Temporal Aggregation for ARMA Model in an Application Perspective
Ze-hui JIN, Ying-ying HU, Jia-liang HUA and Zeng-min WANG
Beijing University of Posts and Telecommunications, China
*Corresponding author

Keywords: ARMA, Temporal aggregation, Monte Carlo, Model optimization.

Abstract. This article perfects ARMA model explanations and discusses regularity results of ARMA model in temporal aggregation. Monte Carlo simulation has been applied to verify accuracy of the theory. And we also put forward possible model applications, that is, integrating shocks to better prediction model in financial markets.

Introduction

Shocks are frequently witnessed by the financial market, to efficiently and timely integrate these unexpected fluctuations into the forecasting model will benefit players for huge profit or less loss. People modify their models according to a certain time unit—for example, players will update their stock forecasting model after every trading day. Suppose that, if the shock happens at the mid of the day, then players have to suffer that afternoon since they couldn’t integrate this shock into the model immediately so they are not able to make moves according to quantitative analysis but intuitions.

Temporal aggregation accounts for this phenomenon, that is, the frequency of data collection is lower than the frequency at which data are generated, thus part of items in the original random process $x = \{x_t\}_{t=0}^{\infty}$ are missing, only items after aggregation $X = \{X_{\tau}\}_{\tau=0}^{\infty}$ can be observed, where $t$ is the original frequency and $\tau$ is the after-aggregation frequency, $X$ is a specific function of $x$ determined by the aggregation scheme (Marcellino, 1999). In the above case, trading data are generated at every minute while people only gather them at every day. Temporal aggregation model is a great method to fix this problem by ‘transferring’ time unit to integrate every single hour data or even every minute data (not realistic though) into the daily model for optimization. So once data is available, players could improve their model to better its forecasting ability.

The study of temporal aggregation was started by Amemiya and Wu (1972), who first proposed the view that the structure of AR model remains unchanged after temporal aggregation. Brewer (1973) derived ARMAX model with exogenous variable $x$ to further improve theories. William (1978) stated that seasonality has to be considered when the temporal aggregation frequency is greater than the nature frequency of data. Weiss (1994) discussed flow and stock variables separately for ARIMA models. Seater (1995) explored the periodic variation of the time series before and after temporal aggregation, and found that the monthly and quarterly data have lower frequency fluctuations and the annual data completely loses the periodic fluctuation; in other words, there exists information loss. The discussion paper of Andrea Silvestrini and David Veredas in 2005 reviewed temporal aggregation of ARIMA, ARMA-GARCH, VARMA and GARCH models, and it was published in 2008 in an official version. Temporal aggregation has been discussed for over 30 years, but most attention has been focused on the theoretical level, little has been done to practical applications.

Temporal aggregation model owns huge potential, few people realized that though. Our essay tries to emphasize this point and correlate it with practice in a new perspective. In section 2, we follow Andrea Silvestrini and David Veredas’ research to perfect ARMA model explanations by illustrating a more complex example and discussing regularity results. In section 3, we verify the model using Monte Carlo simulation, strongly proving that ARMA aggregation model has practical value.
ARMA Model

Model Derivation

In this section, we will focus on the stock case $y_t^* = y_{kt}$, where $k$ is the aggregation frequency. We also mention flow case at the end of this section.

The ARMA model is defined as

$$\phi(L) \cdot y_t = \theta(L) \cdot \epsilon_t.$$  \hspace{1cm} (1)

Where $t = 0, 1, 2, \cdots$, $\phi(L)$ and $\theta(L)$ are both polynomials of lag operator $L$, $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$, $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$, and $\epsilon_t \sim N(0, \sigma^2)$. We assume $\phi(L)$ and $\theta(L)$ have distinct roots that lie outside the unit circle. Let $\delta_j (j = 1, 2, \cdots p)$ be the inverted roots of $\phi(L)$ which are inside the unit circle, then $\phi(L)$ can be expressed as

$$\phi(L) = \prod_{j=1}^{p} \left(1 - \delta_j L \right).$$

We define the model after temporal aggregation as

$$\beta(B) \cdot y_T^* = \eta(B) \cdot \epsilon_T^*.$$  \hspace{1cm} (2)

Where $T = 0, k, 2k, \cdots$, $\beta(B)$ and $\eta(B)$ are both polynomials of lag operator $B = L^k$, $\beta(B) = 1 - \beta_1 B - \cdots - \beta_c B^c$. $\eta(B) = 1 + \eta_1 B + \cdots + \eta_r B^r$, $y_T^*$ is the dependent variable after aggregation and $\epsilon_T^*$ is the white noise after aggregation.

In this case, $\phi(L) \cdot y_t = \theta(L) \cdot \epsilon_t$ model represents model for data gathered per minute, the time unit is short, the information amount is great but the fluctuation is too huge for prediction. While $\beta(B) \cdot y_T^* = \eta(B) \cdot \epsilon_T^*$ model represents daily model which is perfect for reference.

We use polynomial $T(L) = 1 + t_1 L + t_2 L^2 + \cdots + t_h L^h$ to connect these two models, so that once $\phi(L) \cdot y_t = \theta(L) \cdot \epsilon_t$ is updated, $\beta(B) \cdot y_T^* = \eta(B) \cdot \epsilon_T^*$ can be adjusted as well.

$$T(L) \cdot \phi(L) \cdot y_t = \beta(B) \cdot y_T^*.$$  \hspace{1cm} (3)

$$T(L) \cdot \theta(L) \cdot \epsilon_t = \eta(B) \cdot \epsilon_T^*.$$  \hspace{1cm} (4)

According to the first formula above, we have

$$(1 + t_1 L + t_2 L^2 + \cdots + t_h L^h) \cdot \left(1 - \phi_1 L - \cdots - \phi_p L^p\right) \cdot y_t$$

$$= (1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_c B^c) \cdot y_T^*.$$  \hspace{1cm} (5)

Brewer (1973) indicated the matrix form of $T(L)$, thus $T(L) \cdot \phi(L) = \beta(B)$ can be expressed as

$$\begin{bmatrix}
1 \\
t_1 \\
t_2 \\
\vdots \\
t_p \\
t_h \\
0 \\
0 \\
0
\end{bmatrix}
\begin{bmatrix}
\cdots \\
\cdots \\
\cdots \\
1 \\
0 \\
\phi_1 \\
0 \\
0
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_2 \\
\beta_c
\end{bmatrix}.$$  \hspace{1cm} (6)

Note that

1) Matrix $T(L)$ is comprised of coefficients of lag operator $L$, the first line represents $L^0$, the second line is $t_1 L^1 + L^0$, the third line is $t_2 L^2 + t_1 L^1 + L^0$, and so forth. Because the highest power of $T(L)$ is $h$, so the $h+1$ line represents $t_h L^h + t_{h-1} L^{h-1} + \cdots + t_{h-p} L^{h-p}$. And $(-\phi_1)$ is the coefficient of $L^1$ in $\phi(L)$. Only those $L^i (i = 0, 1, 2, \cdots)$ have non-zero coefficients in $\beta(B)$.

2) Because the first item of $T(L)$ is 1, and the first item of $\phi(L)$ is 1, so the first item of
\( \beta(B) \) is 1. Similarly, the last item of \( T(L) \), \( L^h \), has non-zero coefficient \( t_h \), the last item of \( \phi(L) \), \( L^p \), has non-zero coefficient \((-\phi_p)\), so the last item of \( \beta(B) \), \( L^{h+p} \), has non-zero coefficient \( \beta_c \).

3) Matrix \( T(L) \) is \((p + h + 1) \times (p + 1)\), \( \phi(L) \) is \((p + 1) \times 1\), \( \beta(B) \) is \((1 + ck) \times 1\), where \( p \) is known, \( h \) and \( c \) can be calculated by constriction equations. From the view of matrix multiplication, number of rows of \( T(L) \) must be equal with that of \( \beta(B) \), thus the first equation is \( p + h + 1 = 1 + ck \). From the view of matrix solutions, the number of unknown coefficients \( h \) in \( T(L) \) must match with \( c(k - 1) \) conditions in \( \beta(B) \), thus the second equation is \( h = c(k - 1) \). From these two equations we can yield that \( c = p \), \( h = p(k - 1) \).

\( T(L) \) can be calculated through the above matrix. In this case, we give the qualifying \( T(L) \) expression without proof:

\[
T(L) = \prod_{j=1}^{p} \frac{1-\delta_j^k L^j}{1-\delta_j L} = \prod_{j=1}^{p} (\sum_{s=0}^{k-1} \delta_s^j L^s).
\] (7)

Since the polynomial \( T(L) \) is given based on the AR\((P)\) model, so the aggregation of AR\((P)\) model is easy to operate.

\[
T(L) \cdot \phi(L) \cdot y_t = \prod_{j=1}^{p} \frac{1-\delta_j^k L^j}{1-\delta_j L} \cdot \prod_{j=1}^{p} (1 - \delta_j L) \cdot y_t = \prod_{j=1}^{p} (1 - \delta_j^k L^j) \cdot y_t = \prod_{j=1}^{p} (1 - \delta_j^k B) \cdot y_t^*.
\] (8)

By contrast, the operation of MA(q) is quite tough, which has to be solved by an equation system:

\[
T(L) \cdot \theta(L) \cdot \epsilon_t = \prod_{j=1}^{p} \frac{1-\delta_j^k L^j}{1-\delta_j L} \cdot (\sum_{d=0}^{q} \theta_d L^d) \cdot \epsilon_t = \prod_{j=1}^{p} (\sum_{i=0}^{k-1} \delta_j^i L^i) \cdot (\sum_{d=0}^{q} \theta_d L^d) \cdot \epsilon_t = \prod_{j=1}^{p} (\sum_{i=0}^{k-1} L^i) (\sum_{d=0}^{q} \theta_d B) \cdot \epsilon_t.
\] (9)

Where \( \theta_0 = 1 \), the order of \( \epsilon_T^r \) is \( r = p(k - 1) + q \), the order of \( \epsilon_T^* \) is \( r = \lceil \frac{p(k-1)+q}{k} \rceil \), and \([b]\) indicates the integer part of a real number \( b \).

\[
T(L) \cdot \theta(L) \cdot \epsilon_t \equiv \eta(B) \cdot \epsilon_T^*.
\] (10)

The coefficients of \( \eta(B) \) are solved by equalizing the variance and the auto-covariance of R.H.S and L.H.S of the above equation. Since \( \eta(B) \) has the highest \( r \) order, so it has a total of \( r \) auto-covariances. Hence it becomes a problem of solving a system of equations with \( (r + 1) \) unknowns and \( (r + 1) \) nonlinear equations.

Let \( G_t = T(L) \cdot \theta(L) \cdot \epsilon_t \) and \( H_T = \eta(B) \cdot \epsilon_T^* \), then the covariance of \( G_t \) is \( \gamma_0 = E(G_t \cdot G_t) \), its \( r \) auto-covariance are respectively \( \gamma_i = E(G_t \cdot G_{t-i}) \), \( i = 1 \cdots r \). And the covariance of \( H_T \) is \( \Gamma_0 = E(H_T \cdot H_T) \), its \( r \) auto-covariance are respectively \( \Gamma_i = E(H_T \cdot H_{T-i}) \), \( i = 1 \cdots r \). So the equation system can be expressed as

\[
\begin{cases}
G_0 = \gamma_0 \\
G_1 = \gamma_1 \\
\cdots \\
G_r = \gamma_r
\end{cases}
\] (11)

From which coefficients \( \eta_i \) of \( \beta(B) \), \( i = 1 \cdots r \) and covariance \( \sigma_\epsilon^2 \) of \( \epsilon_T^* \) can be yielded.

For flow case, \( T(L) \) is shown below:

\[
T(L) = \left[ \frac{1-\delta_j^k}{1-\delta_j L} \right] \cdot \prod_{j=1}^{p} \frac{1-\delta_j^k L^j}{1-\delta_j L} = (\sum_{s=0}^{k-1} L^s) \cdot \prod_{j=1}^{p} (\sum_{i=0}^{k-1} \delta_j^i L^i).
\] (12)
An Example of ARMA(1,2)

\[(1 - \phi_1 L - \phi_2 L^2)y_t = (1 + \theta L)\epsilon_t. \quad (13)\]

Where the absolute value of \( \theta \) is less than 1. Let \( \delta_1 \) and \( \delta_2 \) be the inverted roots of \((1 - \phi_1 L - \phi_2 L^2)\) which lie inside the unit circle, thus \((1 - \phi_1 L - \phi_2 L^2)\) is equal to \((1 - \delta_1 L)(1 - \delta_2 L)\).

In this case, we aggregate monthly model into quarterly model, that is, \( k = 3 \), and \( B = L^3 \), so \( T(L) \) can be expressed as \( T(L) = \frac{1 - \delta_1^3 L^3}{1 - \delta_1 L} \cdot \frac{1 - \delta_2^3 L^3}{1 - \delta_2 L} \).

Multiply both sides of equation above by \( T(L) \), the AR part can be directly obtained

\[T(L) \cdot \prod_{j=1}^{2} (1 - \delta_j L)y_t = (1 - \delta_1^3 L^3)(1 - \delta_2^3 L^3)y_t. \quad (14)\]

For MA part,

\[T(L) \cdot (1 + \theta L)\epsilon_t = (\sum_{j=0}^{2} \delta_j^3 L^j)(\sum_{j=0}^{2} \delta_j^3 L^j)(1 + \theta L)\epsilon_t = 1 + (\theta + \delta_1 + \delta_2)L + (\delta_1^2 + \delta_1 \delta_2 + \theta \delta_1 + \delta_2^2 + \theta \delta_2) L^2 + (\delta_1^2 \delta_2 + \theta \delta_1^2 \delta_2 + \delta_1 \delta_2^2 + \theta \delta_1 \delta_2 + \theta \delta_2) L^3 + (\delta_1^2 \delta_2^2 + \theta \delta_1^2 \delta_2^2 + \delta_1 \delta_2^2 + \theta \delta_1 \delta_2^2 + \theta \delta_1 \delta_2) L^4 + (\theta \delta_1^2 \delta_2^2 \delta_2 + \theta \delta_1 \delta_2^2 \delta_2 + \theta \delta_1 \delta_2) L^5)\epsilon_t. \quad (15)\]

The order of MA after aggregation is \( r = \frac{p(k-1)+q}{k} = \frac{2^{k-1}+1}{3} = 1 \).

Because \( r = 1 \), so we can suppose that \( \hat{H}_T = \eta(B)\epsilon_t = (1 + \eta B)\epsilon_t \), the covariance and auto-covariance of which are

\[\Gamma_0 = (1 + \eta^2)\sigma_{\epsilon_t}^2, \quad \Gamma_1 = \eta \sigma_{\epsilon_t}^2.\]

While \( G_t = T(L)(1 + \theta L)\epsilon_t \), the covariance and auto-covariance of which are

\[y_0 = [1 + (\theta + \delta_1 + \delta_2)^2 + (\delta_1^2 + \delta_1 \delta_2 + \theta \delta_1 + \delta_2^2 + \theta \delta_2)^2 + (\delta_1^2 \delta_2 + \theta \delta_1^2 \delta_2 + \delta_1 \delta_2^2 + \theta \delta_1 \delta_2)^2 + (\delta_1^2 \delta_2^2 + \theta \delta_1^2 \delta_2^2 + \delta_1 \delta_2^2 + \theta \delta_1 \delta_2) L^2 + (\delta_1^2 \delta_2^2 + \theta \delta_1^2 \delta_2^2 + \delta_1 \delta_2^2 + \theta \delta_1 \delta_2) L^3 + (\delta_1^2 \delta_2^2 \delta_2 + \theta \delta_1^2 \delta_2^2 \delta_2 + \delta_1 \delta_2^2 \delta_2 + \theta \delta_1 \delta_2 \delta_2) L^4 + (\theta \delta_1^2 \delta_2^2 \delta_2 \delta_2 + \theta \delta_1 \delta_2^2 \delta_2 \delta_2 + \theta \delta_1 \delta_2 \delta_2 \delta_2) L^5)]\sigma_{\epsilon_t}^2.
\]

\[y_1 = [(\delta_1^2 \delta_2 + \theta \delta_1^2 \delta_2 + \delta_1 \delta_2^2 + \theta \delta_1 \delta_2 + \theta \delta_2)(\theta + \delta_1 + \delta_2)(\delta_1^2 \delta_2^2 + \theta \delta_1^2 \delta_2^2 + \delta_1 \delta_2^2 + \theta \delta_1 \delta_2 + \theta \delta_1 \delta_2)]\sigma_{\epsilon_t}^2.
\]

Let \( \Gamma_0 = y_0 \) and \( \Gamma_1 = y_1 \), we have

\[
\frac{\eta}{1 + \eta^2} = \frac{y_1}{y_0}.
\]

Let \( \rho_1 = \frac{y_1}{y_0} \), it’s obvious that \( \rho_1 \) is the correlation coefficient. After simplification, we can yield that \( 1 + \eta^2 - \frac{\eta}{\rho_1} = 0 \), and the root is \( \eta = \frac{1}{2} \pm \sqrt{\frac{1}{\rho_1^2} - 4} \), since \( \eta \) is inside the unit circle, thus \( \sigma_{\epsilon_t}^2 = \frac{y_1}{y_0} \).

**Regularity Results**

1) The lag order of AR(p)

The AR model structure remains unchanged after temporal aggregation, and the lag order \( p \) remains the same as well. Models that have unchanged lag order are called stable model, apparently MA model is not stable.

2) The lag order of MA(q)

As \( k \) continues to grow until it tends to infinity, when \( q \geq p, \ r = p \) and when \( q < p, \ r = p - 1 \). The MA(q) model may transfer to MA(p) or MA(p-1) model, which means when \( k \) is large enough, the lag order of MA(q) is determined by the lag order of AR(p), this phenomenon may emerge when we aggregate daily model into yearly and 10-year model.
\[ \lim_{k \to +\infty} r_{t \epsilon_t}^k = \begin{cases} 
\ p, \ q \geq p \\
\ p - 1, \ q < p. \end{cases} \]  

(16)

3) Lag order comparison of MA (q) before and after aggregation

The lag order of MA(q) is \( r = \left\lfloor \frac{p(k - 1) + q}{k} \right\rfloor \), we rewrite it as \( r = \frac{p(k - 1) + q}{k} - \rho \) for convenience, where \( 0 \leq \rho < 1 \). By contrasting \((p + 1 - q)(k - 1)\) and \( \rho k \) we can infer the size of lag order q before aggregation and that r after aggregation. Under the circumstances of \( p > q \), we have \( r > q \).

**Monte Carlo Simulation**

Firstly we make a brief description of the Monte Carlo simulation variables that will be presented below.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Dependent variable</td>
</tr>
<tr>
<td>u</td>
<td>Error term</td>
</tr>
<tr>
<td>MA(1)</td>
<td>The first order of white noise</td>
</tr>
<tr>
<td>MA(2)</td>
<td>The second order of white noise</td>
</tr>
<tr>
<td>Y(-1)</td>
<td>The first order lag item</td>
</tr>
<tr>
<td>Y(-2)</td>
<td>The second order lag item</td>
</tr>
</tbody>
</table>

According to the previous derivation, AR model is stable and its lag order remains unchanged. And it is not difficult to find that the coefficients of B are the k times power that of the operator L. We illustrate three models respectively for simulation, that is, ARMA(1,1), ARMA(2,1) and ARMA(2,2), which verifies the theory of model derivation above, and we also find some special regularities that could not be observed in theory.

<table>
<thead>
<tr>
<th>Coeff.</th>
<th>Monte Carlo Simulation</th>
<th>Theoretical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K=3  K=6  K=12</td>
<td>K=3  K=6  K=12</td>
</tr>
<tr>
<td>ARMA(1,1)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y(-1)</td>
<td>0.511742 0.262522 0.067095</td>
<td>0.512000 0.262100 0.068700</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.048591 0.018760 0.005650</td>
<td>0.048300 0.018800 0.004570</td>
</tr>
<tr>
<td>ARMA(2,1)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y(-1)</td>
<td>0.727157 0.305496 0.078474</td>
<td>0.728000 0.308800 0.089527</td>
</tr>
<tr>
<td>Y(-2)</td>
<td>-0.110075 -0.010910 -0.001130</td>
<td>-0.110600 -0.012231 -0.000312</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.221901 0.156474 0.042024</td>
<td>0.257000 0.152800 0.049670</td>
</tr>
<tr>
<td>ARMA(2,2)***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y(-1)</td>
<td>0.721992 0.397814 0.111100</td>
<td>0.728000 0.308800 0.089527</td>
</tr>
<tr>
<td>Y(-2)</td>
<td>-0.106968 -0.038026 -0.016096</td>
<td>-0.110600 -0.012231 -0.000312</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.277715 0.079809 0.012754</td>
<td>0.313000 0.168400 0.052400</td>
</tr>
<tr>
<td>MA(2)</td>
<td>0.005297 -0.016993 0.011467</td>
<td>0.003300 0.00154 0.015300</td>
</tr>
</tbody>
</table>

* ARMA(1,1) model: \( y_t = 0.8y_{t-1} + u_t + 0.2u_{t-1} \)
** ARMA(2,1) model: \( y_t = 1.4y_{t-1} - 0.48y_{t-2} + u_t + 0.2u_{t-1} \)
*** ARMA(2,2) model: \( y_t = 1.4y_{t-1} - 0.48y_{t-2} + u_t + 0.2u_{t-1} + 0.1u_{t-2} \)

As the aggregation frequency k grows, the coefficient of \( Y(-1) \) decreases gradually, thus the explanatory effect of the lagging part to the dependent variable vanishes gradually. When k tends to
be infinite, coefficients of AR lagging part are all tend to zero, which means they all lose explanation to $y$. Only $y_T^*$ remains, the whole model confirms to AR(1) form.

Coefficients of MA model are based on $\theta$ and $\delta$, we use control variable method to explore whether there is a relationship between $\beta$ and $\delta$. Since the MA part is solved by equation systems, hence we are not capable of expressing it in concrete mathematical form, so we fail to infer the explicit trend relation between $\delta$ and $\beta$.

It is apparent that $\beta$ will influence the accuracy of AR model estimation, which reminds us that $\delta$ may account for the coefficient bias between theory and simulation. Note that when $k$ grows, coefficients may be too small to be significant, such as case $k=12$ showed in the table.

Summary

The paper has considered temporal aggregation of ARMA model, including model derivation and regularity results, and applied Monte Carlo simulation to verify the theory. We showed that temporal aggregation transfer model of ARMA is accurate as long as $k$ is appropriate, but there are times that bias will emerge and researchers have to pay attention to. With this method, unexpected shocks or fluctuations can be timely integrated into the forecasting model for better prediction.

Acknowledgements

We are grateful to our teacher Zengmin Wang for inspiring our thoughts and strongly supporting us along the whole research process. Besides, this research is financially supported by the Research Innovation Fund for College Students of Beijing University of Posts and Telecommunications.

References


