Why ARMAX-GARCH Linear Models Successfully Describe Complex Nonlinear Phenomena: A Possible Explanation

Hung T. Nguyen[1,2], Vladik Kreinovich[3],
Olga Kosheleva[4], and Songsak Sriboonchitta[2]

[1]Department of Mathematical Sciences, New Mexico State University
Las Cruces, New Mexico 88003, USA, hunguyen@nmsu.edu
[2]Faculty of Economics, Chiang Mai University
Chiang Mai, Thailand, songsakecon@gmail.com
[3]Department of Computer Science, [4]Department of Teacher Education
University of Texas at El Paso, El Paso, Texas 79968, USA
vladik@utep.edu, olgak@utep.edu

1. Formulation of the Problem

- Economic and financial processes are very complex, many empirical dependencies are highly nonlinear.
- However, linear models are surprisingly efficient in predicting future values of the corresponding quantities.
- ARMAX model predicts the quantity X affected by the external quantity d:

$$X_t = \sum_{i=1}^{p} \varphi_i \cdot X_{t-i} + \sum_{i=1}^{b} \eta_i \cdot d_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \cdot \varepsilon_{t-i}.$$

- Here, $\varepsilon_t = \sigma_t \cdot z_t$, z_t is white noise with 0 mean and standard deviation 1, and σ_t follows the GARCH model:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{\ell} \beta_i \cdot \sigma_{t-i}^2 + \sum_{i=1}^{k} \alpha_i \cdot \varepsilon_{t-i}^2.$$

- In this paper, we provide a possible explanation for the empirical success of the ARMAX-GARCH models.

2. First Approximation: Closed System

- Let us start with the simplest possible model, in which we ignore all outside effects on the system.
- Such no-outside-influence systems are known as *closed systems*.
- In such a closed system, the future state X_t is uniquely determined by its previous states:
$$X_t = f(X_{t-1}, X_{t-2}, \ldots, X_{t-p}).$$
- So, to describe how to predict the state of a system, we need to describe the corresponding prediction function
$$f(x_1, \ldots, x_p).$$
- We will describe the reasonable properties of this prediction function.
- Then, we will show that these property imply that the prediction function be linear.

3. First Reasonable Property of the Prediction Function $f(x_1, \ldots, x_p)$: Continuity

- In many cases, the values X_t are only approximately known.
- E.g., the existing methods of measuring GDP or unemployment rate are approximate.
- Thus, the actual values X_t^{act} of the quantity X may be, in general, slightly different from observed values X_t.
- It is therefore reasonable to require that:
  - when we apply the prediction function to the observed (approximate) value, then
  - the prediction $f(X_{t-1}, \ldots, X_{t-p})$ should be close to the prediction $f(X_{t-1}^{\text{act}}, \ldots, X_{t-p}^{\text{act}})$ based on X_t^{act}.
- In precise terms, this means that the function $f(x_1, \ldots, x_p)$ should be *continuous.*

4. Second Reasonable Property of the Prediction Function $f(x_1, \ldots, x_p)$: Additivity

- In many practical situations, we observe a joint effect of two (or more) different subsystems $X = X^{(1)} + X^{(2)}$.
- For example, the varying price of the financial portfolio can be represented as the sum of the prices of its parts.
- In this case, we have two possible ways to predict the desired value X_t:
  - we can apply the prediction function $f(x_1, \ldots, x_p)$ to the joint values $X_{t-i} = X^{(1)}_{t-i} + X^{(2)}_{t-i}$:

$$X_t = f\left(X^{(1)}_{t-1} + X^{(2)}_{t-1}, \ldots, X^{(1)}_{t-p} + X^{(2)}_{t-p}\right);$$

  - we can apply this prediction function to both subsystems and add the predictions:

$$X_t = X^{(1)}_t + X^{(2)}_t = f\left(X^{(1)}_{t-1}, \ldots, X^{(1)}_{t-p}\right) + f\left(X^{(2)}_{t-1}, \ldots, X^{(2)}_{t-p}\right).$$

5. Additivity (cont-d)

- It makes sense to require that these two methods lead to the same prediction, i.e., that:

$$f\left(X^{(1)}_{t-1}+X^{(2)}_{t-1},\ldots,X^{(1)}_{t-p}+X^{(2)}_{t-p}\right)=$$

$$f\left(X^{(1)}_{t-1},\ldots,X^{(1)}_{t-p}\right)+f\left(X^{(2)}_{t-1},\ldots,X^{(2)}_{t-p}\right).$$

- In mathematical terms, the predictor function should be *additive*, i.e., that for all possible tuples:

$$f\left(x^{(1)}_{1}+x^{(2)}_{1},\ldots,x^{(1)}_{p}+x^{(2)}_{p}\right)=$$

$$f\left(x^{(1)}_{1},\ldots,x^{(1)}_{p}\right)+f\left(x^{(2)}_{1},\ldots,x^{(2)}_{p}\right).$$

- Every continuous additive function has the form

$$f(x_1,\ldots,x_p)=\sum_{i=1}^{p}\varphi_i\cdot x_i.$$

- Thus, we have justified the use of *linear predictors.*

6. Second Approximation: Taking External Quantities Into Account

- In practice, the desired quantity X may also be affected by some external quantity d.
- For example, the stock price may be affected by the amount of money invested in stocks.
- In this case, to predict X_t, we also need to know the values of d_t, d_{t-1}, ...:

$$X_t = f(X_{t-1}, X_{t-2}, \ldots, X_{t-p}, d_t, d_{t-1}, \ldots, d_{t-b}).$$

- Let us consider reasonable properties of the prediction function $f(x_1, \ldots, x_p, y_0, \ldots, y_b)$.
- Small changes in the inputs should lead to small changes in the prediction.
- Thus, $f(x_1, \ldots, x_p, y_0, \ldots, y_b)$ should be continuous.

7. Second Approximation (cont-d)

- The overall external effect d can be only decomposed into two components corresponding to subsystems:

$$d = d^{(1)} + d^{(2)}.$$

- E.g., $d^{(i)}$ are investments into two sectors of the stock market.

- In this case, just like in the first approximation, we have two possible ways to predict the desired value X_t:
  - we can apply the prediction function $f(x_1, \ldots, x_p, y_0, \ldots, y_b)$ to the joint values

$$X_{t-i} = X^{(1)}_{t-i} + X^{(2)}_{t-i} \text{ and } d_{t-i} = d^{(1)}_{t-i} + d^{(2)}_{t-i};$$

  - we can apply this prediction function to the both systems and add the results: $X_t = X^{(1)}_t + X^{(2)}_t$.

8. Second Approximation: Results

- It makes sense to require that these two methods lead to the same prediction, i.e., that:

$$f\left(X^{(1)}_{t-1}+X^{(2)}_{t-1},\ldots,X^{(1)}_{t-p}+X^{(2)}_{t-p},d^{(1)}_t+d^{(2)}_t,\ldots,d^{(1)}_{t-b}+d^{(2)}_{t-b}\right)=$$
$$f\left(X^{(1)}_{t-1},\ldots,X^{(1)}_{t-p},d^{(1)}_t,\ldots,d^{(1)}_{t-b}\right)+$$
$$f\left(X^{(2)}_{t-1},\ldots,X^{(2)}_{t-p},d^{(2)}_t,\ldots,d^{(2)}_{t-b}\right).$$

- Thus, the prediction function $f(x_1,\ldots,x_n,y_0,\ldots,y_b)$ should be additive.
- We already know that continuous additive functions are linear, so the predictor should be linear:

$$X_t=\sum_{i=1}^{p}\varphi_i\cdot X_{t-i}+\sum_{i=0}^{b}\eta_i\cdot d_{t-i}.$$

9. Third Approximation: Taking Random Effects into Account

- In addition to the external quantities d, the desired quantity X is also affected by many other phenomena.
- In contrast to the explicitly known d_t, we do not know the values ε_t characterizing all these phenomena.
- It is thus reasonable to consider ε_t *random effects*.
- So, to predict X_t, we also need to know ε_t, ε_{t-1}, ...:

$$X_t = f(X_{t-1}, X_{t-2}, \ldots, X_{t-p}, d_t, d_{t-1}, \ldots, d_{t-b}, \varepsilon_t, \ldots, \varepsilon_{t-q}).$$

- It is reasonable to require that the prediction function be continuous and additive:

$$f\left(X^{(1)}_{t-1} + X^{(2)}_{t-1}, \ldots, d^{(1)}_t + d^{(2)}_t, \ldots, \varepsilon^{(1)}_t + \varepsilon^{(2)}_t, \ldots\right) =$$

$$f\left(X^{(1)}_{t-1}, \ldots, d^{(1)}_t, \ldots, \varepsilon^{(1)}_t, \ldots\right) + f\left(X^{(2)}_{t-1}, \ldots, d^{(2)}_t, \ldots, \varepsilon^{(2)}_t, \ldots\right).$$

10. Third Approximation: Result

- Continuous additive functions are linear, so

$$X_t = \sum_{i=1}^{p} \varphi_i \cdot X_{t-i} + \sum_{i=0}^{b} \eta_i \cdot d_{t-i} + \sum_{i=0}^{q} \theta_i \cdot \varepsilon_{t-i}.$$

- If we take $\varepsilon'_t \stackrel{\text{def}}{=} \theta_0 \cdot \varepsilon_t$, this becomes the ARMAX formula:

$$X_t = \sum_{i=1}^{p} \varphi_i \cdot X_{t-i} + \sum_{i=0}^{b} \eta_i \cdot d_{t-i} + \varepsilon'_t + \sum_{i=1}^{q} \theta'_i \cdot \varepsilon'_{t-i}.$$

- Similar arguments lead to a multi-D version of ARMAX, in which:
  - X, d, ε' are vectors, and
  - φ_i, η_i, and θ'_i are corresponding matrices.

11. 4th Approximation: Taking Into Account that Standard Deviations σ_t Change with Time

- In general, the st. dev. σ_t changes with time.
- So, we need to predict both X_t and σ_t:

$$X_t = f(X_{t-1}, \ldots, d_t, \ldots, \varepsilon_t, \ldots, \sigma_{t-1}, \ldots);$$

$$\sigma_t = g(X_{t-1}, \ldots, d_t, \ldots, \varepsilon_t, \ldots, \sigma_{t-1}, \ldots).$$

- Let us consider reasonable properties of these prediction functions.
- It is reasonable to require that small changes in the inputs should lead to small changes in the prediction.
- Thus, the prediction functions should be continuous.
- It also makes sense to consider the case of subsystems.

12. Final Approximation (cont-d)

- There are two cases when σ of the system can be obtained from $\sigma^{(i)}$ of subsystems:
  - when $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ are independent, and
  - when $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ are strongly correlated.
- For independence case, $V = V^{(1)} + V^{(2)}$ for $V = \sigma^2$, so:

$$f'\left(X^{(1)}_{t-1} + X^{(2)}_{t-1}, \ldots, d^{(1)}_t + d^{(2)}_t, \ldots, \varepsilon^{(1)}_t + \varepsilon^{(2)}_t, \ldots, V^{(1)}_{t-1} + V^{(2)}_{t-1}, \ldots\right)$$

$$f'\left(X^{(1)}_{t-1}, \ldots, d^{(1)}_t, \ldots, \varepsilon^{(1)}_t, \ldots, V^{(1)}_{t-1}, \ldots\right) +$$

$$f'\left(X^{(2)}_{t-1}, \ldots, d^{(2)}_t, \ldots, \varepsilon^{(2)}_t, \ldots, V^{(2)}_{t-1}, \ldots\right);$$

$$g'\left(X^{(1)}_{t-1} + X^{(2)}_{t-1}, \ldots, d^{(1)}_t + d^{(2)}_t, \ldots, \varepsilon^{(1)}_t + \varepsilon^{(2)}_t, \ldots, V^{(1)}_{t-1} + V^{(2)}_{t-1}, \ldots\right)$$

$$g'\left(X^{(1)}_{t-1}, \ldots, d^{(1)}_t, \ldots, \varepsilon^{(1)}_t, \ldots, V^{(1)}_{t-1}, \ldots\right) +$$

$$g'\left(X^{(2)}_{t-1}, \ldots, d^{(2)}_t, \ldots, \varepsilon^{(2)}_t, \ldots, V^{(2)}_{t-1}, \ldots\right).$$

13. Final Approximation: Results

- Thus, both prediction f-s are additive, hence linear:

$$X_t = \sum_{i=1}^{p} \varphi_i \cdot X_{t-i} + \sum_{i=0}^{b} \eta_i \cdot d_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \cdot \varepsilon_{t-i} + \sum_{i=1}^{\ell} \beta'_i \cdot \sigma^2_{t-i};$$

$$\sigma^2_t = \sum_{i=1}^{p} \varphi'_i \cdot X_{t-i} + \sum_{i=0}^{b} \eta'_i \cdot d_{t-i} + \sum_{i=0}^{q} \theta'_i \cdot \varepsilon_{t-i} + \sum_{i=1}^{\ell} \beta_i \cdot \sigma^2_{t-i}.$$

- In the strongly correlated case, $\sigma = \sigma^{(1)} + \sigma^{(2)}$.
- Resulting additivity implies $\varphi'_i = \eta'_i = \theta'_i = 0$, so:

$$X_t = \sum_{i=1}^{p} \varphi_i \cdot X_{t-i} + \sum_{i=0}^{b} \eta_i \cdot d_{t-i} + \sum_{i=0}^{q} \theta_i \cdot \varepsilon_{t-i}; \;\; \sigma^2_t = \sum_{i=1}^{\ell} \beta_i \cdot \sigma^2_{t-i}.$$

- Thus, we have justified a (simplified version of) the ARMAX-GARCH formula.
- This version lacks α_0 and terms proport. to ε^2_{t-i}.

14. Conclusions

- In this talk, we analyzed the following problem:
  - on the one hand, economic and financial phenomena are very complex and highly *nonlinear*;
  - on the other hand, in many cases:
    - *linear* ARMAX-GARCH formulas
    - provide a very good empirical description of these complex phenomena.
- We showed that reasonable first principles lead:
  - to the ARMAX formulas and
  - to the (somewhat simplified version of) GARCH formulas.
- Thus, we have provided a reasonable explanation for the empirical success of these formulas.

15. Remaining Problem

- We only explain a simplified version of the GARCH formula.
- It is desirable to come up with a similar explanation of the full GARCH formula.
- Intuitively, the presence of additional terms proportional to ε^2 in the GARCH formula is understandable:
  - when the mean-0 random components $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ are independent,
  - the average value of their product $\varepsilon^{(1)} \cdot \varepsilon^{(2)}$ is zero.
- One can show that this makes the missing term $\sum\limits_{i=1}^{k} \alpha_i \cdot \varepsilon^2_{t-i}$ additive.
- Thus, this term is potentially derivable from our requirements.

16. Remaining Problem (cont-d)

- The term α_0 can also be intuitively explained:
  - there is usually an additional extra source of randomness
  - which constantly adds randomness to the process.
- It is desirable to transform these intuitive arguments into a precise derivation of the full GARCH formula.

17. Acknowledgments

- We acknowledge the partial support of the
  - Center of Excellence in Econometrics, Faculty of Economics,
  - Chiang Mai University, Thailand.
- This work was also supported in part by the National Science Foundation grants:
  - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence), and
  - DUE-0926721.