

BASIC MULTIPLE REGRESSION MODEL

$$Y_t = a_0 + a_1X_{1t} + a_2X_{2t} + \dots + a_pX_{pt} + e_t$$

The Regression Model is:

Linear in a's, X's, and e's.

Static - all variables are from the same time period.

Multivariate - Dependent variable Y is explained by several variable X's.

Limitations of the Basic Regression Model:

Can't Forecast Ahead - the X's aren't available until you have already seen the Y's, which precludes forecasting.

The Error Terms (e's) Must "Behave" - they can't trend, vary widely in size, be correlated with the X's, etc.

BASIC AUTOREGRESSION MODEL

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_p Y_{t-p} + e_t$$

The Autoregression Model:

Uses a Dynamic Structure - by using lags of the dependent variable, forecasts can be made before the current time period occurs.

Has a Reasonable Overall Interpretation - a forecast of Y is a weighted sum of past observed values of the same variable Y, suggesting people make their forecasts by considering recent observed values.

Requires No Sophisticated Procedures - the same methods used to calculate basic multiple regression can solve an autoregression.

Limitations of the Autoregression Model:

The Error Terms (e's) Must "Behave" - they can't trend, vary widely in size, be correlated with the lagged Y's, etc.

No Interpretation to Individual Parameters - in general, the only purpose for autoregression is forecasting. Unlike basic regression, you can't attach economic meaning to the a's.

Easy to Overfit - adding more lagged values of Y often improves the measures of "fit", but may not lead to better forecasts in practice.

BASIC ARMA (aka BOX-JENKINS) MODEL

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_p Y_{t-p} + u_t$$
$$u_t = e_t + b_1 e_{t-1} + b_2 e_{t-2} + \dots + b_q e_{t-q}$$

The ARMA Model:

Is Dynamic in Both Variables and Errors - a lag structure is imposed on both the explanatory variables and the error terms, but the model remains linear in the Y's, a's, b's, and e's.

Has a Reasonable Economic Interpretation - a forecast of Y is the sum of two weighted sums: a weighted sum of past observed values of Y, and a weighted sum of past forecast errors.

Remains Easy to Implement - given past values of Y, the model remains easy to forecast once estimates of the a's and b's have been generated.

Limitations of the ARMA Model:

The Error Terms (u's) Must Still "Behave": certain types of pattern-like behavior can now be dealt with in the u's (e.g., trends and mean-reversions), but they can't vary widely in volatility.

Requires Sophisticated Estimation Procedures - unlike regression and autoregression, requires much more sophisticated optimization algorithms to get estimates of a's and b's.

Easy to Overfit - The choice of how many lags of Y (p) and how many lags of e (q) is very subjective, and it is easy to overfit an ARMA model.

BASIC ARCH/GARCH MODEL

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_r Y_{t-r} + e_t$$

$$e_t \sim N(0, \sigma_t)$$

$$\sigma_t = c_0 + c_1 e_{t-1}^2 + \dots + c_q e_{t-q}^2 + d_1 \sigma_{t-1} + \dots + d_p \sigma_{t-p}$$

The ARCH/GARCH Model:

Is Fundamentally a Basic Regression Model - a GARCH model is simply an ordinary regression (or autoregression) model where allowance is made for very complicated patterns in the error terms.

Has Two Separate Forecasting Equations - the first equation, when estimated, can be used to forecast future values of Y; the last equation, when estimated, can be used to forecast moves in Y's volatility.

Errors Have Changing Volatility - The volatility of the error term is allowed to change via a model that looks like an ARMA equation.

Limitations of the ARCH/GARCH Model:

Requires Sophisticated Estimation Procedures - unlike regression and autoregression, requires much more sophisticated optimization algorithms to get estimates of a's, c's, and d's. Also, the estimation procedures often crash and give no answer.

Easy to Overfit - With three different lag structures to pick (r, q, and p), it is even easier to overfit an ARCH/GARCH model than an ARMA model.

Volatility Forecasts Can be Unusual - Unless special steps are taken, volatility forecasts can be negative, or can grow over time to unrealistic levels.

PERFORMING GARCH IN SAS

The following SAS program performs a GARCH(1,1) model on the log price ratio of the S&P futures contract, from 1/2/85 to 4/27/95.

This file can be found as X:\MARK\MRKEXMPL.SAS. Items in *italics* are comments that do not need to be entered into SAS (and are not in the file). Items in **bold** need to be entered into SAS but they will be different for your specific data set. Everything else must be entered as written.

/ Some preliminary housekeeping things SAS must do */*

filename **smplfile** 'x:\mark\mrkexmpl.prn'; *-> name the datafile that contains your input*
filename **output** 'x:\mark\output.dat'; *-> name the datafile to contain any output*

/ pull in the data and do any necessary transformations */*

```
data mark;  
    infile smplfile flowover;  
    input date :mmddyy. intra impl volm lpr sd20 mon tue wed thu fri;  
    lvolm = log(volm);  
    dlvolm = dif(lvolm);  
run;
```

/ the following procedure performs the actual GARCH estimation */*

```
proc autoreg data = mark ;  
    model lpr = mon tue wed thu fri dlvolm /nlag=2 noint  
        garch=(p=1, q=1, type=noineq) maxit=50;  
    output out=mrkgarch p=lprhat cev=volhat;  
run;
```

/ the rest of the program writes any output or forecasts to the file named output */*

```
data _null_;  
    set mrkgarch;  
    file output lrecl=500;  
    if _n_ =1 then put "date" " lpr" " lprhat" " intra" " impl" " sd20" " volhat";  
    volhat=sqrt(volhat*252)*100;  
    put date lpr lprhat intra impl sd20 volhat;  
    format date date.;;  
run;
```

Your program may use other variables, other filenames, do other variable transformations, use different numbers of GARCH lags, use a different date format, forecast different variables, etc.