FinM 331/Stat 339 Financial Data Analysis

(Applied Statistical Analysis of Financial Data in MATLAB)

Winter 2009

Floyd B. Hanson, Visiting Professor

Email: fhanson@uchicago.edu

Master of Science in Financial Mathematics Program
University of Chicago

Lecture 9 (Last)

6:30-9:30 pm, 02 March 2009, Ryerson 251 in Chicago

7:30-10:30 pm, 02 March 2009 at UBS Stamford

8:30-11:30 am, 03 March 2009, #02-01 Spring Singapore

9. Time Series Continued[a]

- ***9.1 Lecture 8 Postscript — Dueling, Decaying "Exponentials" in AR(2) Power Solutions when $D > 0$:***

This section is in response to a question during Lecture 8 concerning the oscillation in the decay of a linear combination of two real solutions when the discriminant $D = C_1^2 + 4C_2 > 0$. In the illustration of the $\rho(\mathrm{lag})$, **ACF** Figure 4, p. L8.39, corrected version, the selected AR(2) coefficients were $C_0 = 0.0$, $C_1 = 0.2 < 1$ and $C_2 = 0.6 < 1$. See the reprinted and corrected figures on the next two pages, here in Figures 1 and 2.

---

[a] Optional background references are R. S. Tsay ('05, U. of C. BGBS) *Analysis of Financial Time Series*, Chapter 2 and Carmona ('04) Chapter 5.

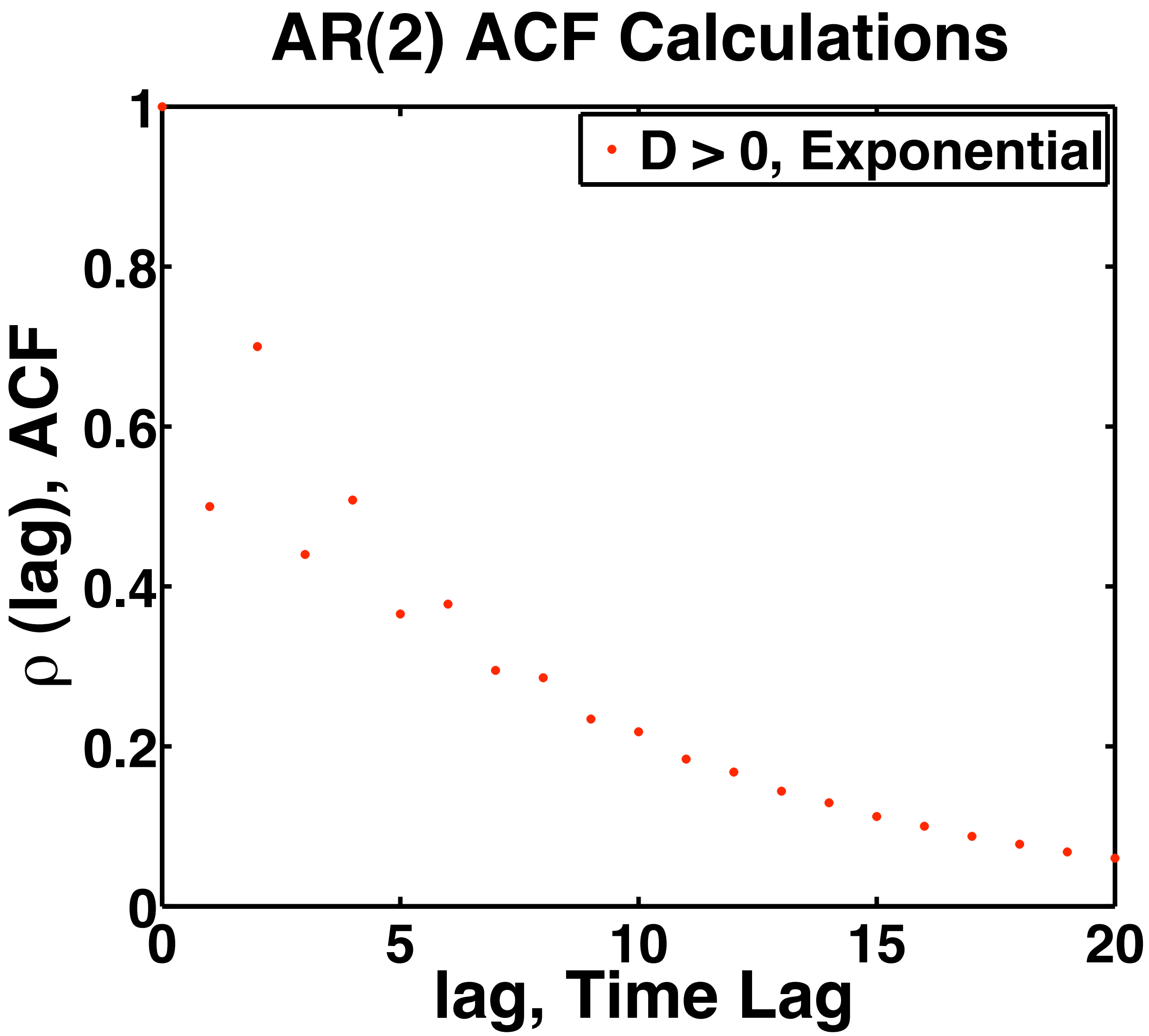

Figure 1: AR(2) ACF, $C1 = 0.2; C2 = 0.6; C0{=}0.0; D = 2.4 > 0;$.

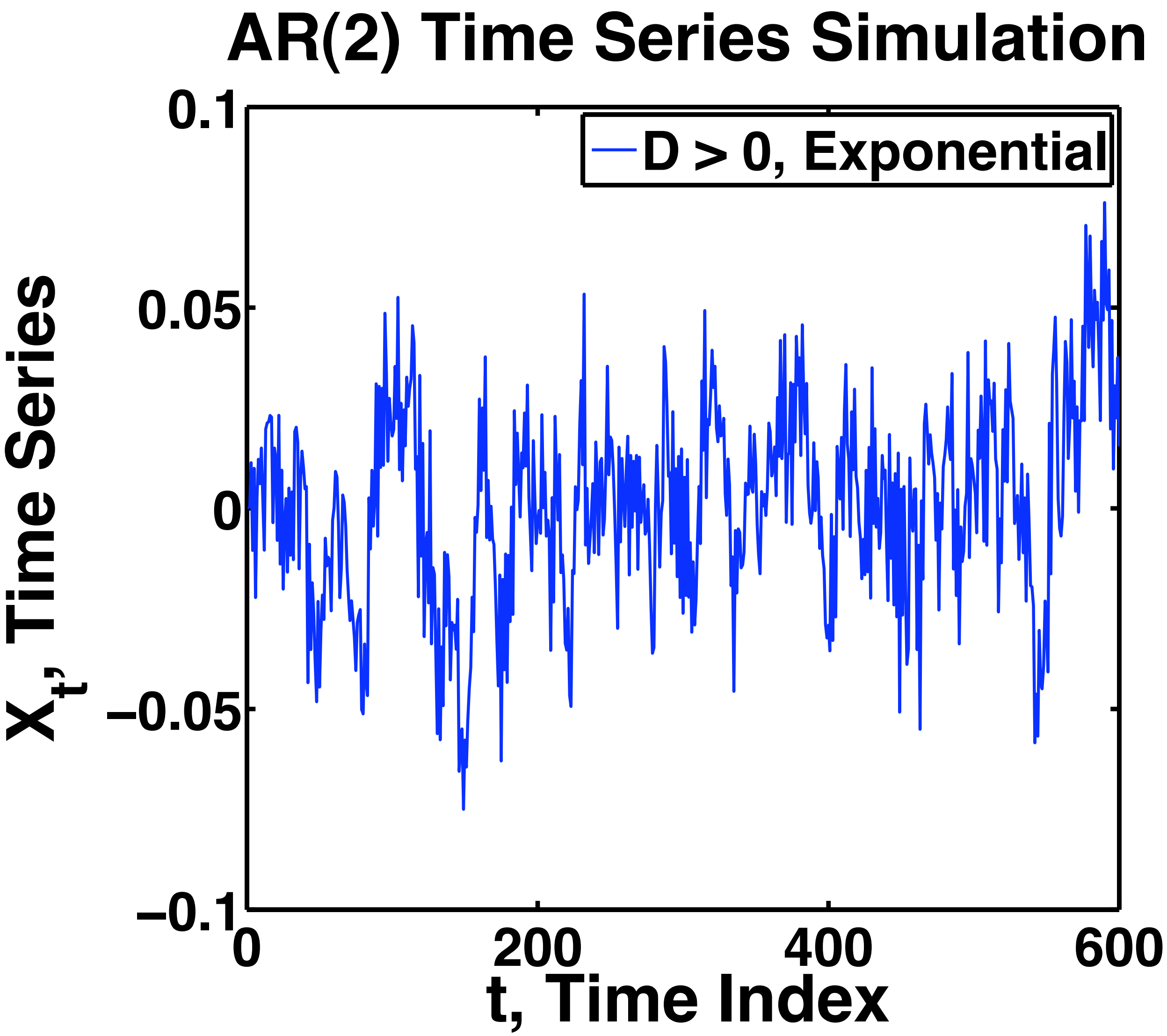

Figure 2: AR(2) time series, $C1 = 0.2; C2 = 0.6; C0 = 0.0; D = 2.4 > 0;$.

So, $D = 2.44 > 0$, with power roots $R_1 \simeq 0.8810 < 1$ and $R_2 \simeq -0.6810 < 0$, appearing in the solution linear combination

$$\rho^\ell = K_1 R_1^\ell + K_2 R_2^\ell,$$

Hence, the dominant root with largest magnitude is positive, while the subdominant root with smallest magnitude is negative, with $\rho_0 = K_1 + K_2 = 1$,

$$\rho_1 = K_1 R_1 + K_2 R_2 = K_1 R_1 - K_2 |R_2| = C_1/(1 - C_2),$$

and $\rho_2 = K_1 R_1^2 + K_2 R_2^2$. Note that there is a alternation in the sign of the subdominant term, giving rise to are discrete oscillations superimposed on the exponential or power decay, but is is unlike the sinusoidal oscillations in the complex root case, noting ℓ is discrete with integer values.

- ***9.2 Time Series Forecasting for AutoRegressive (AR(p)) Models:***

One reason for time series applications to financial data is ***forecasting future returns***, at least in the short term. In addition to finding the estimated best ***AR(p)*** models, for instance, it is important to filter out nonrelevant seasonal and other trends, as well as verifying fast decays for the ACF terms.

Using the estimated lag-p *AR(p)* model at time index $t+1$,

$$X_{p,t+1} = \widehat{C}_{1,0} + \sum_{j=1}^{p} \widehat{C}_{j,j} X_{t+1-j} + W_{p,t+1},$$

the one-step ahead forecast estimate of X_{t+1} is

$$\widehat{X}_{p,t+1}(1) \equiv \mathrm{E}\left[X_{t+1}\,\middle|\,[X_{t+1-i}]_{p\times 1}\right] = \widehat{C}_{1,0} + \sum_{j=1}^{p} \widehat{C}_{j,j} X_{t+1-j},$$

since $\mathrm{E}\left[W_{p,t+1}\,\middle|\,[X_{t+1-i}]_{p\times 1}\right] = 0.$

Similarly, the general ***k-step, multistep ahead forecast estimate*** of X_{t+k}, using prior estimated state values $\tau = t{+}1 : t{+}k{-}1$ for $k \geq 2$, is

$$\widehat{X}_{p,t+k}(k) = \widehat{C}_{1,0} + \left\{ \begin{array}{ll} \left\{ \begin{array}{l} \displaystyle\sum_{j=1}^{k-1} \widehat{C}_{j,j} \widehat{X}_{t+k-j}(k-j) \\ \displaystyle + \sum_{j=k}^{p} \widehat{C}_{j,j} X_{t+k-j} \end{array} \right\}, & 2 \leq k < p \\ \displaystyle\sum_{j=1}^{p} \widehat{C}_{j,j} \widehat{X}_{t+k-j}(k-j), & k \geq p \end{array} \right\},$$

where in the last line the ***AR(p)*** RHS contains only estimated values. It is much simpler to code than it seems.

The associated k-step ahead forecast error for $X_{p,t+k}$, with $\mathrm{km} = \min(p, k-1)$, is

$$\begin{aligned} e_{p,t+k}(k) &= X_{p,t+k} - \widehat{X}_{p,t+k}(k) \\ &= \sum_{j=1}^{\mathrm{km}} \widehat{C}_{j,j}(X_{t+k-j} - \widehat{X}_{t+k-j}(k-j)) + W_{p,t+k} \\ &= W_{p,t+k} + \sum_{j=1}^{\mathrm{km}} \widehat{C}_{j,j} W_{p,t+k-j}. \end{aligned}$$

So $e_{p,t+k}(k)$ has mean zero and variance

$$\sigma_e^2(k) \equiv \mathrm{Var}[e_{p,t+k}(k)] = \sigma_W^2\left(1 + \sum_{j=1}^{\mathrm{km}} \widehat{C}_{j,j}^2\right).$$

Thus, if we have a ***Gaussian***[a] ***AR(p) model with 1-step***[b] and α is the level of significance then the $(1-\alpha)\cdot 100\%$ ***confidence interval*** for $X_{p,t+k}$, is

$$\left[\widehat{X}_{p,t+1}(1)-\sigma_W\cdot\Delta z,\ \widehat{X}_{p,t+1}(1)+\sigma_W\cdot\Delta z\right],$$

where $\Delta z = -(F_Z^{(\mathrm{n})})^{-1}(\alpha/2;0,1) > 0$[c] for $\alpha \in [0,1)$.

---

[a] Yes, we are still talking about relatively simple models, so jumps and stochastic parameters will have to wait, likely in a more advanced course.

[b] The ***k-step*** model is a mixture of independent Gaussians, so technically the rigorous inverse to get Δz would require inverting the joint distribution of $\min(p+1,k)$ Gaussians, but for practical purposes the value of $\sigma_e^2(k)$ might be used for an ***approximate confidence interval***.

[c] In MATLAB Stat. Toolbox, use

`Dz = -norminv(alpha/2,0,1)`.

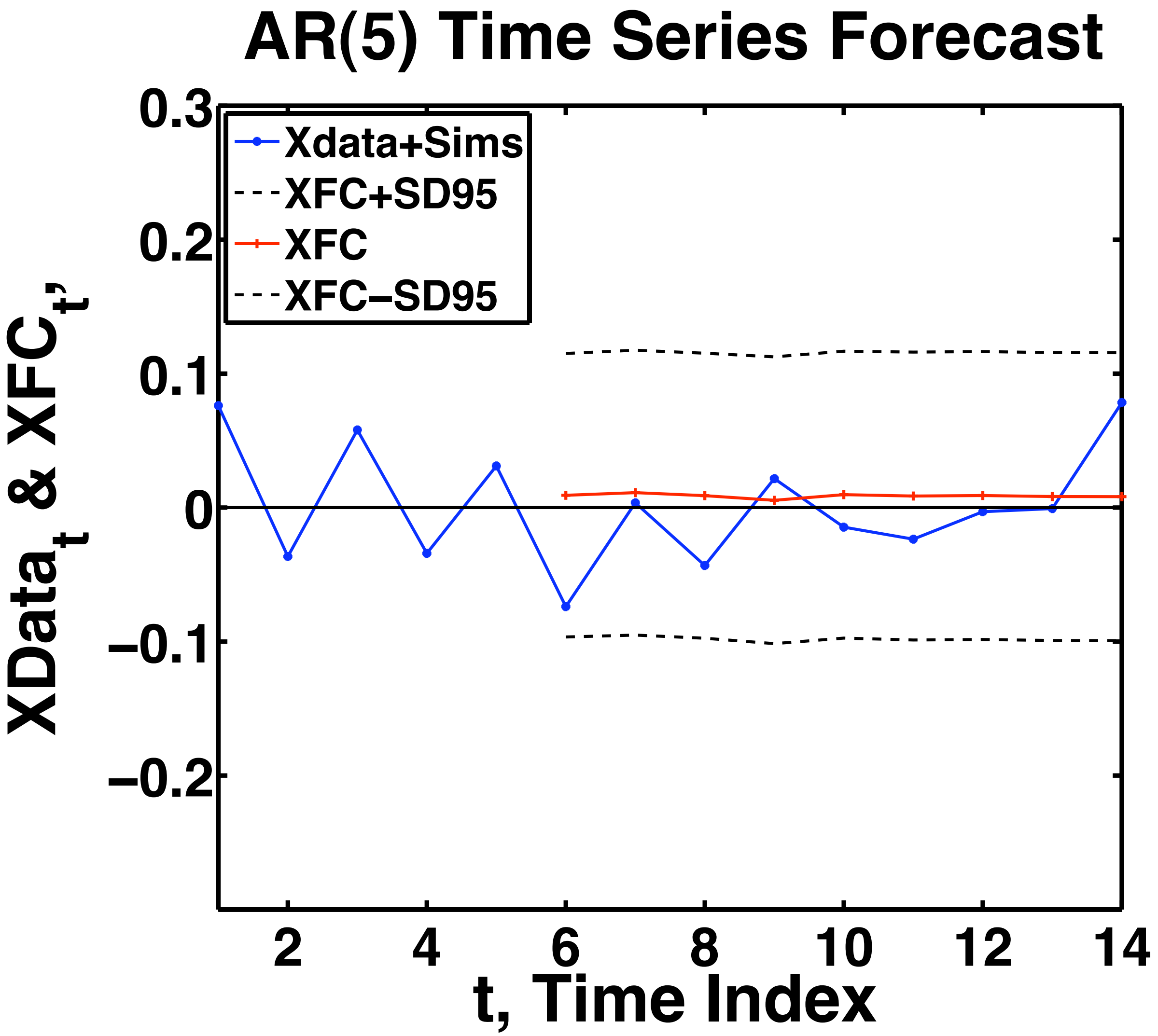

Figure 3: Corrected forecast for AR(5) model and data of Tsay.

```
function timeseries09AR5Forecast
% Time series AR(5) Demo using Tsay (p.49-50 data), Finm331 W09
% CORRECTED 2/7/9 after student question (Jianping Zou).
clc % clear workspace of prior output.
clear % clear variables.
fprintf(’\nfunction timeseries09AR5Forecast.m OutPut:’);
%
n = 14;
C0 = 0.0075; C1 = 0.103; C2 = 0.002;
C3 = -0.114; C4 = 0.032; C5 = 0.084;
CR = [C5 C4 C3 C2 C1]’; % Reverse order got dot product
C = [C1 C2 C3 C4 C5]’; p = 5; ppo = p+1; % C backward: vector;
Xdata = [0.0762 -0.0365 0.0580 -0.0341 0.0311 0.0183]’;
sigw = 0.054; sig2w = sigwˆ2;
fprintf(’\n[C0 C1 C2 C3 C4 C5]=’);
fprintf(’[%7.4f %7.4f %7.4f %7.4f %7.4f %7.4f];’,C0,C);
fprintf(’\nXdata=[%7.4f %7.4f %7.4f %7.4f %7.4f %7.4f];’,Xdata);
fprintf(’\n sigw = %8.3e; sig2w = %8.3e;’,sigw,sig2w);
W = normrnd(0,sigw,1,n);
X = zeros(n,1); X(1:ppo,1) = Xdata;
XFC = zeros(n,1); XFC(1:ppo,1) = Xdata; % add data
```

```
fprintf('\nsize(CR)=[%i,%i]; size(X(1:p))=[%i,%i];'...
    ,size(CR),size(X(1:p)));
for t = p+1:n % Forecasts & Sima:
 XFC(t) = C0+CR'*XFC(t-p:t-1); % CORRECTED 3/7/9
 X(t) = C0+CR'*X(t-p:t-1)+W(t);
end
Xmean = mean(X); Xvar = var(X); Xvol = sqrt(Xvar);
fprintf('\n Xmean = %8.3e; Xvol = %8.3e; Xvar = %8.3e '...
    ,Xmean,Xvol,Xvar);
XFCmean = mean(XFC(ppo:n,1)); XFCvar = var(XFC(ppo:n,1));
XFCvol = sqrt(XFCvar);  % CORRECTED 3/7/9
fprintf('\n XFCmean = %8.3e; XFCvol = %8.3e; XFCvar = %8.3e '...
    ,XFCmean,XFCvol,XFCvar);
nmp = n-p;
sig2FC = zeros(nmp,1); sig2FC(1,1) = sig2w;
for k = 2:nmp % k-Step Forecast Error:
    km = min(p,k-1);
    sig2FC(k) = sig2w*(1+C(1:km,1)'*C(1:km,1));
end
sigFC = sqrt(sig2FC);
alpha = 0.05;
```

```
Dz = -norminv(alpha/2,0,1);
fprintf('\nDz = %7.4f;',Dz);
sig95 = Dz*sigFC;
%%%% Begin Plots:
scrsize = get(0,'ScreenSize'); % target screen figure spacing
ss = [5.0,4.0,3.5]; % figure spacing factors
nfig = 0;
nfig = nfig + 1;
%
fprintf('\n\nFigure(%i):  AR(2) Time Series Simulation\n' ...
    ,nfig)
figure(nfig)
marks = {'b-*','r-+','k-','k--'}; % plot mark list.
Xaxis = zeros(1,n);
%
plot((1:n)',X,marks{1},'linewidth',2);  % CORRECTED 3/7/9
hold on;
% CORRECTED 3/7/9
plot((ppo:n)',XFC(ppo:n,1)+sig95,marks{4},'linewidth',2);
plot((ppo:n)',XFC(ppo:n,1),marks{2},'linewidth',2);
plot((ppo:n)',XFC(ppo:n,1)-sig95,marks{4},'linewidth',2);
```

```
%
plot((1:n)',Xaxis,marks{3},'linewidth',2);
hold off
axis ([1 n -0.3 0.3]) %([xmin xmax ymin ymax])
%
title('AR(5) Time Series Forecast'...
    ,'FontWeight','Bold','Fontsize',44);
ylabel('XData_t & XFC_t,'...
    ,'FontWeight','Bold','Fontsize',44);
xlabel('t, Time Index'...
    ,'FontWeight','Bold','Fontsize',44);
hlegend = legend('Xdata & Sims','XFC+SD95','XFC','XFC-SD95'...
    ,'Location','NorthWest');  % CORRECTED 3/7/9
set(hlegend,'Fontsize',28,'FontWeight','Bold');
set(gca,'Fontsize',36,'FontWeight','Bold','linewidth',3);
set(gcf,'Color','White','Position' ...
   ,[scrsize(3)/ss(nfig) 60 scrsize(3)*0.60 scrsize(4)*0.80]);
%[l,b,w,h]
% End Code
```

Remarks: The forecasts for different trials can be very variable, because there is a single sample of the solution at each k-step. Carmona ('04) suggest using the Monte Carlo method with a series of simulation, but does not really do anything but simulate, so with out taking a sample average at some stage it is not the Monte Carlo Method.

- ***9.3 Moving Average Models — Introduction:***

Moving average models are related to autoregressive models and can be derived from the infinite AR(p) as $p \longrightarrow \infty$,

$$X_t = C_0 + \sum_{j=1}^{\infty} C_j X_{t-j} + W_t,$$

that can be simplified by setting most by relating them to a power of a single coefficient, i.e., $C_j = -c_1^j$ for $j \geq 1$, so

$$X_t = C_0 - \sum_{j=1}^{\infty} c_1^j X_{t-j} + W_t,$$

requiring $|c_1| < 1$ for convergence.

This form can be reduced to a finite form by taking the prior state equation,

$$X_{t-1} = C_0 - \sum_{j=1}^{\infty} c_1^j X_{t-1-j} + W_{t-1},$$

multiplying it my c_1 and subtracting from the current state equation producing the ***MA model of order one or MA(1)***,

$$X_t = c_0 + W_t - c_1 W_{t-1},$$

where $c_0 \equiv C_0(1 - c_1)$, while W_t and W_{t-1} are uncorrelated noise. The ***MA(2) model*** has the form,

$$X_t = c_0 + W_t - c_1 W_{t-1} - c_2 W_{t-2},$$

where $\{W_t, W_{t-1}, W_{t-2}\}$ are an uncorrelated noise set.

The general model, $MA(q)$, is

$$X_t = c_0 + W_t - \sum_{j=1}^{q} c_j W_{t-j},$$

where $q \geq 1$ and the white noise terms are all pairwise uncorrelated with mean zero and common variance σ_W^2[a]. This is the ***pure moving average model*** due to its short memory, here of a single step, but it is not autoregressive since not even X_{t-1} appears on the RHS, although at least its noise W_{t-1} does.

The state mean is $\mathrm{E}[X_t] = c_0$ and variance is

$$\mathrm{Var}[X_t] = \sigma_W^2 \left(1 + \sum_{j=1}^{q} c_j^2 \right).$$

---

[a] This is called ***weak white noise***, while ***strong white noise*** requires noise independence, and neither have to be Gaussian, Carmona ('04), p. 266.

The short term memory of $MA(q)$ is shown in the autocovariance term for lag-ℓ with

$$X_{t-\ell} = c_0 + W_{t-\ell} - \sum_{j=1}^{q} c_j W_{t-\ell-j}.$$

First, consider the simplest case, *MA(1)*, using the noncorrelation of the noise components,

$\mathrm{Cov}[W_t, W_{t-\ell}] = \sigma_W^2 \delta_{|\ell|,0}$[a],

$$\begin{aligned}\gamma_\ell = \mathrm{Cov}[X_t, X_{t-\ell}] &= \mathrm{E}[W_t W_{t-\ell} \\ &\quad - c_1(W_{t-1}W_{t-\ell} + W_t W_{t-1-\ell}) + c_1^2 W_{t-1}W_{t-\ell-1}] \\ &= \sigma_W^2 \left(\delta_{|\ell|,0} - c_1\delta_{|\ell|,1} + c_1^2\delta_{|\ell|,0}\right).\end{aligned}$$

---

[a] Recall, $\delta_{i,j}$ is the Kronecker delta which is 1 only when $j = i$, otherwise it is 0.

Summarizing the MA(1), in terms of the lag-ℓ auto-/correlation function, the ACF[a],

$$\rho_\ell = \frac{\gamma_\ell}{\gamma_0} = \left\{ \begin{array}{ll} 1, & |\ell| = 0 \\ -c_1/(1+c_1^2), & |\ell| = 1 \\ 0, & |\ell| > 1 \end{array} \right\},$$

so the memory is cutoff at $q = 1$ with zero correlation for $\ell > 1$.

---

[a] Tsay ('05) briefly cites an application of Roll ('84) to the use of this model for the bounce in the bid-ask spread having a lag-1 negative correlation to the asset return, pp. 50 & 211.

The ***MA(2)*** model results are similar,

$$\gamma_\ell = \sigma_W^2\big(\delta_{|\ell|,0} - c_1\delta_{|\ell|,1} - c_2\delta_{|\ell|,2} + (c_1^2 + c_2^2)\delta_{|\ell|,0}$$
$$+c_1c_2\delta_{|\ell|,1}\big)$$

or in ACF form,

$$\rho_\ell = \left\{\begin{array}{ll} 1, & |\ell| = 0 \\ -c_1(1-c_2)/(1+c_1^2+c_2^2), & |\ell| = 1 \\ -c_2/(1+c_1^2+c_2^2), & |\ell| = 2 \\ 0, & |\ell| > 2 \end{array}\right\},$$

again the correlation and hence memory cutoff at $\ell = q = 2$ with no correlation for lags beyond lag-q.

The general lag-q *MA(q)* model results are

$$\gamma_\ell = \sigma_W^2\left(\delta_{|\ell|,0} - \sum_{j=1}^{q} c_j\delta_{|\ell|,j} + \sum_{j=1}^{q}\sum_{k=1}^{q} c_j c_k \delta_{\ell+k,j}\right)$$

or in ACF form,

$$\rho_\ell = \left\{\begin{array}{ll} 1, & |\ell| = 0 \\ \dfrac{-c_{|\ell|} + \displaystyle\sum_{j=|\ell|+1}^{q} c_j c_{j-|\ell|}}{1+\sum_{k=1}^{q} c_k^2}, & |\ell| = 1\!:\!q-1 \\ \dfrac{-c_q}{1+\sum_{k=1}^{q} c_k^2}, & |\ell| = q \\ 0, & |\ell| > q \end{array}\right\},$$

again the correlation and hence memory cutoff at the MA(q) order $\ell = q$ with no correlation for lags beyond lag-q.

- ***9.4 Moving Average Models — Identification, Estimation and Forecasting:***

Tsay ('05) presents a number of recommendations for the care and use of the pure moving average models, ***MA(q)***:

1. ***Simple ACF Technique for Identifying MA(q) for the Application***: Computing the sample ACFs from the application, determine which lags have the most significant sample ACF values, e.g., exceeding one standard error (SE) above and below zero, select the most extreme of these values and they do not have to be successive lags, say $\{q_1, q_2, q_3\}$, construct the model:

   $$X_t = c_0 + W_t - c_{q_1} W_{t-q_1} - c_{q_2} W_{t-q_2} - c_{q_3} W_{t-q_1}.$$

   For instance, Tsay, pp. 52-53, selects lags $\{1, 3, 9\}$ making an MA(9) model with 6 zero coefficients.

2. ***Full Maximum Likelihood Estimation for MA(q)***: This includes going back before current time to estimate noise terms, but for large sample Tsay, p. 53-54, finds little difference between full and partial or conditional MLE. The ***Ljung-Box (LB) test***[a] is used to test for presence of serial correlations, with statistics for testing lags up to lag-m given by

$$Q(m) = T(T+2)\sum_{\ell=1}^{m} \widehat{\rho}_\ell^2/(T-\ell),$$

where T is the total time index, $Q(m)$ is compared with a chi-squared distribution[b], rejecting the null hypothesis that there is no correlation if $Q(m) > (\chi_m^2)^{-1}(1-\alpha)$, where $Q_m^* = (\chi_m^2)^{-1}(1-\alpha)$ is the $100(1-\alpha)\%$ percentile of $\chi_m^2(Q)$, α is the significance level and m is the degrees of freedom, while if the test p-value is given, then the null hypothesis is rejected if $p = 1 - \chi_m^2(Q(m)) \leq \alpha$.

---

[a] Biometrika, vol. 66, pp. 67-72, 1978; or Tsay, p. 27

[b] In MATLAB, $(\chi_m^2)^{-1}(1-\alpha) =$ `chi2inv(1-alpha,m)` is a chi-square table entry with test value `p=1-chi2cdf(Q(m),m)`, the complementary probability corresponding to α.

3. ***Forecasting with MA(q) models — Early Reversion to the Sample Mean***: Suppose we have an estimate of the sample mean,

$$\widehat{c}_0 = \frac{1}{n}\sum_{t=1}^{T} X_t,$$

and the c_1 used in the ***MA(1) model***,

$$X_{1,t} = \widehat{c}_0 + W_t - \widehat{c}_1 W_{t-1},$$

then the ***1-step forecast estimate***, conditioned on prior information $\{W_t, W_{t-1}\}$ is

$$\widehat{X}_{1,t+1}(1) = \mathrm{E}[X_{1,t+1}|W_t, W_{t-1}] = \widehat{c}_0 - \widehat{c}_1 W_t,$$

with 1-step forecasting error,

$$e_{1,t+1}(1) = X_{1,t} - \widehat{X}_{1,t+1}(1) = W_{t+1},$$

having variance $e_{1,t+1}(1) = \sigma_W^2$.

In the case of the ***k-step forecast*** for $k \geq 2$ for ***MA(1)*** and conditioned on prior information $\{W_t, W_{t-1}\}$,

$$\widehat{X}_{1,t+k}(k) = \mathrm{E}[X_{1,t+k}|W_t, W_{t-1}]$$

$$= \widehat{c}_0 - \widehat{c}_1\mathrm{E}[W_{t+k-1}|W_t, W_{t-1}] = \widehat{c}_0,$$

reverting to the mean as advertised, with error,

$$e_{1,t+k}(k) = X_{1,t+k} - \widehat{X}_{1,t+k}(k) = W_{t+k} - \widehat{c}_1 W_{t+k-1},$$

having greater variance $\mathrm{Var}[e_{1,t+k}(k)] = \sigma_W^2(1 + c_1^2)$.

In the case of the ***k*-step forecast**, for ***MA(2)*** and conditioned on prior information $\{W_t, W_{t-1}, W_{t-2}\}$,

$$\widehat{X}_{1,t+k}(k) = \mathrm{E}[X_{1,t+k}|W_{t-i}, i = 0\!:\!2]$$
$$= \widehat{c}_0 - \sum_{j=1}^{2} \widehat{c}_j \mathrm{E}[W_{t+k-j}|W_{t-i}, i = 0\!:\!2]$$
$$= \widehat{c}_0 - \widehat{c}_1 W_{t+k-1}\delta_{k,1} - \widehat{c}_2 W_{t+k-2}(\delta_{k,2} + \delta_{k,1}),$$

so $\widehat{X}_{1,t+k}(k) = \widehat{c}_0$ when $k > 2$, ***reverting to the mean as advertised, with error***,

$$e_{1,t+k}(k) = X_{1,t+k} - \widehat{X}_{1,t+k}(k)$$
$$= W_{t+k} - \widehat{c}_1 W_{t+k-1}(1 - \delta_{k,1})$$
$$-\widehat{c}_2 W_{t+k-2}(1 - \delta_{k,2} - \delta_{k,1}),$$

having variance (using the 0-1 law) $\mathrm{Var}[e_{1,t+k}(k)] = \sigma_W^2(1 + c_1^2(1 - \delta_{k,1}) + c_2^2(1 - \delta_{k,2} - \delta_{k,1}))$.

- ***9.5 AutoRegressive Moving Average Models (ARMA = AR + MA):***

The ***autoregressive moving average models*** combine the attractive features of both models, the short memory of the moving average (MA) models and the regression like features of the autoregressive (AR) models, while managing to keep the parameter space sufficiently small. ARMA models are useful for volatility models, for example ***Generalized AutoRegressive Conditional Heteroscedastic***[a] ***(GARCH)*** models are a form of ARMA models for volatility studies.

---

[a] ***Heteroscedasticity*** refers to variable volatility or variance, clearly a property of financial markets, in contrast, an underlying assumption of the Black-Scholes model is ***homoscedastic*** or constant volatility.

For instance, the ***ARMA(1,1)*** time series model combines AR(1) and MA(1) models so has the form,

$$X_t = C_0 + C_1 X_{t-1} + W_t - c_1 W_{t-1},$$

where the $\{W_t, W_{t-1}\}$ represents the prior zero-mean, common variance white noise information and the AR and MA coefficient need to be distinct, i.e., $C_1 \neq c_1$ for a nontrivial ARMA model due to cancellations.

Under the AR assumption of weak stationarity implying a common, time-independent state mean $\mu_X = \mathrm{E}[X_t]$ and zero-mean white noise $\mathrm{E}[W_t] = 0$, $\forall t$,

$$\mu_X = C_0/(1 - C_1),$$

the AR(1) result, provided $C_1 \neq 1$.

Next, the current state-noise covariance, using $\delta X_t \equiv X_t - \mu_X$ as the deviation, is

$$\begin{aligned}\mathrm{Cov}[X_t, W_t] &= \mathrm{E}[(C_1 \delta X_{t-1} + W_t - c_1 W_{t-1}) W_t] \\ &= \sigma_W^2,\end{aligned}$$

since both X_{t-1} and W_{t-1} are independent of W_t.

Similarly, the weak stationary common state variance is

$$\begin{aligned}\mathrm{Var}[X_t] &= \mathrm{E}[(C_1 \delta X_{t-1} + W_t - c_1 W_{t-1})^2] \\ &= C_1^2 \mathrm{Var}[X_{t-1}] - 2C_1 c_1 \mathrm{Cov}[X_{t-1}, W_{t-1}] \\ &\quad + \sigma_W^2 (1 + c_1^2) \\ &= \sigma_W^2 (1 - 2C_1 c_1 + c_1^2)/(1 - C_1^2) \equiv \gamma_0,\end{aligned}$$

provided $C_1^2 < 1$ to preserve positivity and boundedness of the variance. (Note, $1 - 2C_1 c_1 + c_1^2 = 1 - C_1^2 + (c_1 - C_1)^2 > 0$.)

The lag-ℓ autocovariance function recursion is similarly found,

$$\begin{aligned}\gamma_\ell \equiv \mathrm{Cov}[X_t, X_{t-\ell}] \\ &= \mathrm{E}[(C_1 \delta X_{t-1} + W_t - c_1 W_{t-1})\delta X_{t-\ell}] \\ &= C_1 \gamma_{\ell-1} + \sigma_W^2 (\delta_{\ell,0} - c_1 \delta_{\ell,1}) \\ &= C_1 \gamma_{\ell-1} - c_1 \sigma_W^2 \delta_{\ell,1},\end{aligned}$$

for $\ell \geq 1$, so $\gamma_1 = C_1 \gamma_0 - c_1 \sigma_W^2$, and $\gamma_\ell = C_1 \gamma_{\ell-1}$ for $\ell > 1$, recovering the ***AR(1) power recursion result***. In terms of the ACF or lag-ℓ correlation, we have $\rho_1 = C_1 - c_1 \sigma_W^2 / \gamma_0$ and $\rho_\ell = C_1 \rho_{\ell-1} = C_1^{\ell-1} \rho_1$ for $\ell > 1$. Note that there is no correlation cutoff as for MA(1), but ARMA(1,1)'s ACF has a decaying exponential or power law decay for all lags ℓ like AR(1), except it starts at lag of 2.

The general ARMA model has a $p \geq 0$ lag in the state and a $q \geq 0$ lag in the noise and is called an ***ARMA(p,q)*** model, a combination of ***AR(p)*** and ***MA(q)*** models,

$$X_t = C_0 + \sum_{j=1}^{p} C_j X_{t-j} + W_t - \sum_{j=1}^{q} c_j W_{t-j},$$

where again the $\{C_0, [C_i]_{p\times 1}, [c_i]_{q\times 1}\}$ are constant coefficients and the $\{[W_i]_{T\times 1}\}$ are usually zero-mean, σ_w^2-variance, Gaussian white noise, corresponding to model errors with respect to a state data sample, or else the number of terms in a simulation. T is the sample size, we called "n" when sampling early in the course.

Often in the analysis of time series, numerical methods or difference equations, ***shift operator formulations*** are studied to get information on the difference model and its solution. The ***1-step backward shift operator***[a] is defined for time series as $B_1[X_t] \equiv X_{t-1}$ and k-applications lead to $B_1^k[X_t] = X_{t-k}$. Also, the ***1-step forward shift operator*** is defined as $F_1[X_t] \equiv X_{t+1}$ and k-applications lead to $F_1^k[X_t] = X_{t+k}$. If I is the ***identity operator*** then $I[X_t] \equiv X_t$ and then the ***forward increment operation*** is

$$\Delta[X_t] \equiv X_{t+1} - X_t = (F_1 - I)[X_t] = (I - B_1)[X_{t+1}].$$

There is also the constant rule: $F_1[\mu] = B_1[\mu] = \mu$.

---

[a] Here, the arguments of operators are enclosed in square brackets ($[,]$) to set the limits of the scope of the operator, which would otherwise be ambiguous and to distinguish the operator from the ordinary function .

The shift or difference operators are also defined for inverses or negative powers, e.g., $B_1^{-1}[X_t] = X_{t+1} = F_1[X_t]$, so for arbitrary X_t, we write $B_1^{-1} \overset{\text{op}}{=} F_1$ in the sense of ***operator equality***, and similarly $F_1^{-1} \overset{\text{op}}{=} B_1$.

As an example, consider the ACF for ARMA(1,1) is $\rho_\ell = C_1 \rho_{\ell-1}$ for $\ell > 1$ and derive the power solution, though it already has been given:

$$\rho_\ell = C_1 \rho_{\ell-1} \;\; \& \;\; \rho_\ell = F_1[\rho_{\ell-1}] \;\; \& \;\; F_1^{\ell-1}[\rho_1] = \rho_\ell$$

$$\Rightarrow (F_1 - C_1 I)[\rho_{\ell-1}] = 0 \Rightarrow F_1 - C_1 I \overset{\text{op}}{=} 0$$

$$\Rightarrow F_1 \overset{\text{op}}{=} C_1 I \Rightarrow \rho_\ell = (C_1 I)^{\ell-1}[\rho_1] = C_1^{\ell-1}[\rho_1],$$

which was reported previously, where we used the property that $I^k \overset{\text{op}}{=} I$ for any integer k.

In the use of forward shift F_1 in the previous example, the analysis was done in minute details to illustrate application of the operator shift rules. However, in the usual applications, we would use less of the small details, for instance, using the backward shift for the same ACF recursion,

$$\begin{gathered}\rho_\ell = C_1\rho_{\ell-1} \\ \Rightarrow (I - C_1B_1)[\rho_\ell] = 0 \\ \Rightarrow B_1 \overset{\text{op}}{=} I/C_1 \;\;\&\;\; \rho_1 = B_1^{(\ell-1)}\rho_\ell \\ \Rightarrow \rho_\ell = B_1^{-(\ell-1)}[\rho_1] = C^{(\ell-1)}\rho_1.\end{gathered}$$

For time series, the backward shift operator is often used, although may be awkward. Hence, the backward shift formulation of ***ARMA(p,q)*** is

$$\mathcal{B}_{X,p}(B_1)[X_t] = C_0 + \mathcal{B}_{W,p}(B_1)[W_t],$$

where the ***backward state or autoregressive operator coefficient*** is

$$\mathcal{B}_{X,p}(B_1) \equiv I - \sum_{j=1}^{p} C_j B_1^j$$

and the ***backward noise or moving average operator coefficient*** is

$$\mathcal{B}_{W,q}(B_1) \equiv I - \sum_{j=1}^{q} c_j B_1^j.$$

It is assumed that the $(p+1)$ and $(q+1)$ operator polynomials, $\mathcal{B}_{X,p}(B_1)$ and $\mathcal{B}_{W,q}(B_1)$ do not have common algebraic roots or factors, which would imply that the ARMA order (p,q) could be reduced because of redundancies. The formal inverse to obtain the state is by solving the AR operator part of the ARMA model,

$$X_t = \mathcal{B}^{-1}_{X,p}(B_1)[C_0 + \mathcal{B}_{W,q}(B_1)[W_t]] \,.$$

The state operator is said to be ***time series invertible***,

$$\mathcal{B}^{-1}_{X,p}(z) = \left(1 - \sum_{j=1}^{p} C_j z^j\right)^{-1}$$
$$= 1 + \sum_{j=1}^{p} C_j z^j + \left(\sum_{j=1}^{p} C_j z^j\right)^2 + \dots,$$

using the geometric series $(1/(1-\varepsilon) = 1 + \varepsilon + \varepsilon^2 + \dots)$, if the coefficients $|C_j| < 1$ where z is an ordinary polynomial variable, so that the inverted infinite series is formally convergent, but this only means potentially convergent.

However, notice that the terms represented by z^j are in the form $B_1^k X_t = X_{t-k}$, but if the state is bounded $|X_s| < \beta$ for all s and some positive β, then $|B_1^k X_t| < \beta$ will bound all state powers and only the unit bound on C_j matters.

In the case, that the error of fitting is of interest, then relating the noise to the sample data, the user solve the moving average part of the ARMA model,

$$W_t = -\mathcal{B}_{W,q}^{-1}(B_1)[C_0 - \mathcal{B}_{X,p}(B_1)[X_t]]\,,$$

provided the noise coefficients $|c_j| < 1$ for all j for invertibility of $\mathcal{B}_{W,q}(B_1)$.

Since the weakly stationary times stationary time series has a time-independent mean $\mu_X = \mathrm{E}[X_t]$, then

$$\mu_X = C_0 + \sum_{j=1}^{p} C_j \mu_X = \frac{C_0}{1 - \sum_{j=1}^{p} C_j},$$

provided the denominator is nonvanishing. The autocovariance functions including the variance are quite complicated, so they are omitted here. For the that reason, it is no coincidence that the ***ARMA(p,q)*** models that are preferred for forecasting are those with small p and small q.

For an AR solution expansion by infinite series example, the ARMA(1,1) written in deviations,

$$\delta X_t = C_1 \delta X_{t-1} + W_t - c_1 W_{t-1},$$

produces the formal operator solution of the current state,

$$\delta X_t = (1 - C_1 B_1))^{-1}[W_t - c_1 W_{t-1}],$$

when expanded by the geometric series yields

$$\delta X_t = \sum_{k=0}^{\infty} (C_1 B_1)^k [W_t - c_1 W_{t-1}],$$

or with bounded white noise, i.e., non-Gaussian,

$$\delta X_t = \textstyle\sum_{k=0}^{\infty} C_1^k (W_{t-k} - c_1 W_{t-k-1})$$

$$= W_t + (C_1 - c_1) \sum_{k=0}^{\infty} C_1^k W_{t-k-1},$$

convergent so long as $|C_1| < 1$ by weak stationarity.

For an MA solution expansion by infinite series example, the ARMA(1,1) written with focus on the noise,

$$W_t - c_1 W_{t-1} = \delta X_t - C_1 \delta X_{t-1},$$

produces the formal operator solution,

$$W_t = (1 - c_1 B_1))^{-1}[\delta X_t - C_1 \delta X_{t-1}],$$

when expanded by the geometric series yields

$$W_t = \sum_{k=0}^{\infty} (c_1 B_1)^k [\delta X_t - C_1 \delta X_{t-1}],$$

or with bounded white noise,

$$W_t = \sum_{k=0}^{\infty} c_1^k (\delta X_{t-k} - C_1 \delta X_{t-k-1})$$

$$= \delta X_t + (c_1 - C_1) \sum_{k=0}^{\infty} c_1^k \delta X_{t-k-1},$$

convergent so long as $|c_1| < 1$, not weak stationarity.

Final Remarks on ARMA(p,q) Models:

1. The usual PACF method is not informative in identifying the ARMA according to Tsay ('04), but suggests his Extended ACF (EACF[a]) method that starts with a consistent estimate for the AR part of the ARMA model and then uses the ACF to determine the MA part of the ARMA.
2. Forecasting with the ARMA model is quite similar to that AR model except that the forecast is conditional on the the past state and noise terms.

---

[a] Tsay and Tiao, JASA, vol. 79, pp. 84-96, 1984.

- ***9.5 Conditional Stochastic Volatility Models — ARCH and GARCH:***[a]

From our studies of stochastic volatility surfaces and graphs, we should have seen that volatility is variable over time, i.e., $\sigma = \sigma_t$ or $\sigma^2 = \sigma_t^2$ in terms of the variance, and this dependence is technically called ***heteroscedasticity***. The prefix ***hetero*** means ***different*** in contrast to ***homo*** which means the ***same***, as in ***homoscedasticity***, which means the same variance, e.g., the usual Black-Scholes model assumption.

---

[a] The main background reference for this is Tsay ('04), Chapter 3 *Conditional Heteroscedastic Models*, pp. 97ff, logically reorganized by current editor ($\ddot\smile$).

In economics and finance it is specifically the ***conditional variance***, i.e., $\mathrm{Var}[X|Z] = \mathrm{E}[(X - \mathrm{E}[X|Z])^2|Z]$. Hence, time series models of variance are called ***AutoRegressive Conditional Heteroscedasticity (ARCH = AR+CH)*** models.

A better common term would be ***stochastic variance*** models, but again they are more often called ***stochastic volatility*** even the volatility refers to the standard deviation σ_t while the models are usually for the variance σ_t^2, but sometimes volatility is loosely used to refer to general attributes of the model or the application.

Although, the basic variable is the variance since it is based by definition a sum of squares or integrals of a squared integrand ensuring nonnegativity, while loose volatility terms could lead to problems. In the end the results are often presented as volatility, because then that can easily be translated as percentile of the asset once the time scale has been factored out.

The ***variance or the "volatility" cannot be directly obtained*** from a single observation of the underlying asset since a reasonable sized sample is needed. It is like it is an external variable that is not easy to measure, but having a realistic model for it would help.

Other properties of “volatility” need to be considered. Though ***serial correlations over time*** matter in high to daily frequency observations for the volatility, they have little significance for monthly or lower frequencies. Volatility can also ***cluster over time intervals***, either high or low. If the volatility is high or low the next estimation is likely to be low or high, respectively, but not necessarily, this likelihood is usually called the ***leverage effect***.

Usually, ***volatility is thought to be continuous***, so jumps are thought to be ***rare***, except that regressive techniques are notoriously biased against rare events because likelihood or frequency is what counts in regression with not or little weight on the asset value at the jump.

- ***9.6 Combined Asset Log-Return and Volatility Models — ARMA + ARCH:***

The joint models for the asset log-return X_t and the variance σ_t^2 are is usually two variants of the autoregressive (AR) models, the log-return ARMA(p,q;p$_y$) =ARMA$^+$(p,q,p$_y$), assumed stationary, model,

$$X_t = C_0 + \sum_{j=1}^{p} C_j X_{t-j} + W_t - \sum_{j=1}^{q} c_j W_{t-j} + \sum_{k=1}^{p_y} K_k Y_{t-k}$$

and the heteroscedastic, variance AR(m) or ARCH(m) model,

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{m} \alpha_j W_{t-j}^2 \ \ \& \ \ W_t = \sigma_t \varepsilon_t,$$

where Y_{t-k} other explanatory variables and ε_t are IID RVs.

The ***explanatory variables*** $\boldsymbol{Y_{t-k}}$ are those usually not directly observable or extrinsic, unlike the asset observation, and could be something like the market return $\boldsymbol{r_t^{(\mathrm{mkt})}}$ or perhaps the asset volatility itself $\boldsymbol{\sigma_t}$.

The ***IID noise*** $\boldsymbol{\varepsilon_t}$ is mean-zero $\mathbf{E}[\varepsilon_t] = 0$, unit-variance $\mathbf{Var}[\varepsilon_t] = 1$ and $\mathbf{Cov}[\varepsilon_t, \varepsilon_{t-\ell}] = \delta_{\ell,0}$ by independence. The AR variance coefficients must satisfy $\boldsymbol{\alpha_0 > 0}$ and $\boldsymbol{\alpha_j \geq 0}$ for $\boldsymbol{j > 0}$ to guarantee variance positivity, but also must satisfy some regularity conditions so that the variance remains finite, i.e., $\boldsymbol{0 < \sigma_t^2 < \infty}$. Note, that since $\boldsymbol{\varepsilon_t}$ is an IID RV, then it is functionally independent of $\boldsymbol{t}$, except for a single sample.

The variance equation or quadratic noise equation,

$$W_t^2 = \varepsilon_t^2\left(\alpha_0+\sum_{j=1}^{m}\alpha_j W_{t-j}^2\right) = \widetilde{\alpha}_0+\sum_{j=1}^{m}\widetilde{\alpha}_j W_{t-j}^2,$$

is called an AR(m) model because it is considered a state variable that is autocorrelated to past noise, but except for the plus sign (+) instead of a minus sign (−) it might be an MA(m) equation with squared noise that is still noise.

The ***conditional part*** of the ARCH(m) + ARMA(p,q;p$_y$) model comes from defining the prior information set $\mathcal{F}_{t-1}$ and that is

$$\mathcal{F}_{t-1} \equiv \{\{X_{t-j}, j=1{:}p\}; \{W_{t-j}, j=1{:}\max(m,q)\};$$
$$\{Y_{t-j}, j=1{:}p_y\}\}.$$

Further, define the log-return state mean conditioned on the past, in the ***log-return mean equation***,

$$\mu_t \equiv \mu_{t|t-1} \equiv \mathrm{E}[X_t|\mathcal{F}_{t-1}]$$
$$= C_0 + \sum_{j=1}^{p} C_j X_{t-j} - \sum_{j=1}^{q} c_j W_{t-j} + \sum_{k=1}^{p_y} K_k Y_{t-k},$$

so the log-return asset state equation[a] can be simply written,

$$X_t = \mu_t + W_t.$$

---

[a] Tsay ('04). p. 101, and others call the asset noise, W_t the *innovation* or *shock*, but no one using jump, Poisson white noise would call regular or Gaussian white noise a shock, but innovation is used in a similar way in stochastic filtering theory.

Next, the ***conditional heteroscedastic model is defined in general*** as

$$\sigma_t^2 \equiv \sigma_{t|t-1}^2 \equiv \mathrm{Var}[X_t|\mathcal{F}_{t-1}] = \mathrm{E}[(X_t - \mu_t)^2|\,\mathcal{F}_{t-1}]$$

$$= \mathrm{Var}[W_t|\mathcal{F}_{t-1}].$$

If the noise satisfies an ***ARCH (1)***, i.e., $m = 1$, and since ε_t is IID, then $\mathrm{E}[\varepsilon_t|\mathcal{F}_{t-1}] = \mathrm{E}[\varepsilon_t] = 0$, so $\mathrm{E}[W_t] = \mathrm{E}[\sigma_t]\mathrm{E}[\varepsilon_t] = 0$ and using iterated expectations,

$$\mathrm{Var}[W_t] = \mathrm{E}[W_t^2] = \mathrm{E}[\mathrm{E}[W_t^2|\mathcal{F}_{t-1}]]$$

$$= \mathrm{E}[\mathrm{E}[\sigma_t^2\varepsilon_t^2|\mathcal{F}_{t-1}]] = \mathrm{E}[\mathrm{E}[\sigma_t^2|\mathcal{F}_{t-1}]\mathrm{E}[\varepsilon_t^2]]$$

$$= \mathrm{E}[(\alpha_0 + \alpha_1 W_{t-1}^2)\cdot 1] = \alpha_0 + \alpha_1 E[W_{t-1}^2].$$

However, assuming $W(t)$ is stationary, so $\mathrm{Var}[W_t] = \mathrm{Var}[W_{t-1}] = \mathrm{E}[W_{t-1}^2]$ and

$$\mathrm{Var}[W_t] = \alpha_0 + \alpha_1 \mathrm{Var}[W_t] = \sigma_W^2 = \frac{\alpha_0}{1-\alpha_1},$$

provided $0 \le \alpha_1 < 1$.

One important question is does adding stochastic volatility generate ***fat tails***? In order to find out, it is necessary to know more than the mean and the variance, so we need to know more about the distribution and so we assume that the $\varepsilon_t = \mathcal{N}(0,1)$, i.e., a standard normal or Gaussian, and the model is a Gaussian ARCH(1) model (this is ***not*** the GARCH model).

Checking for skewness, using iterated expectations,

$$\mathbf{E}[W_t^3] = \mathbf{E}[\mathbf{E}[W_t^3|\mathcal{F}_{t-1}]] = \mathbf{E}[\mathbf{E}[\sigma_t^3\varepsilon_t^3|\mathcal{F}_{t-1}]]$$

$$= \mathbf{E}[\mathbf{E}[\sigma_t^3|\mathcal{F}_{t-1}]\mathbf{E}[\varepsilon_t^3]] = 0,$$

the third moment is skewless due to ε_t^3 being odd. Checking for kurtosis, high peaks and fat tails,

$$\mathbf{E}[W_t^4] = \mathbf{E}[\mathbf{E}[W_t^4|\mathcal{F}_{t-1}]] = \mathbf{E}[\mathbf{E}[\sigma_t^4\varepsilon_t^4|\mathcal{F}_{t-1}]]$$

$$= \mathbf{E}[\sigma_t^4\mathbf{E}[\varepsilon_t^4]] = 3\mathbf{E}[(\alpha_0 + \alpha_1 W_{t-1}^2)^2]$$

$$= 3\mathbf{E}[\alpha_0^2 + 2\alpha_0\alpha_1 W_{t-1}^2 + \alpha_1^2 W_{t-1}^4]$$

$$= 3(\alpha_0^2 + 2\alpha_0\alpha_1\mathrm{Var}[W_{t-1}] + \alpha_1^2\mathbf{E}[W_{t-1}^4]$$

$$\overset{\text{alg}}{=} 3\alpha_0^2(1+\alpha_1)/((1-\alpha_1)(1-3\alpha_1^2)),$$

provided $0 \le \alpha_1^2 < 1/3 < 1$, where $\mathbf{E}[\varepsilon_t^4] = 3$ was used.

Thus, the ***coefficient of Kurtosis***[a] is

$$\eta_W^{(4)} \equiv \frac{\mathrm{E}[W_t^4]}{(\mathrm{E}[W_t^2])^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} = 3 + \frac{6\alpha_1^2}{1-3\alpha_1^2} > 3,$$

the normal value, verifying the fat tail property if $|\alpha_1| > 0$. For example, if $\alpha_1^2 = 1/6$, then $\eta_W^{(4)} = 5$, or the excess kurtosis is 2.

Since fat tails can be generator even from Gaussian IID noise, the asset noise in non-normal and the likelihood of ***large or outlier log-returns*** is more likely than than if W_t were normal.

---

[a] Tsay, p. 105, has an error in first few lines for $\mathrm{E}[W_t^4]$ derivation, assuming W_t is normal, instead of ε_t, contradicting the kurtosis result for $W(t)$; however, his final answer is correct.

- ***9.7 Test for ARCH Effects:***

By ARCH effect is meant the conditional variance or heteroscedasticity (CH). For the asset mean equation, this means testing the residuals $W_t = X_t - \mu_t$, using the above W_t-squared series on p. L9.50.

One test is the previously mentioned Ljung-Box serial correlation test can be used on W_t^2 and the $Q(m)$ statistic appears here on p. L9.25, while the nulls hypothesis is the $W_{t-j}^2$are zero for $j = 1{:}m$.

Another test that can be used for asset residuals is Engle's[a] heteroscedasticity Lagrange multiplier (LM) test, equivalent to an F test, testing whether whether the null hypothesis that the coefficients $\widetilde{\alpha}^2_{t-j}$ are zero for $j = 1\!:\!m$, where $W_t^2 = \widetilde{\alpha}_0 + \sum_{j=1}^{m} \widetilde{\alpha}_j W_{t-j}^2 + e_t$ for $t = m+1\!:\!T$. Here e_t is the error, another residual, and T is the time sample size, so $\overline{W^2} = \sum_{j=1}^{T} W_t^2/T$ is the sample mean, $\mathrm{SSRes}_0[W_t^2] = \sum_{j=m+1}^{T}(W_t^2 - \overline{W^2})^2$ the residual sum of squares and similarly $\mathrm{SSRes}_1[\widehat{e}_t^2] = \sum_{j=m+1}^{T} W_t^2$, where $\widehat{e}_t$ is residual of the prior least squares residual.

---

[a] R. F. Engle, Econometrica, vol. 50, pp. 987-1007; this is the ARCH model and test paper by its developer and 2003 Noble Laureate in Economics.

The corresponding Engle F statistic[a] is

$$F = \frac{\left(\mathrm{SSRes}_0\left[W_t^2\right] - \mathrm{SSRes}_1\left[\widehat{e}_t^2\right]\right)/m}{\mathrm{SSRes}_1\left[\widehat{e}_t^2\right]/(T-2m-1)} \sim \chi_m^2,$$

such that the null hypothesis is rejected if $F > (\chi_m^2)^{-1}(1-\alpha)$ [b], or $p = 1 - \chi_m^2(F) \le \alpha$, where $F^* = (\chi_m^2)^{-1}(1-\alpha)$ is the upper $100(1-\alpha)$th percentile of $\chi_m^2(F)$, m is the number of degrees of freedom, and α is the significance level .

---

[a] See Tsay's ('04) book, p. 102.

[b] In MATLAB, $(\chi_m^2)^{-1}(1-\alpha)$ = `chi2inv(1-alpha,m)` is the chi-square table entry with Engle F-test value `p = 1-chi2cdf(F,m)`.

Final Remarks on ARCH(m) Models:

1. The response of ARCH(m) model to large gains and losses is restrict by the squared-W_t series the model since the market behavior is very different in response to large positive and negative changes. Use of regression type models also makes the models less sensitive to large deviations or outliers by the averaging out rare events, so may misestimate volatility.
2. The lag-m order identification can be performed through the partial ACF (PACF) techniques that we looked at for AR(p) models in the Lecture 8.

3. Coefficient estimation can be done by maximum likelihood estimation (MLE) using conditional densities corresponding to the ARCH formulation along with much approximations to make the computation feasible[a].

4. Forecasting is similar to either AR(p) models or ARMA(p,q) models[b].

---

[a] See Tsay ('04), pp. 197-108.

[b] See Tsay ('04), p. 109.

- ***9.8 Generalized ARCH (GARCH) Models:***

The ***Generalized ARCH (GARCH) model*** of Bollerslev[a] depends on s prior values of σ_t^2 in the ***GARCH(m,s)*** order model,with the usual asset log-return residual or innovation $W_t = X_t - \mu_t = \sigma_t \varepsilon_t$ for IID mean-zero and unit-variance, is

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{m} \alpha_j W_{t-j}^2 + \sum_{k=1}^{s} \beta_k \sigma_{t-k}^2,$$

where again $\alpha_0 > 0$ and $\alpha_j \geq 0$ for the ARCH parameters, while $\beta_k \geq 0$ for the GARCH parameters, and $\sum_{j=1}^{m} \alpha_j + \sum_{k=1}^{s} \beta_k < 1$, to ensure finiteness. We assume GARCH(m,0) = ARCH(m).

---

[a] T. Bollerslev, Ecnometrics, vol. 31, pp. 307-327.

Following the same procedure*, as used for ARRCH(1) assuming that the IID ε_t are normally distributed*, for simplest, nontrivial ***GARCH(1,1)*** ,

$$\sigma_t^2 = \alpha_0 + \alpha_1 W_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

as for ***ARCH(1,1)***, $\mathrm{E}[W_t] = 0$, $\mathrm{E}[W_t^3] = 0$,

$$\begin{aligned}\mathrm{Var}[W_t] = \quad & \mathrm{E}[W_t^2] = \mathrm{E}[W_t^2] = \mathrm{E}[\sigma_t^2] \\ &= \alpha_0 + \alpha_1 \mathrm{E}[W_{t-1}^2] + \beta_1 \mathrm{E}[\sigma_{t-1}^2] \\ = \quad & \alpha_0/(1-(\alpha_1+\beta_1)),\end{aligned}$$

provided $(1-(\alpha_1+\beta_1)) > 0$

For the 4th order moment,

$$\begin{aligned}\mathrm{E}[W_t^4] = 3\mathrm{E}[\sigma_t^4] &= 3\mathrm{E}[(\alpha_0+\alpha_1 W_{t-1}^2+\beta_1\sigma_{t-1}^2)^2] \\ &= 3(\alpha_0^2+2\alpha_0\alpha_1\mathrm{E}[W_{t-1}^2]+2\alpha_0\beta_1\mathrm{E}[\sigma_{t-1}^2] \\ &\quad +\alpha_1^2\mathrm{E}[W_{t-1}^4]+2\alpha_1\beta_1\mathrm{E}[W_{t-1}^2\sigma_{t-1}^2] \\ &\quad +\beta_1^2\mathrm{E}[\sigma_{t-1}^4]) \\ &= \frac{3\alpha_0^2(1+\alpha_1+\beta_1)}{((1-(\alpha_1+\beta_1))(1-(2\alpha_1^2+(\alpha_1+\beta_1)^2))},\end{aligned}$$

provided $(1-(2\alpha_1^2+(\alpha_1+\beta_1)^2))>0$. Hence the kurtosis coefficient is

$$\begin{aligned}\eta_W^{(4)} &= \frac{3(1-(\alpha_1+\beta_1)^2)}{1-(2\alpha_1^2+(\alpha_1+\beta_1)^2)} \\ &= 3+\frac{6\alpha_1^2}{1-(2\alpha_1^2+(\alpha_1+\beta_1)^2)} > 3,\end{aligned}$$

again finding a fat tail property, ***the same result as Tsay, p. 114***.

That is about all the time we have, but only an introduction to time series was intended. Have a productive financial career.