5-2015

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

Hung T. Nguyen
New Mexico State University - Main Campus, hunguyen@nmsu.edu

Vladik Kreinovich
University of Texas at El Paso, vladik@utep.edu

Olga Kosheleva
University of Texas at El Paso, olgak@utep.edu

Songsak Sriboonchitta
Chiang Mai University, songsakecon@gmail.com

Comments:
Technical Report: UTEP-CS-15-40

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] Department of Economics, Chiang Mai University
Chiang Mai, Thailand, songsakecon@gmail.com
[3] Department of Computer Science, University of Texas at El Paso
500 W. University, El Paso, TX 79968, USA, vladik@utep.edu
[4] University of Texas at El Paso, 500 W. University,
El Paso, TX 79968, USA, olgak@utep.edu

Abstract. Economic and financial processes are complex and highly nonlinear. However, somewhat surprisingly, linear models like ARMAX-GARCH often describe these processes reasonably well. In this paper, we provide a possible explanation for the empirical success of these models.

1 Formulation of the Problem

Economic and financial processes are very complex. It is well know that economic and financial processes are very complex. The future values of the corresponding quantities are very difficult to predict, and many empirical dependencies are highly nonlinear.

Surprising empirical success of ARMAX-GARCH models. In spite of the clearly non-linearity of the economic and financial processes, linear models are surprisingly efficient in predicting the future values of the corresponding quantities. Specifically, if we are interested in the quantity X which is affected by the external quantity d, then good predictions can often be made based on the AutoRegressive-Moving-Average model with eXogenous inputs model (ARMAX) [3, 4]:

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

for appropriate parameters φ_i, η_i, and θ_i. Here, ε_t are random variables of the type $\varepsilon_t = \sigma_t \cdot z_t$, where z_t is white noise with 0 mean and standard deviation 1, and the dynamics of standard deviations σ_t is described by the Generalized

AutoRegressive Conditional Heterosckedasticity (GARCH) model [2–4]:

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

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

Specifically, we start with simplest predictions models, in which many important aspects are ignored, and then show that by appropriately taking these aspects into account, we come up with the ARMAX-GARCH model.

2 First Approximation: Closed System

First approximation: description. Let us start with the simplest possible model, in which we ignore all outside effects on the system, be it deterministic or random. 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:

$$X_t = f(X_{t-1}, X_{t-2}, \ldots, X_{t-p}). \tag{3}$$

So, to describe how to predict the state of a system, we need to describe the corresponding prediction function $f(x_1, \ldots, x_p)$.

In the remaining part of this section, we will describe the reasonable properties of this prediction function, and then we will show that these property imply that the prediction function be linear.

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. For example, if we are interested in predicting Gross Domestic Product (GDP) or unemployment rate, we have to take into account that the existing methods of measuring these characteristics are approximate.

Thus, the actual values X_t^{act} of the quantity X may be, in general, slightly different from the 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 the actual values X_t^{act}.

In other words, if the inputs to the function $f(x_1, \ldots, x_p)$ change slightly, the output should also change slightly. In precise terms, this means that the function $f(x_1, \ldots, x_p)$ should be *continuous*.

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 a sum of the prices corresponding to two different parts of this portfolio. In this case, the desired future value X_t also consists of two components: $X_t = X_t^{(1)} + X_t^{(2)}$.

In this case, we have two possible way to predict the desired value X_t:

- first, we can come up with a prediction X_t by applying the prediction function $f(x_1, \ldots, x_p)$ to the joint values $X_{t-i} = X_{t-i}^{(1)} + X_{t-i}^{(2)}$;
- second, we can apply this prediction function to the first system, then apply it to the second subsystem, and then add the resulting predictions $X_t^{(1)}$ and $X_t^{(2)}$ to come up with the joint prediction $X_t = X_t^{(1)} + X_t^{(2)}$.

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

$$f(X_{t-1}^{(1)}+X_{t-1}^{(2)}, \ldots, X_{t-p}^{(1)}+X_{t-p}^{(2)}) = f(X_{t-1}^{(1)}, \ldots, X_{t-p}^{(1)})+f(X_{t-1}^{(2)}, \ldots, X_{t-p}^{(2)}). \quad (4)$$

In mathematical terms, this means that the predictor function should be *additive*.

Conclusion: we must consider linear predictors. It is known (see, e.g., [1, 5]) that every continuous additive function is a homogeneous linear function, i.e., it has the form

$$f(x_1, \ldots, x_p) = \sum_{i=1}^{p} \varphi_i \cdot x_i \quad (5)$$

for some values φ_i. Thus, we must consider *linear predictors*

$$X_t = \sum_{i=1}^{p} \varphi_i \cdot X_{t-i}. \quad (6)$$

3 Second Approximation: Taking External Quantities Into Account

Second approximation: description. To get a more adequate description of the economic system, let us take into account that 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 determine the future state X_t, we need to know not only the previous states of the system X_{t-1}, X_{t-2}, ..., but also the corresponding values of the external quantity d_t, d_{t-1}, ... Thus, the general prediction formula now takes the following form:

$$X_t = f(X_{t-1}, X_{t-2}, \ldots, X_{t-p}, d_t, d_{t-1}, \ldots, d_{t-b}). \quad (7)$$

So, to describe how to predict the state of a system, we need to describe the corresponding prediction function $f(x_1, \ldots, x_p, y_0, \ldots, y_b)$. Let us consider reasonable properties of this prediction function.

First reasonable property of the prediction function $f(x_1, \ldots, x_p, y_0, \ldots, y_b)$: continuity. Similarly to the previous case, we can conclude that small changes in the inputs should lead to small changes in

the prediction. Thus, the prediction function $f(x_1, \ldots, x_p, y_0, \ldots, y_b)$ should be continuous.

Second reasonable property of the prediction function $f(x_1, \ldots, x_p, y_0, \ldots, y_b)$: additivity. As we have mentioned earlier, in many practical situations, we observe a joint effect of two (or more) different subsystems $X = X^{(1)} + X^{(2)}$. In this case, the overall external effect d can be only decomposed into two components $d = d^{(1)} + d^{(2)}$: e.g., investments into two sectors of the stock market.

In this case, just like in the first approximation, we have two possible way to predict the desired value X_t:

- first, we can come up with a prediction X_t by applying the prediction function $f(x_1, \ldots, x_p, y_0, \ldots, y_b)$ to the joint values $X_{t-i} = X^{(1)}_{t-i} + X^{(2)}_{t-i}$ and $d_{t-i} = d^{(1)}_{t-i} + d^{(2)}_{t-i}$;
- second, we can apply this prediction function to the first system, then apply it to the second subsystem, and then add the resulting predictions $X^{(1)}_t$ and $X^{(2)}_t$ to come up with the joint prediction $X_t = X^{(1)}_t + X^{(2)}_t$.

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

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

$$f(X^{(1)}_{t-1}, \ldots, X^{(1)}_{t-p}, d^{(1)}_t, \ldots, d^{(1)}_{t-b}) + f(X^{(2)}_{t-1}, \ldots, X^{(2)}_{t-p}, d^{(2)}_t, \ldots, d^{(2)}_{t-b}). \quad (8)$$

Thus, the prediction function $f(x_1, \ldots, x_n, y_0, \ldots, y_b)$ should be additive.

Conclusion: we must consider linear predictors. Since every continuous additive function is a homogeneous linear function, we have

$$f(x_1, \ldots, x_p, y_0, \ldots, y_b) = \sum_{i=1}^{p} \varphi_i \cdot x_i + \sum_{i=0}^{b} \eta_i \cdot y_i \quad (9)$$

for some values φ_i and η_i; thus:

$$X_t = \sum_{i=1}^{p} \varphi_i \cdot X_{t-i} + \sum_{i=0}^{b} \eta_i \cdot d_{t-i}. \quad (10)$$

4 Third Approximation: Taking Random Effects into Account

Description of the model. In addition to the external quantities d, the desired quantity X is also affected by many other phenomena. In contrast to the explicitly known quantity d, we do not know the values characterizing all these phenomena, so it is reasonable to consider them *random effects*. Let us denote the random effect generated at moment t by ε_t.

In this case, to determine the future state X_t, we need to know not only the previous states of the system X_{t-1}, X_{t-2}, ..., and the corresponding values of the external quantity d_t, d_{t-1}, ..., we also need to know the values of these random effects ε_t, ε_{t-1}, ... Thus, the general prediction formula now takes the form

$$X_t = f(X_{t-1}, X_{t-2}, \ldots, X_{t-p}, d_t, d_{t-1}, \ldots, d_{t-b}, \varepsilon_t, \ldots, \varepsilon_{t-q}). \tag{11}$$

So, to describe how to predict the state of a system, we need to describe the corresponding prediction function $f(x_1, \ldots, x_p, y_0, \ldots, y_b, z_0, \ldots, z_q)$. Let us consider reasonable properties of this prediction function.

First reasonable property of the prediction function $f(x_1, \ldots, x_p, y_0, \ldots, y_b, z_0, \ldots, z_q)$: continuity. Similarly to the previous cases, we can conclude that small changes in the inputs should lead to small changes in the prediction. Thus, the prediction function $f(x_1, \ldots, x_p, y_0, \ldots, y_b, z_0, \ldots, z_q)$ should be continuous.

Second reasonable property of the prediction function $f(x_1, \ldots, x_p, y_0, \ldots, y_b, z_0, \ldots, z_q)$: additivity. As we have mentioned earlier, in many practical situations, we observe a joint effect of two (or more) different subsystems $X = X^{(1)} + X^{(2)}$. In this case, the overall external effect d can be only decomposed into two components $d = d^{(1)} + d^{(2)}$, and the random effects can also be decomposed into effects affecting the two subsystems: $\varepsilon = \varepsilon^{(1)} + \varepsilon^{(2)}$.

In this case, just like in the first two approximations, we have two possible way to predict the desired value X_t:

- first, we can come up with a prediction X_t by applying the prediction function $f(x_1, \ldots, x_p, y_0, \ldots, y_b)$ to the joint values $X_{t-i} = X^{(1)}_{t-i} + X^{(2)}_{t-i}$, $d_{t-i} = d^{(1)}_{t-i} + d^{(2)}_{t-i}$, and $\varepsilon_{t-i} = \varepsilon^{(1)}_{t-i} + \varepsilon^{(2)}_{t-i}$;
- second, we can apply this prediction function to the first system, then apply it to the second subsystem, and then add the resulting predictions $X^{(1)}_t$ and $X^{(2)}_t$ to come up with the joint prediction $X_t = X^{(1)}_t + X^{(2)}_t$.

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

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

$$f(X^{(1)}_{t-1}, \ldots, d^{(1)}_t, \ldots, \varepsilon^{(1)}_t, \ldots) + f(X^{(2)}_{t-1}, \ldots, d^{(2)}_t, \ldots, \varepsilon^{(2)}_t, \ldots). \tag{12}$$

Thus, the prediction function $f(x_1, \ldots, y_0, \ldots, z_0, \ldots)$ should be additive.

Conclusion: we must consider linear predictors. Since every continuous additive function is a homogeneous linear function, we have

$$f(x_1, \ldots, x_p, y_0, \ldots, y_b, z_0, \ldots, z_q) = \sum_{i=1}^{p} \varphi_i \cdot x_i + \sum_{i=0}^{b} \eta_i \cdot y_i + \sum_{i=0}^{q} \theta_i \cdot z_i \tag{13}$$

for some values φ_i, η_i, and θ_i; thus:

$$X_t = \sum_{i=1}^{p} \varphi_i \cdot X_{t-i} + \sum_{i=0}^{b} \eta_i \cdot d_{t-i} + \sum_{i=0}^{q} \theta_i \cdot \varepsilon_{t-i}. \tag{14}$$

Deriving the original ARMAX formula (1). The formula (14) is almost identical to the ARMAX formula (1), the only difference is that in our formula (14), the value ε_t is multiplied by a coefficient θ_0, while in the ARMAX formula (1), this coefficient is equal to 1.

To derive the formula (1), let us first comment that it is highly improbable that the random quantity ε_t does not have any effect on the current value X_t of the desired quantity; thus, the parameter θ_0 describing this dependence should be non-zero.

Now, to describe the random effects, instead of the original values ε, we can consider the new values $\varepsilon' \stackrel{\text{def}}{=} \theta_0 \cdot \varepsilon$. In terms of thus re-scaled random effects, we have $\varepsilon = \dfrac{1}{\theta_0} \cdot \varepsilon'$. Thus, the corresponding linear combination of random terms takes the form

$$\sum_{i=0}^{q} \theta_i \cdot \varepsilon_{t-i} = \theta_0 \cdot \varepsilon_t + \sum_{i=1}^{q} \theta_i \cdot \varepsilon_{t-i} = \varepsilon'_0 + \sum_{i=1}^{q} \theta_i \cdot \frac{1}{\theta_0} \cdot \varepsilon'_{t-i}, \tag{15}$$

i.e., the form

$$\sum_{i=0}^{q} \theta_i \cdot \varepsilon_{t-i} = \varepsilon'_0 + \sum_{i=1}^{q} \theta'_i \cdot \varepsilon'_{t-i}, \tag{16}$$

where we denoted $\theta'_i \stackrel{\text{def}}{=} \theta_i \cdot \dfrac{1}{\theta_0}$.

Substituting the formula (16) into the expression (14), we get the desired ARMAX formula:

$$X_t = \sum_{i=1}^{p} \varphi_i \cdot X_{t-i} + \sum_{i=0}^{b} \eta_i \cdot d_{t-i} + \varepsilon'_0 + \sum_{i=1}^{q} \theta'_i \cdot \varepsilon'_{t-i}. \tag{17}$$

Similar arguments can be used to explain formulas of Vector ARMAX (VARMAX). Similar arguments lead to a multi-D (version) of the formula (17), in which X, d, ε are vectors, and φ_i, η_i, and θ'_i are corresponding matrices.

5 Fourth Approximation: Taking Into Account that Standard Deviations Change with Time

Description of the model. In the previous sections, we described how the desired quantity X changes with time. In the previous section, we showed how to take into account the random effects $\varepsilon_t = \sigma_t \cdot z_t$ that affect our system.

To complete the description of the system's dynamics, it is necessary to supplement this description with a description of how the corresponding standard deviation σ_t changes with time. So, now, instead of simply predicting the values X_t, we need to predict both the values X_t and the values σ_t.

To predict both values X_t and σ_t, we can use:

- the previous states of the system X_{t-1}, X_{t-2}, ...,
- the corresponding values of the external quantity d_t, d_{t-1}, ...,
- the values of these random effects ε_t, ε_{t-1}, ..., and
- the previous values of the standard deviation σ_{t-1}, σ_{t-2}, ...

Thus, the general prediction formulas now take the form

$$X_t = f(X_{t-1}, \ldots, d_t, \ldots, \varepsilon_t, \ldots, \sigma_{t-1}, \ldots); \tag{18}$$

$$\sigma_t = g(X_{t-1}, \ldots, d_t, \ldots, \varepsilon_t, \ldots, \sigma_{t-1}, \ldots). \tag{19}$$

So, to describe how to predict the state of a system, we need to describe the corresponding prediction functions $f(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$ and $g(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$. Let us consider reasonable properties of this prediction function.

First reasonable property of the prediction functions $f(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$ and $g(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$: continuity. Similarly to the previous cases, we can conclude that small changes in the inputs should lead to small changes in the prediction. Thus, the prediction functions $f(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$ and $g(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$ should be continuous.

Second reasonable property of the prediction functions $f(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$ and $g(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$: independence-based additivity. As we have mentioned earlier, in many practical situations, we observe a joint effect of two (or more) different subsystems $X = X^{(1)} + X^{(2)}$. In this case, the overall external effect d can be only decomposed into two components $d = d^{(1)} + d^{(2)}$, and the random effects can also be decomposed into effects affecting the two subsystems: $\varepsilon = \varepsilon^{(1)} + \varepsilon^{(2)}$.

In our final model, we also need to take into the standard deviations σ; so, we need to know how to compute the standard deviation σ of the sum of two random variables based on their standard deviations $\sigma^{(1)}$ and $\sigma^{(2)}$ of the two components. In general, this is not possible: to know the standard deviation σ of the sum, we need to know not only the standard deviations $\sigma^{(1)}$ and $\sigma^{(2)}$, we also need to know the correlation between the random variables $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$.

However, there are two reasonable cases when σ can be computed based on $\sigma^{(1)}$ and $\sigma^{(2)}$:

- the case when the random variables $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ are independent, and
- the case when the random variables $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ are strongly correlated.

In this section, we will consider both cases; in this subsection, we will consider the first case.

It is known that the variance $V = \sigma^2$ of the sum of two independent random variables is equal to the sum of the variances, so $V = V^{(1)} + V^{(2)}$. To utilize this property, it makes sense to use the variance V instead of standard deviation. In terms of variance, the predictions formulas take the form

$$X_t = f'(X_{t-1}, \ldots, d_t, \ldots, \varepsilon_t, \ldots, V_{t-1}, \ldots); \tag{20}$$

$$V_t = g'(X_{t-1}, \ldots, d_t, \ldots, \varepsilon_t, \ldots, V_{t-1}, \ldots), \tag{21}$$

for appropriate functions $f'(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$ and $g'(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$.

In this case, just like in the first three approximations, we have two possible way to predict the desired values X_t and V_t:

- first, we can come up with predictions X_t and V_t by applying the prediction functions $f'(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$ and $g'(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$ to the joint values $X_{t-i} = X_{t-i}^{(1)} + X_{t-i}^{(2)}$, $d_{t-i} = d_{t-i}^{(1)} + d_{t-i}^{(2)}$, $\varepsilon_{t-i} = \varepsilon_{t-i}^{(1)} + \varepsilon_{t-i}^{(2)}$, and $V_{t-i} = V_{t-i}^{(1)} + V_{t-i}^{(2)}$;
- second, we can apply these prediction functions to the first system, then apply them to the second subsystem, and then add the resulting predictions $X_t^{(i)}$ and $V_t^{(i)}$ to come up with the joint predictions $X_t = X_t^{(1)} + X_t^{(2)}$ and $V_t = V_t^{(1)} + V_t^{(2)}$.

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

$$f'(X_{t-1}^{(1)} + X_{t-1}^{(2)}, \ldots, d_t^{(1)} + d_t^{(2)}, \ldots, \varepsilon_t^{(1)} + \varepsilon_t^{(2)}, \ldots, V_{t-1}^{(1)} + V_{t-1}^{(2)}, \ldots) =$$

$$f'(X_{t-1}^{(1)}, \ldots, d_t^{(1)}, \ldots, \varepsilon_t^{(1)}, \ldots, V_{t-1}^{(1)}, \ldots) + \tag{22}$$

$$f'(X_{t-1}^{(2)}, \ldots, d_t^{(2)}, \ldots, \varepsilon_t^{(2)}, \ldots, V_{t-1}^{(2)}, \ldots);$$

$$g'(X_{t-1}^{(1)} + X_{t-1}^{(2)}, \ldots, d_t^{(1)} + d_t^{(2)}, \ldots, \varepsilon_t^{(1)} + \varepsilon_t^{(2)}, \ldots, V_{t-1}^{(1)} + V_{t-1}^{(2)}, \ldots) =$$

$$g'(X_{t-1}^{(1)}, \ldots, d_t^{(1)}, \ldots, \varepsilon_t^{(1)}, \ldots, V_{t-1}^{(1)}, \ldots) + \tag{23}$$

$$g'(X_{t-1}^{(2)}, \ldots, d_t^{(2)}, \ldots, \varepsilon_t^{(2)}, \ldots, V_{t-1}^{(2)}, \ldots).$$

Thus, both prediction functions $f'(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$ and $g'(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots)$ should be additive.

Since every continuous additive function is a homogeneous linear function, we have

$$f'(x_1, \ldots, y_0, \ldots, z_0, \ldots, t_1, \ldots) = \sum_{i=1}^{p} \varphi_i \cdot x_i + \sum_{i=0}^{b} \eta_i \cdot y_i + \sum_{i=0}^{q} \theta_i \cdot z_i + \sum_{i=1}^{\ell} \beta_i' \cdot t_i \tag{24}$$

and

$$g'(x_1,\dots,y_0,\dots,z_0,\dots,t_1,\dots)=\sum_{i=1}^{p}\varphi_i'\cdot x_i+\sum_{i=0}^{b}\eta_i'\cdot y_i+\sum_{i=0}^{q}\theta_i'\cdot z_i+\sum_{i=1}^{\ell}\beta_i\cdot t_i. \quad (25)$$

for some values φ_i, φ_i', η_i, η_i', θ_i, θ_i', β_i, and β_i'.

Similarly to the previous case, without losing generality, we can take $\theta_0=1$. Thus:

$$X_t=\sum_{i=1}^{p}\varphi_i\cdot X_{t-i}+\sum_{i=0}^{b}\eta_i\cdot d_{t-i}+\varepsilon_t+\sum_{i=1}^{q}\theta_i\cdot\varepsilon_{t-i}+\sum_{i=1}^{\ell}\beta_i'\cdot\sigma_{t-i}^2; \quad (26)$$

$$\sigma_t^2=\sum_{i=1}^{p}\varphi_i'\cdot X_{t-i}+\sum_{i=0}^{b}\eta_i'\cdot d_{t-i}+\sum_{i=0}^{q}\theta_i'\cdot\varepsilon_{t-i}+\sum_{i=1}^{\ell}\beta_i\cdot\sigma_{t-i}^2. \quad (27)$$

Third reasonable property of the prediction functions: dependence-based additivity. In the previous subsection, we considered the case when the random variables corresponding to two subsystems are independent. This makes sense, e.g., when we divide the stocks into groups by industry, so that different random factors affect the stocks from different groups. Alternatively, we can divide the stocks from the same industry by geographic location of the corresponding company, in which case the random factors affecting both types of stocks are strongly positively correlated.

For such random quantities, the standard deviation of the sum is equal to the sum of standard deviations $\sigma=\sigma^{(1)}+\sigma^{(2)}$. In this case, we can similarly use two different ways to predicting X_t and σ_t:

- first, we can come up with predictions X_t and V_t by applying the prediction formulas (26) and (27) to the joint values $X_{t-i}=X_{t-i}^{(1)}+X_{t-i}^{(2)}$, $d_{t-i}=d_{t-i}^{(1)}+d_{t-i}^{(2)}$, $\varepsilon_{t-i}=\varepsilon_{t-i}^{(1)}+\varepsilon_{t-i}^{(2)}$, and $\sigma_{t-i}=\sigma_{t-i}^{(1)}+\sigma_{t-i}^{(2)}$;
- second, we can apply these prediction formulas to the first system, then apply them to the second subsystem, and then add the resulting predictions $X_t^{(i)}$ and $V_t^{(i)}$ to come up with the joint predictions $X_t=X_t^{(1)}+X_t^{(2)}$ and $\sigma_t=\sigma_t^{(1)}+\sigma_t^{(2)}$.

It makes sense to require that these two methods lead to the same prediction.

Let us use the dependence-based additivity property. Let us apply the dependence-based additivity property to the case when the two combined subsystems are identical, i.e., when $X_{t-i}^{(1)}=X_{t-i}^{(2)}$, $d_{t-i}^{(1)}=d_{t-i}^{(2)}$, $\varepsilon_{t-i}^{(1)}=\varepsilon_{t-i}^{(2)}$, and $\sigma_{t-i}^{(1)}=\sigma_{t-i}^{(2)}$. In this case, $X_{t-i}^{(1)}=X_{t-i}^{(2)}=0.5\cdot X_{t-i}$, $d_{t-i}^{(1)}=d_{t-i}^{(2)}=0.5\cdot d_{t-i}$, $\varepsilon_{t-i}^{(1)}=\varepsilon_{t-i}^{(2)}=0.5\cdot\varepsilon_{t-i}$, and $\sigma_{t-i}^{(1)}=\sigma_{t-i}^{(2)}=0.5\cdot\sigma_{t-i}$. Substituting these values $X_{t-i}^{(1)}$, $d_{t-i}^{(1)}$, $\varepsilon_{t-i}^{(1)}$, and $\sigma_{t-i}^{(1)}$ into the formula (26), we conclude that

$$X_t^{(1)}=\sum_{i=1}^{p}\varphi_i\cdot 0.5\cdot X_{t-i}+\sum_{i=0}^{b}\eta_i\cdot 0.5\cdot d_{t-i}+0.5\cdot\varepsilon_t+$$

$$\sum_{i=1}^{q} \theta_i \cdot 0.5 \cdot \varepsilon_{t-i} + \sum_{i=1}^{\ell} \beta_i' \cdot 0.25 \cdot \sigma_{t-i}^2. \tag{29}$$

Thus, for $X_t = X_t^{(1)} + X_t^{(2)} = 2X^{(1)}(t)$, we get

$$X_t = \sum_{i=1}^{p} \varphi_i \cdot X_{t-i} + \sum_{i=0}^{b} \eta_i \cdot d_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \cdot \varepsilon_{t-i} + 0.5 \cdot \sum_{i=1}^{\ell} \beta_i' \cdot \sigma_{t-i}^2. \tag{30}$$

We require that the prediction (26) based on the sums should be equal to the sum (30) of the predictions based on the individual subsystems. Thus, the right-hand sides of the expressions (26) and (30) should be equal for all possible values of the input quantities X_{t-i}, d_{t-i}, ε_{t-i}, and σ_{t-i}. By comparing these right-hand sides, we see that this is possible only if $\beta_i' = 0$.

Similarly, substituting the values $X_{t-i}^{(1)} = 0.5 \cdot X_{t-i}$, $d_{t-i}^{(1)} = 0.5 \cdot d_{t-i}$, $\varepsilon_{t-i}^{(1)} = 0.5 \cdot \varepsilon_{t-i}$, and $\sigma_{t-i}^{(1)} = 0.5 \cdot \sigma_{t-i}$ into the formula (27), we conclude that

$$\left(\sigma_t^{(1)}\right)^2 = \sum_{i=1}^{p} \varphi_i' \cdot 0.5 \cdot X_{t-i} + \sum_{i=0}^{b} \eta_i' \cdot 0.5 \cdot d_{t-i} +$$

$$\sum_{i=0}^{q} \theta_i' \cdot 0.5 \cdot \varepsilon_{t-i} + \sum_{i=1}^{\ell} \beta_i \cdot 0.25 \cdot \sigma_{t-i}^2. \tag{31}$$

Thus, for $\sigma_t^2 = \left(2\sigma_t^{(1)}\right)^2 = 4 \cdot \left(\sigma_t^{(1)}\right)^2$, we get

$$\sigma_t^2 = 2 \cdot \sum_{i=1}^{p} \varphi_i' \cdot X_{t-i} + 2 \cdot \sum_{i=0}^{b} \eta_i' \cdot d_{t-i} + 2 \cdot \sum_{i=0}^{q} \theta_i' \cdot \varepsilon_{t-i} + \sum_{i=1}^{\ell} \beta_i \cdot \sigma_{t-i}^2. \tag{32}$$

We require that the prediction (27) based on the sums should be equal to the sum (32) of the predictions based on the individual subsystems. Thus, the right-hand sides of the expressions (27) and (32) should be equal for all possible values of the input quantities X_{t-i}, d_{t-i}, ε_{t-i}, and σ_{t-i}. By comparing these right-hand sides, we see that this is possible only if $\varphi_i' = 0$, $\eta_i' = 0$, and $\theta_i' = 0$.

Conclusion. Thus, the formulas (26) and (27) take the following form:

$$X_t = \sum_{i=1}^{p} \varphi_i \cdot X_{t-i} + \sum_{i=0}^{b} \eta_i \cdot d_{t-i} + \sum_{i=0}^{q} \theta_i \cdot \varepsilon_{t-i}; \tag{33}$$

$$\sigma_t^2 = \sum_{i=1}^{\ell} \beta_i \cdot \sigma_{t-i}^2. \tag{34}$$

Relation to the ARMAX-GRARCH formula. We can see that the formula (33) is exactly the ARMAX formula, and that the formula (34) is a simplified

version of the GARCH formula (our formula lack a constant term α_0 and the terms proportional to ε_{t-i}^2).

We have derived these empirically successful formulas from first principles. Thus, we indeed provide a reasonable explanation for the empirical success of these formulas.

6 Conclusions and Future Work

Conclusions. In this paper, 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.

Specifically, 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.

Remaining problem. While our approach explains the ARMAX formula, it provides only a partial explanation of the GARCH formula: namely, we only explain a simplified version of the GARCH formula (2). It is desirable to come up with a similar explanation of the full formula (2).

Intuitively, the presence of additional terms proportional to ε^2 in the formula (2) is understandable. Indeed, 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. Let us show that this makes the missing term $\sum\limits_{i=1}^{k} \alpha_i \cdot \varepsilon_{t-i}^2$ additive – and thus, derivable from our requirements. Indeed, we have

$$\sum_{i=1}^{k} \alpha_i \cdot \left(\varepsilon_{t-i}^{(1)} + \varepsilon_{t-i}^{(2)}\right)^2 =$$

$$\sum_{i=1}^{k} \alpha_i \cdot \left(\varepsilon_{t-i}^{(1)}\right)^2 + \sum_{i=1}^{k} \alpha_i \cdot \left(\varepsilon_{t-i}^{(2)}\right)^2 + 2\sum_{i=1}^{k} \alpha_i \cdot \left(\varepsilon_{t-i}^{(1)} \cdot \varepsilon_{t-i}^{(2)}\right).$$

Here, the last term – the average value of the product $\varepsilon^{(1)} \cdot \varepsilon^{(2)}$ – is practically 0:

$$\sum_{i=1}^{k} \alpha_i \cdot \left(\varepsilon_{t-i}^{(1)} \cdot \varepsilon_{t-i}^{(2)}\right) \approx 0,$$

so we indeed have independence-based additivity:

$$\sum_{i=1}^{k} \alpha_i \cdot \left(\varepsilon_{t-i}^{(1)} + \varepsilon_{t-i}^{(2)}\right)^2 \approx \sum_{i=1}^{k} \alpha_i \cdot \left(\varepsilon_{t-i}^{(1)}\right)^2 + \sum_{i=1}^{k} \alpha_i \cdot \left(\varepsilon_{t-i}^{(2)}\right)^2.$$

The term α_0 can also be intuitively explained: since 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 GARCH formula (2).

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.

References

1. J. Aczél and J. Dhombres, *Functional Equations in Several Variables*, Cambridge University Press, 2008.
2. T. Bollerslev, "Generalized Autoregressive Conditional Heteroskedasticitiy", *Journal of Econometrics*, 1986, Vol. 31, No. 3, pp. 307–327.
3. P. J. Brockwell and R. A. Davis, *Time Series: Theories and Methods*, Springer Verlag, New York, 2009.
4. W. Enders, *Applied Econometric Time Series*, Wiley, New York, 2014.
5. V. Kreinovich, H. T. Nguyen, and S. Sriboonchitta, "Prediction in econometrics: towards mathematical justification of simple (and successful) heuristics", *International Journal of Intelligent Technologies and Applied Statistics (IJITAS)*, 2012, Vol. 5, No. 4, pp. 443–460.