What Practitioners Need to Know . . .

by Mark Kritzman

. . . About Estimating Volatility

Part 2

Circle the correct answer:

Autoregressive Conditional Heteroscedasticity (ARCH) is:

(1) a computer programming malfunction in which a variable is assigned more than one definition, causing the program to lock into an infinite loop;
(2) a psychological disorder characterized by a reversion to earlier behavior patterns when confronted with unpleasant childhood memories;
(3) in evolution, a reversion to a more primitive life form caused by inadequate diversity within a species;
(4) a statistical procedure in which the dependent variable in a regression equation is modeled as a function of the time-varying properties of the error term.

To the uninitiated, all the above definitions might seem equally plausible. Moreover, the correct definition, (4), may still not yield an intuitively satisfying description of ARCH.

This column is intended as a child's guide to ARCH, which is to say that it contains no equations. Our goal is to penetrate the cryptic jargon of ARCH so that at the very least you will feel comfortable attending social events hosted by members of the American Association of Statisticians. Of course, you should not expect that familiarity with ARCH will actually cause you to have fun at such events.

Normal Assumptions of Volatility

In the July/August issue of this journal, we reviewed two procedures for estimating volatility—one by which volatility is estimated directly from historical observations and an alternative procedure by which we infer volatility from the prices at which options on the underlying assets trade. One of the implicit assumptions of both these approaches is that volatility is stable.[1] Stability, in this context, does not suggest that volatility remains constant through time. Rather, it implies that volatility changes unpredictably, or that it is uncorrelated with previous levels of volatility.

For example, if we were to calculate the differences between monthly stock returns and their mean and regress the squared values of these differences (called the errors squared) in month t against their values in month t-1, we would not expect to detect a significant relationship. More specifically, the intercept of the regression line should be close to the average value of the errors squared, and the slope of the regression line should not differ significantly from zero.[2] Figure A represents this relationship impressionistically.

The parameters of the linear regression model depicted in Figure A would probably suggest that the errors squared are not autocorrelated, because the slope of the regression line is flat and the intercept seems close to the average value of the observations. We should not necessarily assume, however, that the errors squared are serially independent (uncorrelated with their prior values) simply because the regression parameters from a linear regression are insignificant. We must look further and examine the residuals around the fitted values.

The residuals equal the difference between the actual values for the errors squared and the values predicted from the regression line. If the errors squared were serially independent, the residuals would be distributed randomly around an expected value of zero. If the residuals satisfied this condition, we would describe them as "homoscedastic."

Nonlinearity

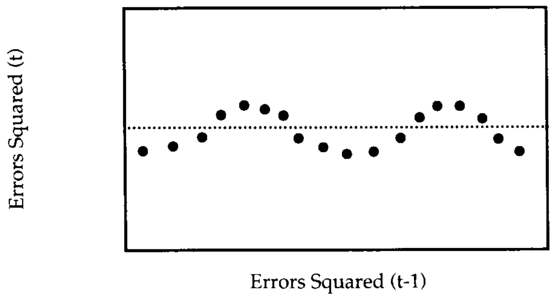
Suppose that high values for the errors squared occur in clusters. It might still be the case that the coefficients from the linear regression of the errors squared on their prior values are insignificant. If, however, the errors squared are related to their prior values in some nonlinear fashion, this nonlinear relationship might be revealed by the patterns formed by the residuals around the fitted values.

Reexamining Figure A, for example, we notice that positive residuals (where the regression line underestimates the actual errors squared) tend to be followed by more positive residuals, and that negative

1. Footnotes appear at end of article.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure A Variance as a Function of its Prior Values



residuals appear in groups as well. We refer to such patterns in the residuals as "heteroscedasticity."

We can correct for this apparent nonlinearity by regressing the residuals in period t on the errors squared in period t-1. We then add the intercept and slope from this regression equation to the intercept and slope from the original regression of the errors squared on their prior values. In effect, we conjecture that variance—the average value of the errors squared—is conditioned on this heteroscedasticity. This explains the term, "Autoregressive Conditional Heteroscedasticity."

The procedure is summarized below.

(1) Subtract the observed returns from their mean and square these differences to calculate the errors squared.
(2) Regress the errors squared in period t on the errors squared in period t-1.
(3) Subtract the fitted errors squared in period t from the observed errors squared in period t-1.
(4) Regress the residuals from Step (3) in period t on the errors squared in period t-1.
(5) Add the intercept and slope from the regression equation in Step (4) to the intercept and slope from the regression equation in Step (2) to form a new prediction equation for variance.

The resulting equation should be a more efficient predictor of variance that the equation resulting from the original regression model of the errors squared on their prior values, to the extent that the residuals from the original model are heteroscedastic. However, the process we have just described is not precisely the process that is used in an ARCH model.

Purists would recommend that one divide both the left-hand side and the right-hand side of the regression equation in Step (4) by the fitted values for the errors squared from the regression equation in Step (2). This is referred to as a generalized-least-squares method. For a more rigorous description of ARCH and related models, see the references at the end of this column.

Summary

Variance may be related to its past values. However, we may fail to detect this autoregressive relationship with a linear regression model of the errors squared if the relationship is nonlinear. If we have reason to suspect a nonlinear relationship, we should test the residuals between the actual values for the errors squared and the fitted values from an autoregressive equation for heteroscedasticity. If there is significant heteroscedasticity in the residuals, we should model the heteroscedasticity and incorporate it in our forecast equation for variance.

As one would expect, ARCH has spawned several other related procedures for modeling heteroscedasticity. We now have GARCH (Generalized ARCH), which includes lagged values for the dependent variable as well as the error term, and EGARCH (Exponential GARCH), which uses the logs of the variables. These are but two of several variations on the ARCH theme. As a contribution to the proliferation of ARCH-related acronyms, it is only fitting that we apply the acronym, PARCH, to the heuristic procedure described above; PARCH, of course, stands for "Poor man's ARCH."[3]

Footnotes

1. Is this an oxymoron?
2. For a review of regression analysis, see M. Kritzman, "What Practitioners Need to Know About Regressions," *Financial Analysts Journal*, May/June 1991.
3. I have benefited from discussions with Stephen Brown, Bruce Lehman, William Margrabe, Dan Nelson and Krishna Ramaswamy, although it is quite likely that none of them would necessarily approve of my attempt to distill the complexities of ARCH. Obviously they are exonerated from any errors or deficiency of rigor in this note.

Further References

Bollerslev, Tim, Ray Chou, Narayaman Jayaraman and Kenneth Kroner, "ARCH Modeling in Finance: A Selective Review of the Theory and Empirical Evidence, with Suggestions for Future Research" (Working paper presented at the UCSD Conference on Modeling Volatility in Financial Markets, April 6–7 1990).

Engle, Robert, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica* 50 (1982), pp. 987–1008.

Greene, William H., *Econometric Analysis* (New York: MacMillan Publishing, 1990), pp. 416–419.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.