

# Quadratic Prediction of Time Series via Auto-Cumulants

Tucker McElroy (U.S. Census Bureau)  
Dhrubajyoti Ghosh (North Carolina State University)  
Soumendra Lahiri (North Carolina State University)

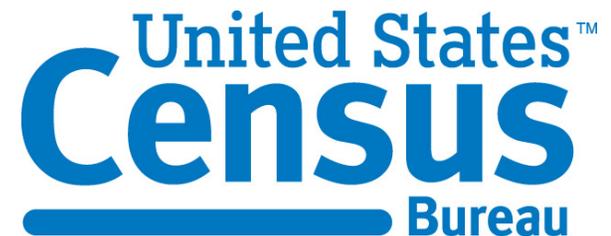
JSM, July 31, 2019



## Disclaimer and Acknowledgement

These slides are released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

This work partially support by NSF Grant Number 1613192.



# Outline

1. Background on Nonlinear Prediction
2. The Quadratic Problem
3. Theory for Quadratic Processes
4. Illustrations of Quadratic Processes



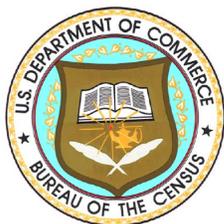
# Background on Nonlinear Prediction

**Goal:** Predict  $Y$  from random vector  $\underline{X}$ .

**Best Prediction:** The mean of  $Y|\underline{X}$ ? Minimizes mean square error (MSE) between  $Y$  and *all* functions of  $\underline{X}$ .

**Gaussian Case:**  $\mathbb{E}[Y|\underline{X}]$  is linear in  $\underline{X}$ .

**Linear Case:** A linear minimizer of MSE depends on means and covariances of  $(Y, \underline{X})$ , i.e., *first* and *second* moments.



## Background on Nonlinear Prediction

When would a nonlinear predictor give lower MSE?

- A non-Gaussian  $(Y, \underline{X})$ ? Maybe; for some non-Gaussian the linear predictor is still best.
- Conjecture: when there's some skewness and/or kurtosis.



# Background on Nonlinear Prediction

**Thesis:** This paper focuses on quadratic predictors in the context of time series forecasting. We provide a general approach involving auto-cumulants (of order 1, 2, 3, and 4).

**Other Literature:** Try to find “universal predictor” by truncating the Volterra expansion. Tsay (1986), Krolzig and Hendry (2001), Castle and Hendry (2010), Teräsvirta et al. (2010), Kock and Teräsvirta (2011).

**What’s New:** A general framework (not based on particular models, or just on forecasting); a classification of processes for which quadratic prediction is helpful; new “quadratic” processes.



# The Quadratic Problem

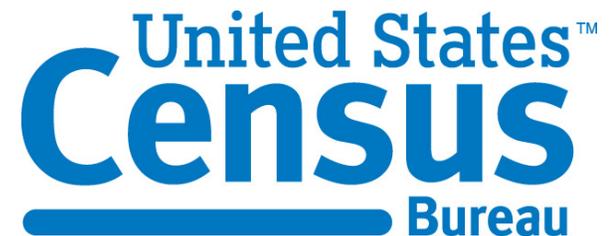
Wlog  $(Y, \underline{X})$  is mean zero. The quadratic problem seeks to minimize

$$\mathbb{E} \left[ (Y - g(\underline{X}))^2 \right],$$

where

$$g(\underline{X}) = \underline{b}' \underline{X} + \underline{X}' B \underline{X} - \mathbb{E} [\underline{X}' B \underline{X}]$$

and the matrix  $B$  is (weakly) lower-triangular.



# The Quadratic Problem

Recall the linear predictor is

$$\hat{Y}^{(1)} = \Sigma_{Y, \underline{X}} \Sigma_{\underline{X}, \underline{X}}^{-1} \underline{X}.$$



# The Quadratic Problem

Using

$$\underline{X}' B \underline{X} = \text{tr}\{\underline{X} \underline{X}' B\} = \text{vec}[B]' \text{vec}[\underline{X} \underline{X}']$$

and replacing  $\text{vec}$  by (weak)  $\text{vech}$ , we obtain

$$g(\underline{X}) = \beta' \{\mathcal{X} - \mathbb{E}[\mathcal{X}]\},$$

where  $\mathcal{X}' = [\underline{X}', \text{vech}[\underline{X} \underline{X}']']$  and  $\beta' = [\underline{b}', \text{vech}[B]']$ .



## The Quadratic Problem

Let  $M = \text{Cov}[\mathcal{X}]$ , i.e.,

$$M = \begin{bmatrix} \Sigma_{\underline{X}, \underline{X}} & \Sigma_{\underline{X}, \text{vech}[\underline{X} \underline{X}']} \\ \Sigma_{\text{vech}[\underline{X} \underline{X}'], \underline{X}} & \Sigma_{\text{vech}[\underline{X} \underline{X}'], \text{vech}[\underline{X} \underline{X}']} \end{bmatrix}.$$

Suppose that both  $\Sigma_{\underline{X}, \underline{X}}$  and the Schur complement

$$S = \Sigma_{\text{vech}[\underline{X} \underline{X}'], \text{vech}[\underline{X} \underline{X}']} - \Sigma_{\text{vech}[\underline{X} \underline{X}'], \underline{X}} \Sigma_{\underline{X}, \underline{X}}^{-1} \Sigma_{\underline{X}, \text{vech}[\underline{X} \underline{X}']}$$

are invertible.



# The Quadratic Problem

The solution to the quadratic problem is

$$\hat{\beta} = \left[ \begin{array}{c} \Sigma_{\underline{X}, \underline{X}}^{-1} \Sigma_{\underline{X}, Y} - \Sigma_{\underline{X}, \underline{X}}^{-1} \Sigma_{\underline{X}, \text{vech}[\underline{X} \underline{X}']} S^{-1} \Sigma_{\text{vech}[\underline{X} \underline{X}'], \hat{E}^{(1)}} \\ S^{-1} \Sigma_{\text{vech}[\underline{X} \underline{X}'], \hat{E}^{(1)}} \end{array} \right],$$

where  $\hat{E}^{(1)} = Y - \hat{Y}^{(1)}$  and

$$\Sigma_{\text{vech}[\underline{X} \underline{X}'], \hat{E}^{(1)}} = \Sigma_{\text{vech}[\underline{X} \underline{X}'], Y} - \Sigma_{\text{vech}[\underline{X} \underline{X}'], \underline{X}} \Sigma_{\underline{X}, \underline{X}}^{-1} \Sigma_{\underline{X}, Y}.$$



## The Quadratic Problem

**Remark:** Quadratic reduces to linear iff  $\Sigma_{\text{vech}[\underline{X} \underline{X}'], \hat{E}^{(1)}} = 0$ . This condition involves *third* moments.

**Re-expression:**

$$\hat{Y}^{(2)} = \hat{Y}^{(1)} + \Sigma_{\hat{E}^{(1)}, \text{vech}[\underline{X} \underline{X}']} S^{-1} \left[ \text{vech}[\underline{X} \underline{X}'] - \Sigma_{\text{vech}[\underline{X} \underline{X}'], \underline{X}} \Sigma_{\underline{X}, \underline{X}}^{-1} \underline{X} \right].$$



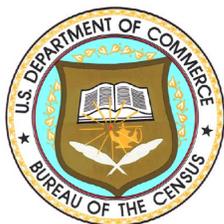
## The Quadratic Problem

**Minimal MSE:** The MSE at this optimum is

$$\Sigma_{Y,Y} - \Sigma_{Y,\underline{X}} \Sigma_{\underline{X},\underline{X}}^{-1} \Sigma_{\underline{X},Y} - \Sigma_{\hat{E}^{(1)},\text{vech}[\underline{X} \underline{X}']} S^{-1} \Sigma_{\text{vech}[\underline{X} \underline{X}'],\hat{E}^{(1)}}.$$

Therefore the efficiency loss of using a linear estimator when a quadratic is warranted is the non-negative quantity

$$\Sigma_{\hat{E}^{(1)},\text{vech}[\underline{X} \underline{X}']} S^{-1} \Sigma_{\text{vech}[\underline{X} \underline{X}'],\hat{E}^{(1)}}.$$



## Theory for Quadratic Processes

Now let  $\underline{X}' = [X_1, X_2, \dots, X_T]$  be a length  $T$  sample from a strictly stationary univariate time series  $\{X_t\}$ .

**Auto-cumulants:** If  $\{X_t\}$  has moments of all orders, then the auto-cumulant functions  $\kappa$  of order  $k + 1$  (for  $k \geq 1$ ) are defined via

$$\kappa(\underline{h}) = \text{cum}[X_{t+h_1}, X_{t+h_2}, \dots, X_{t+h_k}, X_t],$$

where  $\underline{h} = [h_1, h_2, \dots, h_k]'$  is a  $k$ -vector of lags.



# Theory for Quadratic Processes

Wlog  $\mathbb{E}[X_t] = 0$ , and define the auto-moment functions  $\gamma$  of order  $k + 1$  via

$$\gamma(\underline{h}) = \mathbb{E}[X_{t+h_1} X_{t+h_2} \cdots X_{t+h_k} X_t].$$

For  $k = 1, 2$  we have  $\kappa = \gamma$ , but for  $k \geq 3$  the auto-cumulant and auto-moment functions are different.



# Theory for Quadratic Processes

Assume auto-cumulant functions are absolutely summable over  $\underline{h} \in \mathbb{Z}^k$ . Then the polyspectral density of order  $k + 1$  is given by

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

where  $\underline{\lambda} = [\lambda_1, \dots, \lambda_k]'$  denotes a  $k$ -vector of frequencies.



## Theory for Quadratic Processes

**Polyspectra for Linear Processes:** If  $X_t = \Psi(B) Z_t$ , where  $\Psi(z)$  is a power series such that  $\Psi(0) = 1$ , and  $\{Z_t\}$  is i.i.d. (with  $k$ th cumulant  $\mu_k$ ), we say the process  $\{X_t\}$  is *linear*. In such a case the polyspectra of order  $k + 1$  is given by

$$f(\underline{\lambda}) = \mu_{k+1} \prod_{j=1}^k \Psi(e^{-i\lambda_j}) \Psi(e^{i \sum_{j=1}^k \lambda_j}).$$



## Theory for Quadratic Processes

For nonlinear processes, we can (under conditions, see Tekalp and Erdem (1989)) factor the polyspectra: for each  $k \geq 1$  there exists a constant  $\mu_{k+1}$  and power series  $\Psi_{k+1}(z)$  such that  $\Psi_{k+1}(0) = 1$  and

$$f(\underline{\lambda}) = \mu_{k+1} \prod_{j=1}^k \Psi_{k+1}(e^{-i\lambda_j}) \Psi_{k+1}(e^{i\sum_{j=1}^k \lambda_j}).$$

**Remark:** For a linear process  $\Psi_{k+1} = \Psi$  for all  $k \geq 1$ , and  $\mu_k$  corresponds to the  $k$ th cumulant of  $Z_t$ .



# Theory for Quadratic Processes

**Definition of Quadratic Process:** There is some MSE gain to using a quadratic predictor over a linear predictor (based on an infinite past).

**Theorem 1.** *Let  $\{X_t\}$  be strictly stationary with fourth moments, and absolutely summable auto-cumulants of order 2, 3, 4. Suppose that  $\mu_3 \neq 0$ , and the polyspectra of order 2 and 3 can be factored into the form*

$$f(\underline{\lambda}) = \mu_{k+1} \prod_{j=1}^k \Psi_{k+1}(e^{-i\lambda_j}) \Psi_{k+1}(e^{i\sum_{j=1}^k \lambda_j}),$$

*yielding  $\Psi_2$  and  $\Psi_3$ . Then  $\{X_t\}$  is a quadratic process if and only if  $\Psi_3(z) \neq \Psi_2(z)$  for some  $z \in \mathbb{C}$ .*



# Theory for Quadratic Processes

**Implications:** Given some stationary process with known parameters

1. Compute auto-cumulants (analytically or by Monte Carlo)
2. Compute polyspectra for each  $k$
3. Factorize polyspectra (algorithm in paper) to obtain  $\Psi_k(z)$  and  $\mu_k$
4. Check whether  $\Psi_3(z) \neq \Psi_2(z)$



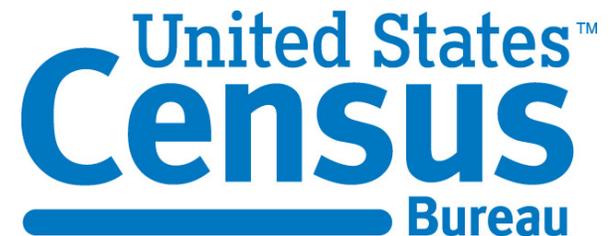
# Illustrations of Quadratic Processes

**ARCH and GARCH:** Suppose

$$X_t = \sigma_t Z_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j X_{t-j}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2.$$

It is known to be a white noise, i.e.,  $\mu_2 = 0$  and  $\Psi_2(z) \equiv 1$ .



## Illustrations of Quadratic Processes

**ARCH and GARCH:** Set  $\omega(x) = \sum_{j=1}^p \alpha_j x^j$  and  $\theta(x) = 1 - \sum_{j=1}^q \beta_j x^j$ . With  $\pi(z) = 1 - \omega(z)/\theta(z)$ , the bi-spectrum ( $k = 2$ ) is

$$f(\underline{\lambda}) = \mathbb{E}[X^3] \left( \pi(yz)^{-1} + \pi(\bar{y})^{-1} + \pi(\bar{z})^{-1} - 2 \right),$$

where  $y = e^{-i\lambda_1}$  and  $z = e^{-i\lambda_2}$ . Hence GARCH is a quadratic process when  $\mathbb{E}[X^3] \neq 0$  and  $\pi(z)$  is non-trivial.

**Open Problem:** Compute tri-spectrum!



## Illustrations of Quadratic Processes

**Hermite Processes:** Let  $\{Z_t\}$  be a mean zero, stationary Gaussian with autocovariance  $c(h)$  such that  $c(0) = 1$ . The Hermite polynomials are defined for  $k \geq 1$  ( $H_0 \equiv 1$ ) as

$$H_k(x) = \frac{(-1)^k}{\sqrt{k!}} e^{x^2/2} \partial_x^k e^{-x^2/2}.$$

For a sequence of coefficients  $\{J_k\}_{k \geq 1}$  that are square summable, let  $h(x) = \sum_{k=1}^{\infty} J_k H_k(x)$  and define a Hermite process via  $X_t = h(Z_t)$ .



## Illustrations of Quadratic Processes

**Hermite Lognormal:** Let  $h(x) = e^x - e^{c(0)/2}$ , so that  $\{X_t\}$  is a lognormal process. The auto-moments (with  $\mu = e^{c(0)/2}$ ) are

$$\gamma(k_1) = \mu^2 (\exp\{c(k_1)\} - 1)$$

$$\gamma(k_1, k_2) = \mu^3 (\exp\{c(k_1) + c(k_2) + c(k_3)\} - \exp\{c(k_1)\} - \exp\{c(k_2)\} - \exp\{c(k_3)\} + 2)$$

$$\begin{aligned} \gamma(k_1, k_2, k_3) = & \mu^4 (\exp\{c(k_1) + c(k_2) + c(k_3) + c(k_1 - k_2) + c(k_1 - k_3) + c(k_2 - k_3)\} \\ & - \exp\{c(k_1 - k_2) + c(k_1 - k_3) + c(k_2 - k_3)\} - \exp\{c(k_2) + c(k_3) + c(k_2 - k_3)\} \\ & - \exp\{c(k_1) + c(k_3) + c(k_1 - k_3)\} - \exp\{c(k_1) + c(k_2) + c(k_1 - k_2)\} \\ & + \exp\{c(k_2 - k_3)\} + \exp\{c(k_1 - k_3)\} + \exp\{c(k_1 - k_2)\} \\ & + \exp\{c(k_3)\} + \exp\{c(k_2)\} + \exp\{c(k_1)\} - 3.) \end{aligned}$$

**Verify:** Show this is a quadratic process when  $\{Z_t\}$  is a moving average.



## Illustrations of Quadratic Processes

**Hermite Quadratic:** Say only  $J_1$  and  $J_2$  are non-zero, so that the process is expressed as

$$X_t = J_1 H_1(Z_t) + J_2 H_2(Z_t) = J_1 Z_t + J_2 Z_t^2 - J_2.$$

If we divide through by  $J_2$ , we obtain a rescaled process of the form  $\alpha H_1(Z_t) + H_2(Z_t)$ , for  $\alpha = J_1/J_2$ . The auto-moments are given in terms of  $\alpha$  (complicated).



# Illustrations of Quadratic Processes

**Numerical Example:** Suppose

$$(1 - 2\rho \cos(\omega)B + \rho^2 B^2)Z_t = \epsilon_t \sim \text{WN}(0, \sigma^2),$$

with  $\rho = .99$ ,  $\omega = \pi/6$ , and  $\sigma = .1$ , a Gaussian process. Consider the Hermite Quadratic process: set  $J_k = 0$  for  $k > 2$ ,  $J_2 = 1$ , and  $J_1 = \alpha = 2$ . This results in a strong pattern of cyclical persistence.

**Prediction Results:** With a sample of size  $T = 5$  we compute the linear and quadratic MSE, obtaining .154 and .124 respectively, so that inclusion of the quadratic estimation results in a **19.6 % reduction** in MSE.



# Conclusion

## Key Advances:

- Formulation and solution of quadratic prediction problem: relies on third and fourth cumulants.
- Conditions derived under which quadratic is better than linear.
- Algorithms to compute and factor polyspectra.
- New classes of processes, including Hermite Quadratic.

Email: [tucker.s.mcelroy@census.gov](mailto:tucker.s.mcelroy@census.gov)

