Received: 23 July 2019 | Accepted: 14 September 2021

DOI: 10.1111/rssb.12480

ROYAL STATISTICAL SOCIETY
Series B Statistical Methodology

ORIGINAL ARTICLE

Model identification via total Frobenius norm of multivariate spectra

Tucker S. McElroy[1] | **Anindya Roy**[1,2]

[1]U.S. Census Bureau, Washington, District of Columbia, USA

[2]University of Maryland, Baltimore County, Baltimore, Maryland, USA

Correspondence
Tucker S. McElroy, Research and Methodology Directorate, U.S. Census Bureau, 4600 Silver Hill Road, Washington, District of Columbia 20233-9100, USA.
Email: tucker.s.mcelroy@census.gov

Abstract

We study the integral of the Frobenius norm as a measure of the discrepancy between two multivariate spectra. 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 integrated 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 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. The numerical results show that the proposed estimator can fit moderate- to large-dimensional structural time series in real time, an option that is lacking in current literature.

KEYWORDS

Kullback-Leibler Discrepancy, Multivariate Modeling, Spectral Density, Structural Models

1 | INTRODUCTION

1.1 | Overview

When modelling vector time series data, it is of interest to know whether the model fits the data well. This problem can be addressed through a criterion function that provides a distance measure between spectral 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 (2005) discuss time-domain gof tests. There has been less literature on frequency-domain model fitting and gof testing; see Akashi et al. (2018) for recent work. Our paper proposes a total (i.e. integrated) Frobenius norm of the multivariate spectral density as a criterion function for time series, and develops the applications of model fitting and gof testing accordingly.

The main contributions of this paper are: (i) a fast method for fitting structural models to multivariate time series; (ii) a gof test for separable time series models; (iii) new limit theory for linear and quadratic functionals of the multivariate periodogram. This work can therefore be seen as a development of ideas summarized in Taniguchi and Kakizawa (2000). Taniguchi (1980) provided theory for nonlinear functionals of the periodogram, and Taniguchi (1981) treated model fitting; also see Keenan (1985). Taniguchi (1987) developed the Frobenius norm functional for univariate time series, and Chiu (1988) provided asymptotics for a weighted version of such; also see Dahlhaus and Wefelmeyer (1996). Paparoditis (2000) proposed gof testing for univariate time series based on the Frobenius norm, and Paparoditis (2005) considered the multivariate extension to a VARMA model. In recent work, Velasco and Lobato (2018) study minimum distance estimators based on a Frobenius norm distance for higher order spectra (for univariate time series), extending the framework of Anh et al. (2007).

The older literature on model fitting and gof testing generally uses a smoothed periodogram, which is a way of removing the impact of parameter uncertainty at the cost of lowering the asymptotic rate of convergence for the test statistic (through introduction of a bandwidth parameter). Some prior literature considers non-quadratic functions of the spectral density, but we focus more narrowly on the quadratic functional (through the Frobenius norm) and dispense—like Velasco and Lobato (2018)—with smoothing the periodogram; this forces us to obtain new convergence results for quadratic functions (with complicated variance expressions) but the benefit is that user-defined bandwidths are not necessary. This is because our gof statistic (for separable models, such as VARMA) has a square root rate with a simple variance that does not depend on nuisance parameters; the asymptotic variance is superior to that of Paparoditis (2005). For structural model fitting, our estimates have an analytical formula that makes modelling moderate-dimensional time series tractable.

1.2 | Discussion

Conventionally, model fitting and gof testing are handled through the Kullback-Leibler (KL) criterion function (see McElroy & Findley, 2015; Taniguchi & Kakizawa, 2000), 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 Rivers & Vuong, 2002; Vuong, 1989 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 focusing upon those frequencies emphasized by one-step ahead forecast performance, as with the KL criterion—we can expect superior performance when the model is misspecified, and when the analysis' objectives require the autocovariance function at many different lags. For example, longer term forecasts will rely on larger lag autocovariances (cf. McElroy & Wildi, 2013); we demonstrate below (see discussion after Proposition 1) that the total Frobenius norm can yield improved performance over KL when the model is misspecified.

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 Appendix A, are an attractive facet of the total Frobenius norm framework when working with moderate dimension time series. When models have a fairly low dimension, so that there are less than a hundred parameters, maximum likelihood estimation (or Bayesian methods) based on a Gaussian likelihood may be feasible computationally, and would be preferable to the Frobenius methods when the data is Gaussian. However, when the data is non-Gaussian it can be important to assess the model across all frequencies—and not merely focus on one-step ahead forecasting performance, as the Gaussian likelihood implies. Moreover, for modelling moderate- or high-dimensional structural time series the Frobenius norm approach becomes appealing to the extent that likelihood approaches become impossible.

A second application is model gof testing. Here the relevant spectra are a fitted model's spectral density and some nonparametric measure of the process' spectra, such as the periodogram or a tapered autocovariance spectral estimator (McElroy & Politis, 2014). Whereas conventional diagnostics (e.g. Ljung-Box (LB) statistics; Ljung & Box, 1978) assess residual autocorrelation, the total Frobenius norm encompasses these by insisting on complete agreement 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 goodness-of-fit is not restricted to performance at particular autocorrelation lags.

An essentially equivalent formulation of gof testing can be constructed by assessing whether the residual spectral density corresponds to white noise, that is, 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 (cf. the portmanteau of Lütkepohl, 2005). 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.

Section 2 provides the basic properties of the new criterion function, with two motivational illustrations developed in Section 3. (For the model diagnosis results, new asymptotic theory for quadratic functionals of the multivariate periodogram is developed and discussed in Appendix A.) Section 4 develops the applications to model fitting, and includes 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. Asymptotic results for Sections 4 and 5, including a new asymptotic theory for the MOM estimator of structural time series models, are gathered together separately in Section 6. We make two empirical applications in Section 7, 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 results (Appendices A and B), all proofs (Appendix C) and additional tables (Appendix D) pertaining to the empirical applications. The R code and data files needed for the simulation studies and data analysis are also available as supplementary material, and can be obtained from https://github.com/tuckermcelroy/frobnorm.

2 | FRAMEWORK

For a complex (possibly non-square) matrix A, the Frobenius norm is defined via $\|A\| = \sqrt{\mathrm{Tr}(AA^*)}$, where * denotes conjugate transpose. We will abbreviate the trace of a matrix by $[[A]]$. Consider a stationary m-dimensional vector time series $\{x_t\}$, where for each time t we denote the components by $x_t' = [x_{t,1}, \ldots, x_{t,m}]$. The process' autocovariance function is $\Gamma(h) = \mathrm{Cov}(x_{t+h}, x_t)$, so that for any $1 \leq j, k \leq m$, $\Gamma_{jk}(h) = \mathrm{Cov}(x_{t+h,j}, x_{t,k})$. When moments of suitable order exist, the autocumulant functions are defined via

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

for $a_1, \ldots, a_k \in \{1, \ldots, m\}$. Clearly, $\gamma_{a_1,a_2}(t_1) = \Gamma_{a_1,a_2}(t_1)$. To establish the theoretical results of this paper, we will adopt Assumption 1(k), where $k \geq 2$:

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

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

The spectral density f is defined via $f(\lambda) = \{f_{jk}(\lambda)\}_{j,k=1}^m$, where $f_{jk}(\lambda) = \sum_{h\in\mathbb{Z}} \Gamma_{jk}(h)e^{-ih\lambda}$. The spectral density is a Hermitian matrix-valued function, which implies that $f(\lambda)^* = f(\lambda)$ for each λ; hence $f(\lambda)f(\lambda)^* = f(\lambda)^2$. Evidently, the Frobenius norm of $f(\lambda)$ depends on λ:

$$\|f(\lambda)\|^2 = \mathrm{Tr} f(\lambda)f(\lambda)^* = \mathrm{Tr} f(\lambda)^2 = [[f(\lambda)^2]].$$

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 of 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 $\Gamma_{jk}(h) = (2\pi)^{-1}\int_{-\pi}^{\pi} f_{jk}(\lambda)\, e^{i\lambda h}\, d\lambda$ for any $1 \le j, k \le m$. This says that $\Gamma(h)$ is the hth Fourier coefficient of f. Henceforth we denote via $\langle f\rangle_h$ such an integral divided by 2π, so that for any $1 \le j, k \le m$, $\langle f_{jk}\rangle_h = \Gamma_{jk}(h)$. 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 $|f|$, that is, the total Frobenius norm of f is $|f| := \sqrt{[[\langle f^2\rangle_0]]}$. Given two m-variate spectra f and g, we define the Frobenius Discrepancy (FD) as the squared total Frobenius spectral norm of their difference:

$$\mathrm{FD}(f, g) = |f - g|^2. \tag{1}$$

The Frobenius norm has the property that $\|A\| = 0$ if and only if $A_{jk} = 0$ for all entries $1 \le j, k \le m$. 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 $1 \le j, k \le m$. 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,

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

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*.

Either f or g can be the multivariate periodogram, which we define next. In applications 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 $\bar{x}$. Let the sample autocovariance be defined as $\widehat{\Gamma}(h) = T^{-1}\sum_{t=1}^{T-h} (x_{t+h} - \bar{x})(x_t - \bar{x})'$ for $h \ge 0$, 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 $d(\lambda) = \sum_{t=1}^{T} (x_t - \bar{x})\, e^{-i\lambda t}$. 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 $\langle I\rangle_h = \widehat{\Gamma}(h)$.

Another 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 denoted by $f\, g^{-1}$, and takes a more specific form if the model is separable (discussed in Section 5). 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 (McElroy & Findley, 2015); then

$$\mathrm{KL}(f, g) = \langle [[f\, g^{-1}]]\rangle_0 + \langle \log\det g\rangle_0.$$

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–Kolmogorov filter (McElroy & Trimbur, 2015).

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, that is, 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 $fg^{-1} \equiv \langle fg^{-1}\rangle_0$. This is equivalent a.e. to the formulation in terms of the total Frobenius norm: $|fg^{-1} - \langle fg^{-1}\rangle_0| = 0$. Although the spectral residual fg^{-1} is not a spectral density (it is not Hermitian, in general), by direct calculation

$$|fg^{-1} - \langle fg^{-1}\rangle_0|^2 = [[\langle (fg^{-1})^2\rangle_0]] - [[\langle fg^{-1}\rangle_0^2]] = \sum_{h\neq 0} \|\langle fg^{-1}\rangle_h\|^2.$$

Hence, this quantity equals zero if and only if all the residual autocovariances—that is, the quantities $\langle fg^{-1}\rangle_h$—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 $|fg^{-1} - \langle fg^{-1}\rangle_0| = 0$, 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 model non-Gaussian processes as well. Denote the spectral density of the data process by $\tilde{f}$. If $\tilde{f} \in \mathcal{F}$, then the model is correctly specified, and there exists some true $\tilde{\theta}$ such that $\tilde{f} = f_{\tilde{\theta}}$. 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 $\tilde{\theta}$. If the model is misspecified, then $\tilde{f} \notin \mathcal{F}$, but we still obtain a minimizer $\tilde{\theta}$, which is called the pseudo-true value (PTV); see McElroy (2016) for background. The PTV yields the element of $\mathcal{F}$ that is closest to $\tilde{f}$, according to the distance metric corresponding to the fitting criterion.

From this discussion, it is apparent that $\mathrm{FD}(f_\theta, \tilde{f})$ could be used as a model fitting criterion, with the value zero attained at the minimizer $\tilde{\theta}$ corresponding to complete agreement, that is, $f_{\tilde{\theta}} = \tilde{f}$ *a.e.*. 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, $\tilde{f}$ 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 $\widehat{f} f_{\widehat{\theta}}^{-1}$ (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 $\tilde{\xi}$. Potentially both models are misspecified. 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, that is, $\text{FD}(f_{\tilde{\theta}}, f_{\tilde{\xi}}) = 0$, and this would be tested by fitting both models and computing $\text{FD}(f_{\hat{\theta}}, f_{\hat{\xi}})$. We shall not formally pursue this application any further here.

3 | APPLICATIONS AND ILLUSTRATIONS

This section provides a more concrete discussion of the preceding topics, focusing on Vector Autoregression (VAR) forecasting and fitting structural models.

3.1 | VAR forecasting

To further motivate the use of the total Frobenius norm criterion, we next explore an illustration involving forecasting. Consider fitting a pth order VAR, which has spectral density $f_\theta(\lambda) = \Phi(e^{-i\lambda})^{-1}\,\Sigma\,\Phi(e^{i\lambda})^{-1\prime}$, where $\Phi(z) = 1_m - \sum_{j=1}^{p} \Phi_j z^j$, Φ_p is invertible and 1_m denotes the identity matrix of dimension m. Here Σ is the error covariance matrix of VAR(p) model, and θ describes the parameter matrices $\Phi_1, \ldots, \Phi_p, \Sigma$ in a single vector. Suppose this model is misspecified, and the true spectral density is $\tilde{f}(\lambda) = \Psi(e^{-i\lambda})\,\Omega\,\Psi(e^{i\lambda})'$, where $\Psi(z) = 1_m + \sum_{j=1}^{\infty} \Psi_j z^j$ and Ω is the innovation covariance matrix. Let $\phi = \text{vec}[\Phi_1, \ldots, \Phi_p]$ denote the vectorized form of the coefficient matrices of the VAR(p) model, and let $\sigma = \text{vec}\,\Sigma$, so that $\theta' = [\phi', \sigma']$ (we write θ with the redundancies in σ for easier exposition). Then the PTVs are

$$\tilde{\theta} = \arg\min_{\phi,\sigma} \{[[\Sigma\Gamma_\Phi(0)]] - 2[[\Sigma\Gamma_\Psi(0)]]\}, \tag{2}$$

where $\Gamma_\Phi(0)$ is the variance of a process with spectral density

$$\Phi(e^{i\lambda})^{-1\prime}\,\Phi(e^{-i\lambda})^{-1}\,\Sigma\,\Phi(e^{i\lambda})^{-1\prime}\,\Phi(e^{-i\lambda})^{-1},$$

and $\Gamma_\Psi(0)$ is the variance of a process with spectral density $\Phi(e^{i\lambda})^{-1\prime}\,\Psi(e^{-i\lambda})\,\Omega\,\Psi(e^{i\lambda})'\,\Phi(e^{-i\lambda})^{-1}$. For complex z of unit magnitude, $(1_m - \Phi_1'\bar{z} - \ldots - \Phi_p'\bar{z}^p)^{-1} = -z^p(1_m + \Phi_p^{-1\prime}\Phi_{p-1}'z + \ldots + \Phi_p^{-1\prime}\Phi_1' z^{p-1} - \Phi_p^{-1\prime} z^p)^{-1}\,\Phi_p^{-1\prime}$, and therefore $\Gamma_\Phi(0)$ is the variance of a VAR($2p$) model with VAR polynomial

$$(1_m - \Phi_1 z - \ldots - \Phi_p z^p)(1_m + \Phi_p^{-1\prime}\Phi_{p-1}'z + \ldots + \Phi_p^{-1\prime}\Phi_1' z^{p-1} - \Phi_p^{-1\prime} z^p).$$

Similarly, $\Gamma_\Psi(0)$ can be identified as the variance of a linear process with appropriate coefficient matrices. The PTVs for ϕ and σ can be computed separately by concentrating the objective function.

Proposition 1 *Let $\Phi(z) = 1_m - \sum_{j=1}^{p} \Phi_j z^j$ be the matrix polynomial corresponding to a VAR(p) model and let $\phi = \text{vec}[\Phi_1, \ldots, \Phi_p]$. Define matrices $A(\phi)$ and $B(\phi)$ via*

$$A(\phi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi(e^{-i\lambda})^{-1\prime}\,\Phi(e^{i\lambda})^{-1} \otimes \Phi(e^{i\lambda})^{-1\prime}\,\Phi(e^{-i\lambda})^{-1}\,d\lambda,$$
$$B(\phi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi(e^{-i\lambda})^{-1\prime}\,\Psi(e^{i\lambda}) \otimes \Phi(e^{i\lambda})^{-1\prime}\,\Psi(e^{-i\lambda})\,d\lambda.$$

When fitting a VAR(p) model to a stationary process via the FD criterion (1), the PTVs defined in Equation (2) are given by the formulas

$$\tilde{\phi} = \arg\max_{\phi} \{vec[\Omega]' B(\phi)' A(\phi)^{-1} B(\phi) vec[\Omega]\}, \tag{3}$$

$$\tilde{\sigma} = A(\tilde{\phi})^{-1} B(\tilde{\phi}) vec[\Omega]. \tag{4}$$

For a specific illustration, consider the case of a univariate AR(1) given by $x_t = \phi x_{t-1} + z_t$ with $\{z_t\}$ a white noise process of variance σ^2; suppose that the true process is an invertible MA(1) process $x_t = w_t + \theta w_{t-1}$ with $\{w_t\}$ a white noise process of variance ω^2. Then $\sigma^2 A(\phi) = \sigma^2(1+\phi^2)(1-\phi^2)^{-3}$ is the variance of an AR(2) process with roots (ϕ^{-1}, ϕ^{-1}) and error variance σ^2. Also, $\omega^2 B(\phi) = \omega^2(1+\theta^2+2\phi\theta)/(1-\phi^2)$ is the variance of an ARMA(1,1) with parameters ϕ and θ, and error variance ω^2. Let $\rho = \theta/(1+\theta^2)$ be the lag one autocorrelation of the true process. By the assumption of invertibility, $\rho \in (-1/2, 1/2)$. Then from Equation (3), $\tilde{\phi}$ (the maximizer of $\omega^4(1+\theta^2+2\phi\theta)^2(1-\phi^2)/(1+\phi^2)$), is a solution to the equation

$$q(\phi) := \rho - \phi - 2\rho\phi^2 - \rho\phi^4 = 0. \tag{5}$$

Equation (5) is a depressed quartic equation with a negative discriminant. Hence it has two real roots and two complex roots. Looking at the signs of the equation at $\pm\rho$ and $\pm\infty$, we see that if $\rho > 0$ then the real roots are in the intervals $[-\rho, \rho]$ and $(-\infty, -\rho)$ and if $\rho < 0$ then the real roots are in the intervals $[\rho, -\rho]$ and $(-\rho, \infty)$. If $\rho = 0$ then all roots are zero. Thus, there is a unique root $\tilde{\phi}$ in the interval $(-\rho, \rho)$, and that is the PTV for ϕ.

On the other hand, ρ is also the Yule-Walker solution for this model fitting exercise, and is the PTV obtained from minimizing the KL distance between the true MA(1) process and the assumed AR(1) model. Suppose one is forecasting using the assumed AR(1) model with a coefficient ϕ. Then the h-step ahead forecast error variance will be

$$\omega^2(1+\theta^2)(1+\phi^{2h} - 2\theta\phi^h 1_{\{h=1\}}).$$

Since $|\tilde{\phi}| \le |\rho|$, for $h > 1$ the prediction error will be smaller using the Frobenius norm minimizer than that obtained from KL, whereas for $h = 1$ the KL-minimizer (which is known to be optimal for one-step ahead forecast) will yield a smaller forecast Mean Squared Error (MSE). This shows how the total Frobenius norm can yield superior results to KL when models are misspecified, but high lag information about the process' serial dynamics is needed.

More generally, suppose we wish to fit a multivariate VAR(1), but suppose this is a misspecification. In this case, the Yule-Walker PTV is $\tilde{\Phi}_1^{YW} = R(1)$, where $R(h) = \Gamma(h)\Gamma(0)^{-1}$ is the multivariate autocorrelation function. From McElroy and Findley (2015) we obtain the h-step ahead forecast error MSE (for any $h \ge 1$) for a VAR(1) process with parameter Φ_1:

$$\begin{aligned}
&\langle (1_m - \Phi_1^h e^{-ih\cdot}) \tilde{f} (1_m - \Phi_1^h e^{ih\cdot})' \rangle_0 \\
&\quad = \Gamma(0) - \Phi_1^h \Gamma(-h) - \Gamma(h)\Phi_1^{h\prime} + \Phi_1^h \Gamma(0) \Phi_1^{h\prime} \\
&\quad = \Gamma(0) - R(h)\Gamma(0)R(h)' + (\Phi_1^h - R(h))\Gamma(0)(\Phi_1^h - R(h))'.
\end{aligned}$$

This MSE is minimized (in the sense of the ordering of positive definite matrices—see McElroy & Findley, 2015) if we can find Φ_1 such that $\Phi_1^h = R(h)$; when $h = 1$, this is satisfied by the

Yule-Walker PTV $\tilde{\Phi}_1^{YW}$. In the case that $h > 1$, it is possible for the Frobenious Norm PTV for Φ_1 to yield a lower MSE than the Yule-Walker PTV. To illustrate this point, for simplicity take $\Gamma(0) = 1_m$ and suppose this is known; then

$$|f_\theta - \tilde{f}|^2 = 2 \sum_{k \geq 1} \|\Phi_1^k - R(k)\|^2,$$

using the fact that the lag k autocovariance for a VAR(1) model equals Φ_1^k (for $k \geq 1$) when the variance matrix is assumed to be 1_m. Evaluating at the Yule-Walker PTV, the expression becomes $2 \sum_{k \geq 2} \|[\tilde{\Phi}_1^{YW}]^k - R(k)\|^2$, which must dominate the corresponding expression for the Frobenius Norm PTV (since this PTV minimizes the Frobenius Norm criterion), denoted $\tilde{\Phi}_1^{FN}$. We conclude that there exists $k \geq 2$ such that $\| [\tilde{\Phi}_1^{FN}]^k - R(k) \| \leq \| [\tilde{\Phi}_1^{YW}]^k - R(k) \|$. The trace of the k-step ahead forecast MSE is

$$[[1_m - R(k)R(k)']] + \|\Phi_1^k - R(k)\|^2,$$

so we see that the trace MSE evaluated at the Frobenius Norm PTV is less than or equal to the trace MSE evaluated at the Yule-Walker PTV, that is, k-step ahead forecast performance is superior (in the sense of summing the MSE for each component) for the Frobenius Norm PTV.

3.2 | Structural models

Another application arises from the fitting of structural models (Harvey, 1989). For a time series $\{x_t\}$ that has been differenced to stationarity, the model spectrum (as shown in McElroy, 2017) takes the form

$$f_\theta(\lambda) = \sum_{k=1}^{K} g_k(\lambda) \Theta_k, \tag{6}$$

for scalar parameter-free real-valued even functions g_k, and positive definite (p.d.) parameter matrices Θ_k. For example, consider a structural model with a trend $\{\mu_t\}$, seasonal $\{\xi_t\}$ (of period s), and irregular $\{\iota_t\}$ component, which are related to the m-dimensional observed process $\{x_t\}$ via

$$x_t = \mu_t + \xi_t + \iota_t. \tag{7}$$

Letting B denote the backshift operator, we define the component processes via $(1 - B)\mu_t = \eta_t$, $(1 + B + B^2 + \ldots + B^{s-1})\xi_t = \zeta_t$, 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. Then application of $1 - B^s = (1 - B)(1 + B + B^2 + \ldots + B^{s-1})$ reduces the process to stationarity, and the quantities in Equation (6) are $g_1(\lambda) = |1 + e^{-i\lambda} + e^{-i2\lambda} + \ldots + e^{-i(s-1)\lambda}|^2$ (for the trend component), $g_2(\lambda) = |1 - e^{-i\lambda}|^2$ (for the seasonal component), and $g_3(\lambda) = |1 - e^{-is\lambda}|^2$ (for the irregular component), and such that $\Theta_1 = \Sigma_\mu$, $\Theta_2 = \Sigma_\xi$, and $\Theta_3 = \Sigma_\iota$.

In general, by writing $\Theta = [\Theta_1, \Theta_2, \ldots, \Theta_K]$ we can then define the parameter $\theta = \text{vec}[\Theta]$. (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 $|\tilde{f} - f_\theta|$, where we allow the components of θ to be real-valued. That is, we do not enforce the p.d. constraints on θ.

Proposition 2 *The MOM estimator for the structural model (6) has formula*

$$\widehat{\Theta} = \langle g' \otimes I \rangle_0 \cdot \left[G^{-1} \otimes 1_m \right] = \langle g' G^{-1} \otimes I \rangle_0, \tag{8}$$

where I is the periodogram and $G_{ik} = \langle g_i g_k \rangle_0$.

Note that the MOM estimator in Equation (8) 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 $\tilde{\theta}$, and does indeed correspond to p.d. $\tilde{\Theta}_k$. Replacing I by the true spectral density in Proposition 2, it is immediate that the PTV exists whenever $\tilde{\Theta}_k$ is p.d., and in such a case the PTV is unique. A further application of these MOM estimators is a test for cointegration: if a particular variance matrix Θ_k has reduced rank, then the kth component (if non-stationary) can exhibit collinearity (McElroy, 2017).

4 | MODEL FITTING

4.1 | General theory

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, which are discussed in Appendix A of the Supplementary Material. For continuous real matrix-valued functions $\varphi_1, \varphi_2, \varphi_3$ of frequency λ we define quadratic and linear functionals via

$$Q_{\varphi_1,\varphi_2}(f,g) = \langle [[\varphi_1 f \, \varphi_2 g]] \rangle_0 \qquad L_{\varphi_3}(f) = \langle [[\varphi_3 f]] \rangle_0. \tag{9}$$

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 $j = 1, 2, ..., T$ is in the set $[-\pi, \pi]$, and define

$$\widehat{Q}_{\varphi_1,\varphi_2}(f,g) = T^{-1} \sum_{j=1}^{T} [[\varphi_1 f \, \varphi_2 g]](\lambda_j) \quad \widehat{L}_{\varphi_3}(f) = T^{-1} \sum_{j=1}^{T} [[\varphi_3 f]](\lambda_j). \tag{10}$$

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 $\widehat{L}_{\varphi_3}(I)$; the technical details are provided in Appendix A.

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

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

where the subscript function id refers to the case where the function is the identity matrix. Replacing the quadratic functionals by their empirical versions yields the criterion function

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

We define the estimator $\widehat{\theta}$ obtained by minimizing this criterion. In Section 6, it is shown that $\widehat{\text{FD}}(f_\theta, I)$ converges in probability to $\text{FD}(f_\theta, \tilde{f})$, plus a positive bias term. So long as $\widehat{\text{FD}}(f_\theta, I)$ is continuous with respect to θ, then $\widehat{\theta}$ is consistent for the PTV, which is the minimizer of $|\tilde{f} - f_\theta|$. (Note that the minimizer of $|\tilde{f} - f_\theta|$ is equal to the minimizer of $|\tilde{f} - f_\theta|^2 + \langle [[\tilde{f}]]^2 \rangle_0$ as well, because the second term is free of θ.) The PTV is denoted $\tilde{\theta}$. The asymptotic normality of $\widehat{\theta}$ for $\tilde{\theta}$ is developed in Section 6, and the special case of a structural model is also treated there.

4.2 | MOM estimator simulation: Large number of parameters

To evaluate the finite sample performance of the MOM estimator (8) under different levels of model complexity, we used a structural model with trend and seasonal components, and evaluated the efficiency of estimators via root mean squared error (RMSE). Although it is desirable to compare the MOM estimator with the maximum likelihood estimator (MLE), or other likelihood-based estimators, even for moderate dimensions (at least six) the computation proved to be prohibitive. Maximum likelihood procedures compute the likelihood usually using the Kalman filter (Durbin & Koopman, 2001) and then use optimizers with the likelihood as input. Bayesian structural time series computations also use the Kalman filter based likelihood, and proceed by using Markov chain Monte Chain computation (Petris et al., 2009; Scott & Varian, 2014, 2015).

Most available methods either focus on univariate structural time series or allow for parameter-restricted multivariate time series. The MARSS package in R (Holmes et al., 2012) could be used for fitting multivariate processes, but computation of high-dimensional processes can be time consuming, if not infeasible. A recent paper on multivariate structural time series (Qiu et al., 2018) does analyse an eight-dimensional time series, but none of the literature reports any repeated sampling simulation results for moderate-dimensional structural time series. In essence, while there is extensive literature available for estimating univariate structural time series, fast methods providing adequate estimates for multivariate structural time series of moderate to high dimension are non-existent. The proposed MOM estimator fills this gap.

For univariate series we compared the efficiency of the MLE (computed using STSM in R) with the MOM estimator, and indeed for univariate process the MLE was in general 20–30% more efficient for the cases considered. For higher dimensional processes we computed the MLE using the MARSS package in R as well as the *Ecce Signum* R code developed for multivariate signal extraction. For dimensions 2 and 3 the MLE, when it could be reliably computed, was more efficient than the MOM estimator; but even for moderate dimensions the MLE computation time

became infeasible—and also the optimization routine did not always converge. The routines in MARSS were faster than those in *Ecce Signum* but still for a series of dimension seven, one single fit took 4930 s. For a 50-dimensional process with nearly 4000 parameters the computation of the MOM estimator required <4 s.

The purpose of this simulation experiment is to demonstrate that even for processes with a large number of parameters (e.g. 1000 parameters), the MOM estimator can be computed in real time with reliable accuracy. Specifically, we consider the trend-seasonal-irregular structural model defined via (7). We varied the process dimension m between 1 and 25. At $m = 25$ the process has up to 975 parameters. The sample sizes were chosen to be $T = 100, 1000$. We set $\Sigma_\iota = 1_m$, and for Σ_μ and Σ_ξ we replicated four instances for each dimension $1 \le m \le 25$. For each instance, we randomly sampled the ranks of the variance matrices, r_μ and r_ξ, uniformly from $\{1, ..., m\}$. Thus we had 100 different scenarios with the parameter dimension potentially ranging from 3 to 975. To generate a positive definite variance matrix of rank r we randomly sampled $m \times r$ matrices C whose entries were i.i.d. standard Gaussian, and set the variance matrix to CC'. Thus, $\Sigma_\mu = C_\mu C_\mu'$ and $\Sigma_\xi = C_\xi C_\xi'$, where C_μ and C_ξ were $m \times r_\mu$ and $m \times r_\xi$ full column rank matrices.

For each reduced rank variance matrix, we also noted the effective number of free parameters. An $m \times m$ nonnegative definite matrix of rank r has $r(2m - r + 1)/2$ number of free parameters. Thus, for a parameter configuration where the ranks of Σ_μ and Σ_ξ were r_μ and r_ξ, respectively, the model had

$$p_{\text{eff}} = r_\mu(2m - r_\mu + 1)/2 + r_\xi(2m - r_\xi + 1)/2 + m(m+1)/2$$

number of free parameters. For example, for a process with dimension $m = 20$, where the ranks for the trend and seasonal components were 14 and 3, respectively, the model had 456 parameters. While the maximum dimension considered in the simulation was $m = 25$, the maximum number of effective parameters in the randomly sampled instances was $p_{\text{eff}} = 939$.

We generated $n = 5000$ Monte Carlo replications (using Gaussian innovations for each structural component) for each of the hundred randomly generated parameter configurations and recorded the efficiency of the MOM estimators, as measured by their Monte Carlo RMSE; this measure for a MOM estimator $\widehat{\Theta}_k$ of Θ_k, normalized by $\sqrt{T}$, is defined as

$$\text{Average RMSE}(\widehat{\Theta}_k) = T^{1/2} \begin{pmatrix} m+1 \\ 2 \end{pmatrix}^{-1} n^{-1} \sum_{i=1}^{n} \|\widehat{\Theta}_k - \Theta_k\|^2,$$

where i is the index of the simulation. The Average RMSE for the entries of the different component matrices are given in the left panels of Figure 1. For each parameter configuration the RMSE is plotted against the number of free parameters, p_{eff}; this illustrates the sensitivity of estimation error to model complexity. From the figure it is clear that the estimation error, when normalized by the sample size and the number of parameters, is very consistent across different degrees of model complexity and sample sizes. There is a slight upward trend in the RMSE for the entries as the models become more complex, indicating that $\sqrt{T}$ approximation may not be sufficient in finite samples when the parameter dimension is large; however, overall the estimation error seems to be remarkably stable even for higher dimensional processes with a large number of parameters.

The asymptotic variance expression for the MOM estimators is derived under Assumption 1 of Section 2. To evaluate the performance of the estimators when these moment conditions are

violated, we repeated the simulation experiment using Student's t vector white noise processes (with degrees of freedom equal to 4, and standardized to have unit variance) for the error processes $\{\eta_t\}$ and $\{\zeta_t\}$. These error processes are extremely heavy-tailed. The Average RMSE of the different variance components are given in the right panels of Figure 1. The RMSE values are slightly more variable across different scenarios, with occasional large values. However, the overall behavior for the RMSE values is nearly the same as in the Gaussian case.

4.3 | Rank testing

An interesting application of the asymptotic theory (developed in Theorem 2 below) is that of testing for reduced rank structures of the variance matrices. Following McElroy and Jach (2019), 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 & 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 2 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 & Jach, 2019 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 $m = 2$ 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 Remark 1 and Theorem 2 (see Section 6) associated with the entries of Θ_k, with parameters replaced by their MOM estimates.

For simulation, we used model (7) with the following parameter specifications:

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

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 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 $\rho_\xi \in \{0, 0.4, 0.8, 0.9, 1\}$. 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 unity the power of the test will increase. Table 1 shows the size and power of the test for a reduced rank component

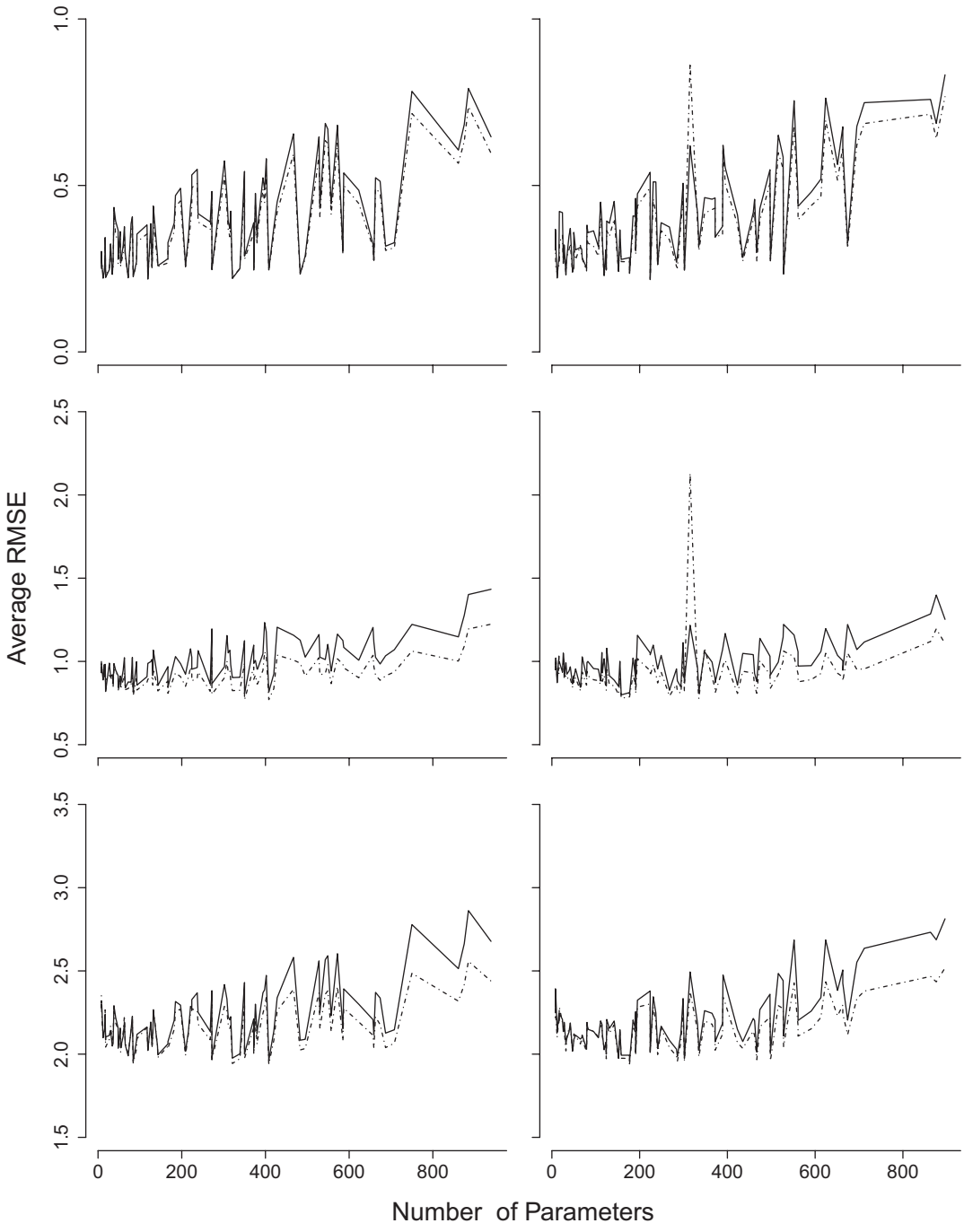

FIGURE 1 The method-of-moments estimator's average root mean squared error is shown along the *y*-axis, and the number of effective parameters in the model (p_{eff}) is shown along the *x*-axis. The solid line corresponds to $T = 1000$, and the dashed line corresponds to $T = 100$. The left and right panels are respectively for Gaussian and Student's *t* (4 degrees of freedom) distributions, while the top, middle, and bottom panels correspond, respectively, to the trend, seasonal and the irregular variance components

for both the trend and the seasonal components, based on a nominal $\alpha = 0.05$. The values corresponding to the null hypothesis 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.

5 | MODEL EVALUATION

In this section, we consider testing model goodness-of-fit using the total Frobenius norm. We assume that the models are separable, and have been fitted with a concentrated Whittle likelihood (described below), but one wants to evaluate the model fit by examining the residuals in frequency-domain, across all frequencies via the total Frobenius norm. Appendix B of the Supplementary material considers the case of non-separable models.

5.1 | Methodology

A separable model for $\{x_t\}$ takes the form $x_t = \Psi_\theta(B)\epsilon_t$, where $\{\epsilon_t\}$ is a multivariate white noise with covariance matrix Σ, and $\Psi_\theta(z) = \sum_{j=0}^{\infty} \Psi_j z^j$ such that $\Psi_0 = 1_m$ and the matrices Ψ_j each depend on the parameter vector θ—but Σ is separately parameterized. For example, VARMA models are separately parameterized. The spectral density is

$$f_\theta(\lambda) = \Psi_\theta(e^{-i\lambda})\,\Sigma\,\Psi_\theta(e^{i\lambda})', \tag{11}$$

and we let $\underset{\sim}{f}_\theta(\lambda) = \Psi_\theta(e^{-i\lambda})\Psi_\theta(e^{i\lambda})'$ correspond to setting $\Sigma = 1_m$ in Equation (11). McElroy (2018) shows that the concentrated Whittle likelihood for a separable model is (up to a constant) the log determinant of the forecast error variance matrix

$$\widehat{\Sigma}(\theta) = \langle \Psi_\theta(e^{-i\cdot})^{-1}\, I\, \Psi_\theta(e^{i\cdot})^{-1\prime} \rangle_0.$$

In particular, the minimizer (when it exists and is unique) of $\log\det\widehat{\Sigma}(\theta)$ is the Whittle estimator $\widehat{\theta}$, and the noise covariance matrix is estimated via $\widehat{\Sigma}(\widehat{\theta})$. In this estimate, we integrate (and divide by 2π) the quantity

$$\widehat{J} = \Psi_{\widehat{\theta}}(e^{-i\cdot})^{-1}\, I\, \Psi_{\widehat{\theta}}(e^{i\cdot})^{-1\prime},$$

which can be viewed as an estimate of the spectral density of the residual process $\Psi_{\widehat{\theta}}(B)^{-1} X_t$, that is,

$$\widehat{g} = \Psi_{\widehat{\theta}}(e^{-i\cdot})^{-1}\, \tilde{f}\, \Psi_{\widehat{\theta}}(e^{i\cdot})^{-1\prime}.$$

If we replace I by the true spectral density $\tilde{f}$ in the forecast error variance matrix, then the minimizer $\tilde{\theta}$ is the PTV for θ based on the Whittle likelihood. Then by replacing $\widehat{\theta}$ by $\tilde{\theta}$ in $\widehat{J}$ and $\widehat{g}$, we obtain 'asymptotic versions' (they are only asymptotic in the sense that the impact of parameter uncertainty has been removed) defined respectively as

$$\begin{aligned} \tilde{J} &= \Psi_{\tilde{\theta}}(e^{-i\cdot})^{-1}\, I\, \Psi_{\tilde{\theta}}(e^{i\cdot})^{-1\prime} \\ \tilde{g} &= \Psi_{\tilde{\theta}}(e^{-i\cdot})^{-1}\, \tilde{f}\, \Psi_{\tilde{\theta}}(e^{i\cdot})^{-1\prime}. \end{aligned}$$

TABLE 1 Power and size of the test of reduced rank structure based on method-of-moments (MOM) estimates

		$H_0: \det(\Sigma_\mu) = 0$					$H_0: \det(\Sigma_\xi) = 0$				
		ρ_ξ					ρ_ξ				
		0.00	**0.40**	**0.80**	**0.90**	**1.00**	**0.00**	**0.40**	**0.80**	**0.90**	**1.00**
		$T = 200$					$T = 200$				
ρ_μ	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**
	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**
		$T = 500$					$T = 500$				
ρ_μ	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**
	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**

Note: Type I errors at the null values are given in bold.

Note that if the model is correctly specified, then $\tilde{f} = f_{\tilde{\theta}}$ and plugging in Equation (11) yields $\tilde{g} = \Sigma$. This motivates the testing problem $|\tilde{g}|^2 - \|\langle \tilde{g} \rangle_0\|^2 = 0$, because $|\tilde{g}|^2 - \|\langle \tilde{g} \rangle_0\|^2$ is non-negative and equals zero if and only if $\tilde{g}$ is constant, that is, corresponds to white noise. Rewriting this functional as

$$|\tilde{g}|^2 - \|\langle \tilde{g} \rangle_0\|^2 = Q_{\underline{f}_{\tilde{\theta}}^{-1}, \underline{f}_{\tilde{\theta}}^{-1}}(\tilde{f}, \tilde{f}) - [[\langle \tilde{g} \rangle_0^2]],$$

it can be seen that substituting $\hat{J}$ for $\tilde{g}$ yields a statistic that converges to $|\tilde{g}|^2 + \langle [[\tilde{g}]]^2 \rangle_0 - [[\langle \tilde{g} \rangle_0^2]]$ (this is shown in the proof of Proposition 4, below). In the case that $\tilde{g} \equiv \Sigma$, 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 $|\tilde{g}|^2 - \|\langle \tilde{g} \rangle_0\|^2$), we propose to subtract the quantity $[[\langle \hat{J} \rangle_0]]^2$, which converges to $[[\Sigma]]^2$ under the null. These arguments suggest defining a model evaluation functional $\mathrm{Eval}(\tilde{g})$ and estimator $\widehat{\mathrm{Eval}}(\hat{J})$ as follows:

$$\mathrm{Eval}(\tilde{g}) = |\tilde{g}|^2 - \|\langle \tilde{g} \rangle_0\|^2 + \langle [[\tilde{g}]]^2 \rangle_0 - [[\langle \tilde{g} \rangle_0]]^2 = Q^2_{\underline{f}_{\tilde{\theta}}^{-1}, \underline{f}_{\tilde{\theta}}^{-1}}(\tilde{f}, \tilde{f}) + \langle [[\tilde{g}]] \rangle_0 - [[\langle \tilde{g} \rangle_0^2]] - [[\langle \tilde{g} \rangle_0]]^2$$

$$\widehat{\mathrm{Eval}}(\hat{J}) = |\hat{J}|^2 - \|\langle \hat{J} \rangle_0\|^2 - [[\langle \hat{J} \rangle_0]]^2 = \hat{Q}_{\underline{f}_{\hat{\theta}}^{-1}, \underline{f}_{\hat{\theta}}^{-1}}(I, I) - [[\langle \hat{J} \rangle_0^2]] - [[\langle \hat{J} \rangle_0]]^2.$$

The estimator $\widehat{\mathrm{Eval}}(\hat{J})$ can be alternatively expressed with the final two terms, $[[\langle \hat{J} \rangle_0^2]]$ and $[[\langle \hat{J} \rangle_0]]^2$, 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 (cf. Chen & Deo, 2000). The evaluation estimator converges in probability to the correct quantity for the testing problem, using Assumption 1 of Section 2; this is shown in Section 6 below.

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 $[[\langle \widehat{J} \rangle_0]]^2$. The asymptotic normality of $\widehat{\text{Eval}}(\widehat{J})$ is provided in Section 6. Whereas the limiting variance matrix in general is somewhat complicated, under the null hypothesis of the correct model specification there is a helpful simplification, such that $\sqrt{T}\widehat{\text{Eval}}(\widehat{J})$ is asymptotically normal with mean zero and variance $4[[\Sigma^4]] + 4[[\Sigma^2]]^2$ (Corollary 1 below).

5.2 | Simulation results

To test the utility of Corollary 1 (see Section 6) 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

$$x_t = \begin{bmatrix} 0.3 & -0.3 \\ 0 & 0.4 \end{bmatrix} x_{t-1} + \begin{bmatrix} -0.01 & -0.1 \\ -0.1 & 0.25 \end{bmatrix} x_{t-2} + z_t, \tag{12}$$

where $\{z_t\}$ are Gaussian white noise with mean zero and variance 1_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 $p \in \{1, 2, ..., 8\}$. Thus, for $p = 1$ the model will be misspecified (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 2 provides the proportion of empirical rejections for two-sided tests of correct model order (with nominal $\alpha = 0.05$), using the limiting distribution of $\widehat{\text{Eval}}(\widehat{J})$ given in Corollary 1, where the error variance Σ is estimated from the residuals. Of course in the misspecified models the estimate of Σ will be also affected by the misspecification.

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 (12) 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 2; the proportions are remarkably close to the corresponding values for the Gaussian case, indicating the robustness of the proposed test against distributional assumptions.

6 | ASYMPTOTIC RESULTS

6.1 | Model fitting

Here we summarize the asymptotic results needed for model fitting (Section 4.1), and begin with the convergence of $\widehat{\text{FD}}(f_\theta, I)$. (This result relies on Assumption 1 of Section 2.)

Proposition 3 *Assume that* $\{x_t\}$ *is strictly stationary with spectral density* $\tilde{f}$*, and satisfies Assumption 1*(k) *for* $2 \leq k \leq 8$*. Then for all* $\theta \in \Theta$

TABLE 2 Size and power of the model evaluation test, computed based on residuals from a VAR(p) to the VAR(2) model (12), for different model orders and different sample sizes

	Gaussian innovations			Student's t innovations		
p	$T = 200$	$T = 500$	$T = 1000$	$T = 200$	$T = 500$	$T = 1000$
1	0.089	0.217	0.697	0.062	0.183	0.376
2	0.024	0.043	0.045	0.029	0.048	0.051
3	0.027	0.050	0.055	0.050	0.066	0.059
4	0.038	0.059	0.059	0.062	0.080	0.077
5	0.062	0.073	0.064	0.085	0.100	0.084
6	0.078	0.089	0.064	0.100	0.127	0.098
7	0.096	0.108	0.066	0.132	0.139	0.130
8	0.103	0.132	0.082	0.146	0.176	0.139

Notes: 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 either a Gaussian distribution or a multivariate Student's t with 4 degrees of freedom and identity as the scale matrix.

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

Next, we examine asymptotic normality; define the matrix M_θ with jkth entry given by $[[\langle \frac{\partial f_\theta}{\partial \theta_j} \frac{\partial f_\theta}{\partial \theta_k} \rangle_0]]$.

Theorem 1 *Assume that* $\{x_t\}$ *is strictly stationary with spectral density* $\tilde{f}$ *and satisfies Assumption 1*(k) *for* $\mathrm{k} \geq 2$. *Also suppose that* $\widehat{\mathrm{FD}}(f_\theta, I)$ *is a twice continuously differentiable function of* θ, *that the PTV* $\tilde{\theta}$ *exists and is unique, and is in the interior of the parameter space. If* $M_{\tilde{\theta}}$ *is invertible, then as* $\mathrm{T} \to \infty$

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

where V_θ *is a matrix with jkth entry given by*

$$V_{jk} = 2\left[\left[\left\langle \tilde{f} \frac{\partial f_\theta}{\partial \theta_j} \tilde{f} \frac{\partial f_\theta}{\partial \theta_k} \right\rangle_0\right]\right] + \left\langle \left\langle \left[\left[\frac{\partial f_\theta}{\partial \theta_j} \tilde{f}(\lambda, -\lambda, \omega) \frac{\partial f_\omega}{\partial \theta_k} \right]\right] \right\rangle_0 \right\rangle_0,$$

and $\tilde{f}(\lambda, -\lambda, \omega)$ *is the tri-spectral density (defined in Appendix A).*

Next, we apply this result to the case of structural models (cf. Equation 6); when the structural model is correctly specified, the secondary conditions of Theorem 1 are satisfied, and the asymptotic theory for $\widehat{\theta}$ is fairly straightforward.

Theorem 2 *Assume that* $\{x_t\}$ *is strictly stationary with spectral density* $\tilde{f}$ *and satisfies Assumption 1*(k) *for* $\mathrm{k} \geq 2$. *If the structural model (6) is correctly specified, then* $\widehat{\theta}$ *is consistent for* $\tilde{\theta}$ *and*

$$\sqrt{T}(\widehat{\theta} - \tilde{\theta}) \xRightarrow{\mathcal{L}} \mathcal{N}\left(0, 2\langle A(\tilde{f}' \otimes \tilde{f}) A' \rangle_0 + \langle \langle A(\lambda) \tilde{F}(\lambda, -\lambda, \omega) A(\omega)' \rangle_0 \rangle_0\right),$$

where $A = G^{-1} g \otimes 1_{m^2}$ *and* $\tilde{F}$ *is a block-matrix form of the multivariate tri-spectrum defined in Appendix A.*

Remark 1 The asymptotic variance matrix in Theorem 2 can be written in terms of the M and V matrices of Theorem 1, which in the case that the tri-spectral density is zero work out to be $M = G \otimes 1_{m^2}$ and $V = 2\langle gg' \otimes (\tilde{f}' \otimes \tilde{f}) \rangle_0$. From Equation (6) 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 4-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 2 (so long as the tri-spectrum is zero).

6.2 | Model evaluation

Here we summarize the asymptotic results needed for model evaluation (Section 5.1).

Proposition 4 *Assume that* $\{x_t\}$ *is strictly stationary with spectral density* $\tilde{f}$*, and satisfies Assumption 1*(k) *for* k = *2, 3, 4. Suppose that the model for* $\{x_t\}$ *is separable, and* $\Psi_\theta(z)$ *is twice continuously differentiable in* θ*. Let the PTV* $\tilde{\theta}$ *be the minimizer of* $\mathrm{KL}(\tilde{f}, f_{-\theta})$*; if it exists uniquely in the interior of the parameter space, then as* $\mathrm{T} \to \infty$, $\widehat{\mathrm{Eval}}(\widehat{J}) \xrightarrow{P} \mathrm{Eval}(\tilde{g})$*, and*

$$\mathbb{E}[|\tilde{J}|^2] = O(T^{-1}) + |\tilde{g}|^2 + \langle [[\tilde{g}]]^2 \rangle_0. \tag{13}$$

Remark 2 Proposition 4 does not assume that the null hypothesis is true. But if the residual is white noise, then $\tilde{g} \equiv \Sigma$, and

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

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

In the following result on asymptotic normality, for any matrix-valued function $a(\theta)$ we let $\nabla' a(\theta)$ denote the block row matrix $[\frac{\partial a(\theta)}{\partial \theta_1}, \ldots, \frac{\partial a(\theta)}{\partial \theta_q}]$ (where θ has q components). Note that if c is an $m \times m$ matrix, then $[[c \nabla' a(\theta)]]$ is $1 \times q$, with kth entry given by $[[c \frac{\partial a(\theta)}{\partial \theta_k}]]$. Analogously, the $q \times q$ matrix $H(\theta)$ is defined via $H(\theta) = \langle [[\tilde{f} \nabla \nabla' f_{-\theta}]] \rangle_0$, where the jkth entry is given by $\langle [[\tilde{f} \frac{\partial^2 f_{-\theta}}{\partial \theta_j \partial \theta_k}]] \rangle_0$.

Theorem 3 *Assume that* $\{x_t\}$ *is strictly stationary with spectral density* $\tilde{g}$ *and satisfies Assumption 1*(k) *for* k ≥ *2. Suppose that the model for* $\{x_t\}$ *is separable, and* $\Psi_\theta(z)$ *is twice continuously differentiable in* θ*. Let the PTV* $\tilde{\theta}$ *be the minimizer of* $\mathrm{KL}(\tilde{f}, f_{-\theta})$*; if it exists uniquely in the interior of the parameter space, then as* $\mathrm{T} \to \infty$

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

where W and v are defined as

$$W = \begin{bmatrix} V_{b|b} & V_{b|f_{-\tilde{\theta}}^{-1}} & V_{b|f_{-\tilde{\theta}}^{-1}, f_{-\tilde{\theta}}^{-1}} & V_{b|\nabla f_{-\tilde{\theta}}^{-1}} \\ V_{f_{-\tilde{\theta}}^{-1}|b} & V_{f_{-\tilde{\theta}}^{-1}|f_{-\tilde{\theta}}^{-1}} & V_{f_{-\tilde{\theta}}^{-1}|f_{-\tilde{\theta}}^{-1}, f_{-\tilde{\theta}}^{-1}} & V_{f_{-\tilde{\theta}}^{-1}|\nabla f_{-\tilde{\theta}}^{-1}} \\ V_{f_{-\tilde{\theta}}^{-1}, f_{-\tilde{\theta}}^{-1}|b} & V_{f_{-\tilde{\theta}}^{-1}, f_{-\tilde{\theta}}^{-1}|f_{-\tilde{\theta}}^{-1}} & V_{f_{-\tilde{\theta}}^{-1}, f_{-\tilde{\theta}}^{-1}|f_{-\tilde{\theta}}^{-1}, f_{-\tilde{\theta}}^{-1}} & V_{f_{-\tilde{\theta}}^{-1}, f_{-\tilde{\theta}}^{-1}|\nabla f_{-\tilde{\theta}}^{-1}} \\ V_{\nabla f_{-\tilde{\theta}}^{-1}|b} & V_{\nabla f_{-\tilde{\theta}}^{-1}, f_{-\tilde{\theta}}^{-1}} & V_{\nabla f_{-\tilde{\theta}}^{-1}|f_{-\tilde{\theta}}^{-1}, f_{-\tilde{\theta}}^{-1}} & V_{\nabla f_{-\tilde{\theta}}^{-1}|\nabla f_{-\tilde{\theta}}^{-1}} \end{bmatrix}$$

$$\begin{aligned} v' = \{ & -2, -2[[\langle \tilde{g} \rangle_0]], 1, \\ & -4[[\langle \tilde{g}\Psi_{\tilde{\theta}}(e^{-i\cdot})^{-1}\tilde{f}\nabla'\Psi_{\tilde{\theta}}(e^{i\cdot})^{-1\prime}\rangle_0]]H^{-1} \\ & +4[[\langle \tilde{g}\rangle_0\langle\Psi_{\tilde{\theta}}(e^{-i\cdot})^{-1}\tilde{f}\nabla'\Psi_{\tilde{\theta}}(e^{i\cdot})^{-1\prime}\rangle_0]]H^{-1} \\ & +4[[\langle \tilde{g}\rangle_0]][[\langle\Psi_{\tilde{\theta}}(e^{-i\cdot})^{-1}\tilde{f}\nabla'\Psi_{\tilde{\theta}}(e^{i\cdot})^{-1\prime}\rangle_0]]H^{-1}\}. \end{aligned}$$

Here $b(\lambda) = \Psi_{\tilde{\theta}}(e^{i\lambda})^{-1\prime}\langle \tilde{g} \rangle \Psi_{\tilde{\theta}}(e^{-i\lambda})^{-1}$ and $H = H(\tilde{\theta})$; the entries of W are described in Theorem A.1 of Appendix A.

The entries of W are described using the notation of Theorem A.1, where $V_{\nabla a(\theta)|c}$ is shorthand for a block matrix with jth block row given by $V_{\frac{\partial a(\theta)}{\partial \theta_j}|c}$. These block matrices depend on the tri-spectrum; calculation of W can be accomplished by using plug-in estimators, as discussed in Taniguchi (1999). However, under the null hypothesis of correct model specification 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 spectral density* $\tilde{g}$ *and satisfies Assumption 1*(k) *for* $k \geq 2$. *Suppose that the model for* $\{x_t\}$ *is separable, and* $\Psi_\theta(z)$ *is twice continuously differentiable in* θ. *If the model is correctly specified, then as* $T \to \infty$

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

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

7 | EMPIRICAL ANALYSES

7.1 | 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 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 (7) 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 related trends model) and the common trends restriction, whereby Σ_μ is enforced to have rank one. These results are summarized in Appendix D of the Supplementary Material. While parameter estimates differ between the MOM and MLE, the former approach does effect a whitening transformation of 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 several minutes are required for the MLE optimization routine.

Both the MOM and MLE estimates of Σ_μ indicate the possibility of common trends, as the former estimate has a cross-correlation of 0.784, whereas the latter is 0.875 (in the related trends specification). 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 trends based on the MOM estimate, the results are 741.385 for $\det(\Theta_k)$, with an estimated variance of 500, 704.5 and a normalized test statistic of 1.0477. Hence the null hypothesis of common trends cannot be rejected.

7.2 | 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 (7) with second-order stochastic trend is used along with a seasonal $\{\xi_t\}$ that is additively composed of six atomic seasonal processes, one for each principal monthly seasonal frequency; see Appendix D of the Supplementary Material for a discussion. 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 D 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: over an hour for the MLE optimization, versus <1 s 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 <1 s, whereas a single Gaussian likelihood evaluation (with an efficient Durbin–Levinson implementation) takes more than a second.

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 those of the U.S. Census Bureau.

ORCID

Tucker S. McElroy https://orcid.org/0000-0002-2991-9067

REFERENCES

Akashi, F., Odashima, H., Taniguchi, M. & Monti, A.C. (2018) A new look at portmanteau tests. *Sankhya A*, 80(1), 121–137.

Anh, V.V., Leonenko, N.N. & Sakhno, L.M. (2007) Minimum contrast estimation of random processes based on information of second and third orders. *Journal of Statistical Planning and Inference*, 137(4), 1302–1331.

Brillinger, D.R. (2001) *Time series: data analysis and theory*. Philadelphia: Siam.

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

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

Chiu, S.T. (1988) Weighted least squares estimators on the frequency domain for the parameters of a time series. *The Annals of Statistics*, 16, 1315–1326.

Dahlhaus, R. & Wefelmeyer, W. (1996) Asymptotically optimal estimation in misspecified time series models. *The Annals of Statistics*, 24, 952–974.

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

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

Durbin, J. & Koopman, S.J. (2001) *Time series analysis by state space methods*. Oxford: Oxford University Press.

Golub G. & Van Loan, C. (1996) *Matrix computations*. Baltimore: Johns Hopkins University Press.

Harvey, A. (1989) *Forecasting, structural time series models, and the Kalman filter*. Cambridge: Cambridge University Press.

Holmes, E.E., Ward, E.J. & Wills, K. (2012) MARSS: multivariate autoregressive state-space models for analyzing time-series data. *The R Journal*, 4, 11–19.

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

Keenan, D.M. (1985) Asymptotic properties of minimization estimators for time series parameters. *The Annals of Statistics*, 13, 369–382.

Kohli P. & Pourahmadi, M. (2012) Nonparametric estimation of the innovation variance and judging the fit of ARMA models. In: Bell, W., Holan, S. & McElroy, T. (Eds.) *Economic time series: modeling and seasonality*. New York: Chapman and Hall, pp. 459–476.

Li, W.K. (2004) *Diagnostic checks in time series*. Boca Raton, FL: CRC Press.

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

Ljung, G. & Box, G. (1978) On a measure of lack of fit in time series models. *Biometrika*, 65, 297–303.

Lütkepohl, H. (2005) *New introduction to multiple time series analysis*. Berlin/Heidelberg: Springer Science & Business Media.

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

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

McElroy, T.S. (2018) Recursive computation for block nested covariance matrices. *Journal of Time Series Analysis*, 39(3), 299–312.

McElroy, T.S. & Findley, D.F. (2015) Fitting constrained vector autoregression models. In: Beran, J., Feng, Y. & Hebbel, H. (Eds.) *Empirical economic and financial research – theory, methods, and practice*. New York: Springer, pp. 451–470.

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

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

McElroy, T. & Jach, A. (2019) Testing collinearity of vector time series. *The Econometrics Journal*, 22(2), 97–116.

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

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

McElroy, T.S. & 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.Available from: http://www.federalreserve.gov/pubs/feds/2012/201245/201245abs.html

McElroy, T. & Wildi, M. (2013) Multi-step ahead estimation of time series models. *International Journal of Forecasting*, 29, 378–394.
Paparoditis, E. (2000) Spectral density based goodness-of-fit tests for time series models. *Scandinavian Journal of Statistics*, 27, 143–176.
Paparoditis, E. (2005) Testing the fit of a VARMA model. *Journal of Time Series Analysis*, 26, 543–568.
Peña, D. & Rodriguez, J. (2002) A powerful portmanteau test of lack of fit for time series. *Journal of the American Statistical Association*, 97, 601–610.
Petris, G., Petrone, S. & Campagnoli, P. (2009) Dynamic linear models. In: *Dynamic linear models with R*. New York, NY: Springer, pp. 31–84.
Qiu, J., Rao Jammalamadaka, S. & Ning, N. (2018) Multivariate Bayesian structural time series model. *The Journal of Machine Learning Research*, 19, 2744–2776.
Rivers, D. & Vuong, Q. (2002) Model selection tests for nonlinear dynamic models. *Econometrics Journal*, 5, 1–39.
Scott, S.L. & Varian, H.R. (2014) Predicting the present with Bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, 5, 4–23.
Scott, S.L. & Varian, H.R. (2015) Bayesian variable selection for nowcasting economic time series. In: *Economic analysis of the digital economy*. Chicago: University of Chicago Press, pp. 119–135.
Taniguchi, M. (1980) On estimation of the integrals of certain functions of spectral density. *Journal of Applied Probability*, 17, 73–83.
Taniguchi, M. (1981) An estimation procedure of parameters of a certain spectral density model. *Journal of the Royal Statistical Society: Series B (Methodological)*, 43, 34–40.
Taniguchi, M. (1987) Minimum contrast estimation for spectral densities of stationary processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 49, 315–325.
Taniguchi, M. (1999) Statistical analysis based on functionals of nonparametric spectral density estimators. In: Ghosh, S. (Ed.) *Asymptotics, nonparametrics, and time series*. New York: Marcel Dekker, pp. 351–394.
Taniguchi, M. & Kakizawa, Y. (2000) *Asymptotic theory of statistical inference for time series*. New York: Springer-Verlag.
Velasco, C. & Lobato, I. (2018) Frequency domain minimum distance inference for possibly noninvertible and noncausal ARMA models. *The Annals of Statistics*, 46, 555–579.
Vuong, Q. (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307–333.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: McElroy, T.S. & Roy, A. (2022) Model identification via total Frobenius norm of multivariate spectra. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84, 473–495. https://doi.org/10.1111/rssb.12480