is a fairly close agreement between the MLE and MOM based on the nine-year span; the MOM covariances based on the full span are also quite close.

	9-year MLE				9-year MOM				49-year MOM			
	South	West	NE	MW	South	West	NE	MW	South	West	NE	MW
South	0.0954	0.0471	0.0170	0.0326	0.0875	0.0397	0.0124	0.0280	0.0941	0.0380	0.0111	0.0214
West	0.0471	0.0274	0.0101	0.0196	0.0397	0.0209	0.0064	0.0138	0.0380	0.0240	0.0054	0.0110
NE	0.0170	0.0101	0.0049	0.0081	0.0124	0.0064	0.0027	0.0047	0.0111	0.0054	0.0027	0.0039
MW	0.0326	0.0196	0.0081	0.0170	0.0280	0.0138	0.0047	0.0096	0.0214	0.0110	0.0039	0.0088

Table B.3: Estimates of Σ_μ, the trend covariance matrix for four-variate Starts data, based on MLE (9-year span), MOM (9-year span), and MOM (49-year span).

	9-year MLE				9-year MOM				49-year MOM			
	South	West	NE	MW	South	West	NE	MW	South	West	NE	MW
South	0.0797	0.0698	0.0177	0.0399	0.0651	0.0744	0.0159	0.0511	0.1072	0.0868	0.0276	0.0644
West	0.0698	0.0747	0.0201	0.0457	0.0744	0.0897	0.0163	0.0568	0.0868	0.0801	0.0201	0.0485
NE	0.0177	0.0201	0.0080	0.0150	0.0159	0.0163	0.0046	0.0131	0.0276	0.0201	0.0089	0.0172
MW	0.0399	0.0457	0.0150	0.0385	0.0511	0.0568	0.0131	0.0406	0.0644	0.0485	0.0172	0.0401

Table B.4: Estimates of the first atomic seasonal covariance matrix for four-variate Starts data, based on MLE (9-year span), MOM (9-year span), and MOM (49-year span).

	9-year MLE				9-year MOM				49-year MOM			
	South	West	NE	MW	South	West	NE	MW	South	West	NE	MW
South	0.0176	0.0017	0.0035	0.0149	0.0154	-0.0000	0.0111	0.0280	0.0233	0.0142	0.0064	0.0285
West	0.0017	0.0094	0.0048	0.0072	-0.0000	0.0233	-0.0062	-0.0013	0.0142	0.0119	-0.0006	0.0153
NE	0.0035	0.0048	0.0117	0.0158	0.0111	-0.0062	0.0122	0.0204	0.0064	-0.0006	0.0079	0.0107
MW	0.0149	0.0072	0.0158	0.0449	0.0280	-0.0013	0.0204	0.0508	0.0285	0.0153	0.0107	0.0362

Table B.5: Estimates of the second atomic seasonal covariance matrix for four-variate Starts data, based on MLE (9-year span), MOM (9-year span), and MOM (49-year span).

	9-year MLE				9-year MOM				49-year MOM			
	South	West	NE	MW	South	West	NE	MW	South	West	NE	MW
South	0.0471	0.0435	-0.0187	0.0115	0.0512	0.0649	-0.0007	0.0585	0.0747	-0.0133	0.0225	0.0410
West	0.0435	0.0489	-0.0204	0.0138	0.0649	0.1116	-0.0155	0.0585	-0.0133	0.0317	-0.0042	-0.0078
NE	-0.0187	-0.0204	0.0155	-0.0004	-0.0007	-0.0155	0.0611	0.0205	0.0225	-0.0042	0.0200	0.0113
MW	0.0115	0.0138	-0.0004	0.0143	0.0585	0.0585	0.0205	0.0786	0.0410	-0.0078	0.0113	0.0226

Table B.6: Estimates of the third atomic seasonal covariance matrix for four-variate Starts data, based on MLE (9-year span), MOM (9-year span), and MOM (49-year span).

	9-year MLE				9-year MOM				49-year MOM			
	South	West	NE	MW	South	West	NE	MW	South	West	NE	MW
South	0.1256	0.0330	-0.0228	0.0188	0.0151	0.0074	-0.0142	0.0143	0.0418	-0.0065	-0.0018	-0.0074
West	0.0330	0.0349	-0.0088	0.0106	0.0074	0.0679	-0.0137	0.0236	-0.0065	0.0020	-0.0004	0.0015
NE	-0.0228	-0.0088	0.0078	-0.0045	-0.0142	-0.0137	0.0141	-0.0152	-0.0018	-0.0004	0.0005	0.0001
MW	0.0188	0.0106	-0.0045	0.0088	0.0143	0.0236	-0.0152	0.0178	-0.0074	0.0015	0.0001	0.0014

Table B.7: Estimates of the fourth atomic seasonal covariance matrix for four-variate Starts data, based on MLE (9-year span), MOM (9-year span), and MOM (49-year span).

	9-year MLE				9-year MOM				49-year MOM			
	South	West	NE	MW	South	West	NE	MW	South	West	NE	MW
South	0.0107	0.0151	-0.0001	-0.0024	0.0139	0.0246	0.0014	0.0086	0.0090	0.0107	-0.0022	0.0038
West	0.0151	0.0304	0.0001	0.0079	0.0246	0.0440	0.0024	0.0136	0.0107	0.0127	-0.0026	0.0043
NE	-0.0001	0.0001	0.0009	0.0005	0.0014	0.0024	0.0002	0.0022	-0.0022	-0.0026	0.0008	-0.0026
MW	-0.0024	0.0079	0.0005	0.0169	0.0086	0.0136	0.0022	0.0256	0.0038	0.0043	-0.0026	0.0106

Table B.8: Estimates of the fifth atomic seasonal covariance matrix for four-variate Starts data, based on MLE (9-year span), MOM (9-year span), and MOM (49-year span).

	9-year MLE				9-year MOM				49-year MOM			
	South	West	NE	MW	South	West	NE	MW	South	West	NE	MW
South	0.1179	-0.0117	0.0138	0.0186	0.1109	0.0034	0.0094	0.0434	0.0298	-0.0055	0.0027	0.0134
West	-0.0117	0.0277	-0.0023	0.0030	0.0034	0.0034	-0.0012	0.0002	-0.0055	0.0022	0.0009	-0.0028
NE	0.0138	-0.0023	0.0077	0.0040	0.0094	-0.0012	0.0015	0.0042	0.0027	0.0009	0.0020	0.0008
MW	0.0186	0.0030	0.0040	0.0311	0.0434	0.0002	0.0042	0.0174	0.0134	-0.0028	0.0008	0.0061

Table B.9: Estimates of the sixth atomic seasonal covariance matrix for four-variate Starts data, based on MLE (9-year span), MOM (9-year span), and MOM (49-year span).

	9-year MLE				9-year MOM				49-year MOM			
	South	West	NE	MW	South	West	NE	MW	South	West	NE	MW
South	6.5252	0.8111	0.1909	-0.5699	10.4195	-0.2665	0.2241	-3.5556	7.8623	0.0752	0.1554	-0.5432
West	0.8111	1.2563	0.2305	0.4889	-0.2665	0.6848	0.2621	0.4431	0.0752	2.7924	0.0972	0.2862
NE	0.1909	0.2305	0.3038	0.1529	0.2241	0.2621	0.5119	-0.0478	0.1554	0.0972	0.9486	0.1649
MW	-0.5699	0.4889	0.1529	1.0762	-3.5556	0.4431	-0.0478	2.3841	-0.5432	0.2862	0.1649	1.8240

Table B.10: Estimates of Σ_ι, the irregular covariance matrix for four-variate Starts data, based on MLE (9-year span), MOM (9-year span), and MOM (49-year span).