Inference and Prediction for Quadratic Processes

Tucker McElroy[1] Dhrubajyoti Ghosh[2] Soumendra Lahiri[3]

[1]US Census Bureau

[2]Washington University

[3]Washington University

Abstract

Economic time series published by the U.S. Census Bureau often exhibit non-Gaussian behavior, and in such a case non-linear forecasts can be superior to linear forecasts. This paper develops the theoretical background for a new quadratic (non-linear) prediction method for time series through three main contributions: (1) a new theory for infinite arrays and their factorization is developed, generalizing multivariate spectral factorization theory; (2) the formula for the quadratic *h*-step ahead forecast filter (based on an infinite past) is derived; (3) a necessary and sufficient condition involving the bispectrum is derived, for discerning when quadratic prediction offers a benefit over linear prediction. Upon this chassis of new factorization and forecasting theory we lay the machinery of model fitting, residual analysis, and simulation.

[1]US Census Bureau, [2]Washington University, [3]Washington University

Disclaimer

This presentation 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. All time series analyzed in this presentation are from public or external data sources.

Funding

This work partially supported by NSF research grant 1811998, 2019-2021 (DMS Co-PI). *Development of a general framework for nonlinear prediction using auto-cumulants: Theory, methodology and computation*.

Outline

- Introduction
- Framework and Background
- Theory for Infinite Arrays
- The Quadratic Forecast Filter
- Applications
- Conclusion

Introduction

Motivation

- Monthly and quarterly economic time series published by the U.S. Census Bureau often exhibit non-Gaussian characteristics, such as *asymmetry* and *extreme values* (kurtosis).
- Asymmetry and kurtosis in a time series can be measured through the bispectral and trispectral density, respectively.
- For such processes, linear forecasts need not be optimal; a non-linear forecast can yield lower mean squared error.
- Non-linear forecast formulas can depend on the bispectrum and trispectrum, and not just on the spectrum (as in the case of linear forecasts).

Introduction

Goals

- We aim to derive the formula for the *h*-step ahead forecast of a stationary time series.
- If the time series is non-Gaussian, the forecast formula can be non-linear; we seek the quadratic approximation to the Volterra expansion of the predictor.
- In order to derive the formula, we need a new theory for infinite arrays associated with autocumulants of the stochastic process.
- We also want a necessary and sufficient condition for the quadratic forecast formula to be superior to using a linear formula.

Introduction

Main Objective

For a nonlinear stationary time series $\{X_t\}$ with autocovariance function $\{\gamma_k\}$, we provide explicit formulas for optimal filters $\Pi(z) = \sum_{k\geq 0} \pi_k z^k$ and $\Pi(z,y) = \sum_{j,k\geq 0} \pi_{j,k} z^j y^k$, where

$$\widehat{X}_{t+h|t} = \sum_{k\geq 0} \pi_k X_{t-k} + \sum_{j,k\geq 0} \pi_{j,k} (X_{t-j}X_{t-k} - \gamma_{k-j}), \qquad (1)$$

and $h \geq 1$.

Framework and Background

Autocumulants Definition

- Let $\{X_t\}$ be a $k+1$th order stationary time series with $k+1$ moments for given $k \geq 1$.
- So the $k+1$th order moments are finite and the $k+1$th order autocumulant function is defined by

$$\gamma_{h_1,\ldots,h_k} = \text{cum}\left(X_{t+h_1}, \ldots, X_{t+h_k}, X_t\right).$$

- Further, assume that the autocumulant function is absolutely summable for $\underline{h} = [h_1, \ldots, h_k]' \in \mathbb{Z}^k$.

Framework and Background

Polyspectrum Definition

- The absolute summability condition suffices to define the $k+1$th order polyspectral density

$$f(\lambda_1, \ldots, \lambda_k) = \sum_{\underline{h} \in \mathbb{Z}^k} \gamma_{\underline{h}} \exp\{-i \underline{h}' \underline{\lambda}\},$$

  where we set $\underline{\lambda} = [\lambda_1, \ldots, \lambda_k]'$, and each of these are frequencies in $[-\pi, \pi]$.
- We may append a k subscript to f: f_2 $(k=1)$ is the *spectral* density, f_3 $(k=2)$ is the *bispectral* density, and f_4 $(k=3)$ is the *trispectral* density.

Framework and Background

Relation of Autocumulant and Polyspectrum

The coefficients are recovered via integration over the unit torus:

$$\gamma_{\underline{h}} = \frac{1}{(2\pi)^k} \int_{[-\pi,\pi]^k} \exp\{i \underline{h}' \underline{\lambda}\} f(\underline{\lambda}) \, d\underline{\lambda}.$$

Framework and Background

Autocumulant Generating Function

We extend the polyspectral density to the autocumulant generating function (acgf), which is defined over $\mathbb{C}^k$, as follows:

$$f(z_1, \ldots, z_k) = \sum_{\underline{h} \in \mathbb{Z}^k} \gamma_{\underline{h}} z_1^{h_1} \cdots z_k^{h_k}.$$

This is the $k+1$th order acgf, and evaluating at $z_j = e^{-i\lambda_j}$ for $1 \leq j \leq k$ clearly yields its restriction to the polyspectral density.

Framework and Background

Shorthand

Let $\langle z^h g(z) \rangle_z = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{-i\lambda h} g(e^{-i\lambda}) \, d\lambda$ as shorthand, so that

$$\gamma_h = \langle z^{-h} f(z) \rangle_z$$
$$\gamma_{h_1,h_2} = \langle \langle z^{-h_1} y^{-h_2} f(z,y) \rangle_z \rangle_y .$$

We may consider certain portions of a Laurent series, using the following notation: for any $-\infty \leq r \leq s \leq \infty$, let

$$[f(z)]_r^s = \sum_{j=r}^{s} \langle z^{-j} f(z) \rangle_z z^j .$$

Sometimes we write this with B, the backshift operator, inserted for z when we are considering a filter.

Framework and Background

Polyspectral Factorization

The polyspectral density f_{k+1} can be factored into a product of functions Ψ_{k+1} that are k-fold power series, of the form

$$\Psi_{k+1}(z_1, \ldots, z_k) = \sum_{j_1, \ldots, j_k \geq 0} \psi^{(k+1)}_{j_1, \ldots, j_k} \, z_1^{j_1} \cdots z_k^{j_k} = \Psi_{k+1}(\underline{z}).$$

In the case $k = 1$, the factorization is the classical result

$$f(z) = \sigma^2 \, \Psi_2(z) \, \Psi_2(z^{-1}). \tag{2}$$

Theory for Infinite Arrays

Laurent Series Arrays

- We extend 1-arrays and 2-arrays to Laurent series arrays by allowing each component to be a Laurent series.
- A 1-array Laurent series $\underline{\xi}(z)$ has rth component $\xi^{(r)}(z)$ for $r \geq 0$.
- A 2-array Laurent series $B(z)$ has component r, s given by $B^{(r,s)}(z)$ for $r, s \geq 0$.
- A power series array is special case where Laurent is causal in z.
- Multiplication: $B(z)\underline{\xi}(z)$ is a 1-array with rth component $\sum_{s \geq 0} B^{(r,s)}(z) \xi^{(s)}(z)$.
- Similarly define multiplication of 2-arrays, transpose, identity array (equals 1 if $r = s$ for all z and 0 otherwise), and invertibility.

Theory for Infinite Arrays

Definition

A Laurent series 2-array $B(z)$ has a forward Wiener-Hopf factorization if there exist power series 2-arrays $B^{-}(z)$ and $B^{+}(z)$ such that

$$B(z) = B^{-}(z^{-1})B^{+}(z)'. \tag{3}$$

If instead

$$B(z) = B^{+}(z)B^{-}(z^{-1})', \tag{4}$$

we say that $B(z)$ has a backward Wiener-Hopf factorization.

Theory for Infinite Arrays

Theorem

Consider a system

$$[\underline{\tau}(z)]_0^\infty = [B(z)\,\underline{\xi}(z)]_0^\infty,$$

where $\underline{\tau}(z)$ *is a Laurent series 1-array,* $\underline{\xi}(z)$ *is a power series 1-array, and* $B(z)$ *is a Laurent series 2-array. Suppose* $B(z)$ *has a forward Wiener-Hopf factorization (3), and that* $B^+(z)$ *and* $B^-(z)$ *are invertible. Then the system is solved by*

$$\underline{\xi}(z) = B^+(z)^{-1\prime}[B^-(z^{-1})^{-1}\underline{\tau}(z)]_0^\infty.$$

Theory for Infinite Arrays

Positive Definite Laurent Series Arrays

- We say that a Hermitian Laurent series 2-array $B(z)$ is *positive definite*, denoted $B(z) > 0$, if every Schur complement is a positive definite (scalar) function.
- For any $m \geq 1$ the $(m+1)$th Schur complement is defined to be the Schur complement of the upper left $(m+1) \times (m+1)$-dimensional matrix Laurent series with respect to the upper left $m \times m$-dimensional matrix Laurent series.
- A principle example of Hermitian positive definite Laurent series 2-arrays is the autocovariance generating function of a stationary 1-array process.

Theory for Infinite Arrays

Forward WH Factorization of PD Arrays

Any positive definite Laurent series 2-array has a forward Wiener-Hopf factorization.

Theorem

Suppose that $B(z)$ is a Hermitian Laurent series 2-array that is positive definite. Then there exist power series 2-arrays $B^-(z)$ and $B^+(z)$ such that (4) holds.

The Quadratic Forecast Filter

Problem Framing

- Consider the problem of h-step ahead forecasting of a mean zero stationary process $\{X_t\}$ based on linear and quadratic functions of the infinite past $\underline{X} = \{X_s, s \leq t\}$.
- We denote this predictor by $\widehat{X}_{t+h|t}$, and it follows that it takes the form given by (1).
- The centering by γ_{k-j} in this formula is necessary to ensure the forecast has mean zero.
- Also, because $X_{t-j}X_{t-k}$ and $X_{t-k}X_{t-j}$ are equal, without loss of generality we impose that $\pi_{jk} = 0$ for $k < j$.
- Our information sets are $\{X_{t-\ell}\}_{\ell \geq 0}$ and $\{X_{t-j}X_{t-k}\}_{k \geq j \geq 0}$.

The Quadratic Forecast Filter

Review of Linear Forecast Filter

In the case of linear h-step ahead forecasting, the optimal filter is given by $\eta(B)$ (where B is the backshift operator):

$$\eta(z) = [z^{-h}\Psi_2(z)]_0^\infty \, \Psi_2(z)^{-1} = \sum_{j \geq 0} \psi_{j+h}^{(2)} z^j \, \Psi_2(z)^{-1}. \qquad (5)$$

The Quadratic Forecast Filter

Quadratic Verus Linear Forecast

The quadratic forecast reduces to a purely linear forecast if and only if

$$\text{Cov}[X_{t+h}, X_{t-j}X_{t-k} - \widehat{X_{t-j}X_{t-k}}] = 0 \qquad (6)$$

for all $k \geq j \geq 0$, where $\widehat{X_{t-j}X_{t-k}}$ denotes the linear prediction of $X_{t-j}X_{t-k}$ on the basis of $\{X_{t-\ell}\}_{\ell \geq 0}$.

The Quadratic Forecast Filter

Linear Prediction of Quadratic Terms

- We describe $\widehat{X_{t-j}X_{t-k}}$.
- For any $r \geq 0$ define the process $\{Y_t^{(r)}\}$ by $Y_t^{(r)} = X_t X_{t-r} - \gamma_r$, which is stationary with mean zero.
- Then the prediction target $X_{t-j}X_{t-k} - \gamma_{k-j}$ can be expressed as $Y_{t-j}^{(k-j)}$.
- Setting $\widehat{Y_{t-j}^{(r)}} = \Phi^{(j,r)}(B) X_t$,

$$\Phi^{(j,r)}(z) = \sigma^{-2} [z^j \Psi_2(z^{-1})^{-1} \langle y^r f(zy^{-1}, y) \rangle_y]_0^\infty \Psi_2(z)^{-1}, \quad (7)$$

where σ^2 is the innovation variance given in (2).

The Quadratic Forecast Filter

Reframe the Problem

- With (7) we can rewrite (1) in a more convenient form.
- The upper triangular form of $\Pi(z, y)$ yields

$$\Pi(z, y) = \sum_{j \geq 0} \sum_{r \geq 0} \pi_{j,j+r} z^j y^{j+r} = \sum_{r \geq 0} \Pi^{(r)}(zy) y^r$$

  by setting $\Pi^{(r)}(x) = \sum_{j \geq 0} \pi_j^{(r)} x^j$ with $\pi_j^{(r)} = \pi_{j,j+r}$.
- Then the h-step ahead quadratic predictor can be written as

$$\widehat{X}_{t+h|t} = \eta(B) X_t \oplus \sum_{r \geq 0} \Pi^{(r)}(B)(Y_t^{(r)} - \widehat{Y_t^{(r)}}). \tag{8}$$

  The first summand of (8) is orthogonal to the linear filter error.

The Quadratic Forecast Filter

Autocovariance Generating Function of $\{\underline{Y}_t\}$

- In view of (8) our objective is to compute the Laurent series 1-array $\underline{\Pi}(z)$ with rth component $\Pi^{(r)}(z)$ for $r \geq 0$.
- We define the autocovariance generating function $A(z)$ of the 1-array process $\{\underline{Y}_t\}$, which has rth component $\{Y_t^{(r)}\}$:

$$\begin{aligned} A^{(r,s)}(z) &= \langle \langle y^r x^{-s} f(zy^{-1}, y, z^{-1}x) \rangle_y \rangle_x \\ &\quad + \langle \left(y^{r-s} + y^{s+r} z^{-s}\right) f(zy^{-1}) f(y) \rangle_y \end{aligned}$$

for any $r, s \geq 0$, and satisfies

$$\langle z^{-k} A^{(r,s)}(z) \rangle_z = \mathrm{Cov}[Y_{t+k}^{(r)}, Y_t^{(s)}].$$

The Quadratic Forecast Filter

Non-redundancy Condition

- $A(z)$ is a function of the spectrum and trispectrum.
- Under the following non-redundancy condition $A(z)$ is positive definite.
- **Assumption P**: The autocovariance generating function for the linear prediction of $Y_t^{(r)}$ from $\{Y_t^{(r-1)}, \ldots, Y_t^{(0)}\}$ is positive definite for all $r \geq 1$.
- Assumption P states that for each $Y_t^{(r)}$ is not perfectly linearly predictable on the basis of the other product pair time series.

The Quadratic Forecast Filter

Linearity Condition

- Define the Laurent series 1-array $\underline{L}(z)$, with sth component $L^{(s)}(z) = \langle y^{-s} f(zy, y^{-1}) \rangle_y$ for $s \geq 0$.
- This 1-array encodes the "forward-looking" portion of the bispectrum.
- **Assumption L**: $\underline{L}(z)$ is non-zero.
- If Assumption L is violated, the quadratic filter reduces trivially to a linear filter, because $\Phi^{(j,r)}(z) = 0$.

The Quadratic Forecast Filter

Defined Expressions

$$B(z) = \left[\begin{array}{cc} 1 & -\Psi_2(z)^{-1} \underline{L}(z^{-1})' \\ -\Psi_2(z^{-1})^{-1} \underline{L}(z) & \sigma^2 A(z) \end{array} \right].$$

Define $R(z)$ to be a 1-array Laurent series with first component zero, and latter components given by

$$\sigma^2 z^{-h} [\Psi_2(z)]_0^{h-1} \, \Psi_2(z)^{-1} \underline{L}(z^{-1}).$$

The Quadratic Forecast Filter

Theorem

Let $\{X_t\}$ be a fourth order stationary time series satisfying Assumptions P and L. The Laurent series 1-array $\underline{\Pi}(z)$ appearing in the quadratic forecasting problem (8) is given by

$$\underline{\Pi}(z) = [0, I]\, B^{+}(z)^{-\prime} \left[B^{-}(z^{-1})^{-1}\, \underline{R}(z) \right]_0^{\infty}, \tag{9}$$

where the $[0, I]$ operator denotes a forward row shift acting on a Laurent series 2-array. The MSE of quadratic filter is equal to the linear MSE minus the quantity $\langle \underline{Q}'(z^{-1}) \underline{Q}(z) \rangle$, where

$$\underline{Q}(z) = \sigma^{-1} \left[B^{+}(z^{-1})^{-1}\, \underline{R}(z) \right]_0^{\infty}.$$

The Quadratic Forecast Filter

Quadratic Versus Linear

- When $\underline{R}(z) = 0$ the quadratic filter reduces to a linear filter. This is equivalent to

$$0 = [(z^{-h} - \eta(z))\underline{L}(z^{-1})]_0^{\infty}. \tag{10}$$

- This condition involves the linear forecast error $(z^{-h} - \eta(z))$ and the bispectral density through $\underline{L}(z)$.
- In the special case that the process $\{X_t\}$ is a causal linear process, it can be shown that (10) holds.

Applications

Whittle Likelihood

- The Whittle likelihood is based upon expressing the *linear* one-step ($h = 1$) ahead forecast MSE in frequency domain.
- This can be generalized to h-step ahead forecast error:

$$\begin{aligned}\epsilon_{t+h} &= X_{t+h} - B^{-h}[\Psi_2(B)]_h^{\infty}\,\Psi_2(B)^{-1}\,X_t \\ &= [\Psi_2(B)]_0^{h-1}\,\Psi_2(B)^{-1}\,X_{t+h}.\end{aligned} \tag{11}$$

Applications

Whittle Likelihood

- When considering the Whittle likelihood, we wish to allow for the model to differ from the true spectral density, denoted $\widetilde{f}$.
- So $\Psi_{k+1}(\underline{z})$ will be viewed as a model-based function, possibly parameterized by a finite-dimensional parameter vector.
- It then follows from (11) that the forecast MSE is

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{j=0}^{h-1} \psi_j^{(2)} e^{-i\lambda j} \, \Psi_2(e^{-i\lambda})^{-1} \right|^2 \widetilde{f}(\lambda) \, d\lambda. \qquad (12)$$

- Setting $h = 1$ in (12) yields the exponential of the concentrated Whittle likelihood. For estimation of models, the periodogram is substituted for $\widetilde{f}$.

Applications

Applications of Linear Forecast Error Formula

The $h = 1$ case of (11) and (12) allows us to:

1. Given $\Psi_2(z)$, recursively simulate a process $\{X_t\}$ by filtering past values with the 1-step ahead forecast filter and adding i.i.d. $\{\epsilon_t\}$, i.e.,

$$X_{t+1} = B^{-1}[\Psi_2(B)]_1^{\infty} \, \Psi_2(B)^{-1} \, X_t + \epsilon_{t+1}.$$

2. Given data, fit a parametric model (which entails a specified form for $\Psi_2(z)$) via the Whittle likelihood, i.e., replace $\widetilde{f}$ in (12) with the periodogram and set $h = 1$.
3. Given a proposed model $\Psi_2(z)$, compute residuals via (11), and check whether residual serial dependence is present in the autocumulants.

Applications

Extension to Quadratic Whittle

- To construct a quadratic prediction analogue of the Whittle likelihood, we should compute the MSE while allowing for model mis-specification.
- The quadratic forecast error is

$$\epsilon_{t+h} = (B^{-h} - \Pi(B))X_t - \underline{\Pi}(B)' \left(\underline{Y}_t - \mathbb{E}[\underline{Y}_t]\right).$$

- Given the numerical challenges of computing $\underline{\Pi}(z)$, it may be more convenient to use a finite-sample analogue.
- We can compute forecast errors in-sample based on a specified model.
- Then for $t = 1, 2, \ldots, T - h$, with sample size T, we compute the in-sample forecast errors $X_{t+h} - \widehat{X}_{t+h|1:t}$, square, and sum.

Extension to Quadratic Whittle

- Such a quantity is then viewed as a function of the parameter vector, given the data, and can be minimized numerically.
- Setting $h = 1$:

$$\sum_{t=1}^{T-1} \left(X_{t+1} - \widehat{X}_{t+1|t:1}\right)^2.$$

This is a finite-sample approximation to the analogue to the Whittle likelihood, and we call it the **Quadratic Whittle likelihood**.

Applications

Maximum Entropy Transformation

- Set $h = 1$:

$$X_{t+1} = \widehat{X}_{t+1|t} + \epsilon_{t+1}. \tag{13}$$

- The first term on the right side is a linear and quadratic function of present and past data, and with which ϵ_{t+1} is uncorrelated.
- So $\{\epsilon_t\}$ is a white noise process, and hence (13) is a causal quadratic transformation of $\{X_t\}$ to a white noise process.
- Hence $X_{t+1} - \widehat{X}_{t+1|t}$ is a **maximum entropy transformation**.

Applications

Residual Analysis

- Having fitted a model, we can compute residuals via

$$e_{t+1} = X_{t+1} - \widehat{X}_{t+1|t:1}$$

  for $t = 1, \ldots, T-1$. Here the 1-step ahead forecast is based on the quadratic filter determined by the fitted model.
- Because the true forecast errors $\{\epsilon_t\}$ are a white noise, we can expect that the residuals $e_2, \ldots, e_T$ resemble a sample from a white noise sequence.
- So we can use standard inference tools such as Portmanteau statistics to evaluate model goodness-of-fit.

Applications

Simulation

- Consider (13) viewed as a recursive prescription to simulate a process.
- With a specified model (and selected parameters) we generate the 1-step ahead forecast at time t via computing $\widehat{X}_{t+1|t:1}$, supposing we already have $X_1, \ldots, X_t$ simulated; independently we draw ϵ_{t+1} from a pdf g with variance given by the quadratic forecast MSE.
- Use a burn-in period, because this procedure is recursive.

Conclusion

New results:

1. Polyspectral factorization
2. Theory for Laurent Series Arrays: solving systems, Wiener-Hopf factorization
3. Quadratic filter formula
4. Linearity condition involving bispectrum
5. Quadratic Whittle Likelihood, residual analysis, simulation

Contact

- tucker.s.mcelroy@census.gov