ARCH and MGARCH models

Christopher F Baum

EC 823: Applied Econometrics

Boston College, Spring 2013

ARCH models

Heteroskedasticity can occur in time series models, just as it may in a cross-sectional context. It has the same consequences: the OLS point estimates are unbiased and consistent, but their standard errors will be inconsistent, as will hypothesis test statistics and confidence intervals.

We may prevent that loss of consistency by using heteroskedasticity-robust standard errors. The "Newey–West" or HAC standard errors available from `newey` in the OLS context or `ivreg2` in the instrumental variables context will be robust to arbitrary heteroskedasticity in the error process as well as serial correlation.

The most common model of heteroskedasticity employed in the time series context is that of *autoregressive conditional heteroskedasticity*, or ARCH. As proposed by Nobel laureate Robert Engle in 1982, an ARCH model starts from the premise that we have a static regression model

$$y_t = \beta_0 + \beta_1 z_t + u_t$$

and all of the Gauss–Markov assumptions hold, so that the OLS estimators are BLUE. This implies that $Var(u_t|Z)$ is constant. But even when this unconditional variance of u_t is constant, we may have time variation in the *condiitional variance* of u_t:

$$E(u_t^2|u_{t-1}, u_{t-2}, \dots) = E(u_t^2|u_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2$$

so that the conditional variance of u_t is a linear function of the squared value of its predecessor.

If the original u_t process is serially uncorrelated, the variance conditioned on a single lag is identical to that conditioned on the entire history of the series. We can rewrite this as

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2$$

where $u_t = \sqrt{h_t}\, v_t$, $v_t \sim (0, 1)$. This formulation represents the ARCH(1) model, in which a single lagged u^2 enters the ARCH equation. A higher-order ARCH equation would include additional lags of u^2. To ensure a positive variance, $\alpha_0 > 0$ and $\alpha_1 > 0$. When $\alpha_1 > 0$, the squared errors are positively serially correlated even though the u_t themselves are not.

Since we could estimate this equation and derive OLS *b* which are BLUE, why should we be concerned about ARCH? First, we could derive consistent estimates of *b* which are asymptotically more efficient than the OLS estimates, since the ARCH structure is no longer a linear model.

Second, the dynamics of the conditional variance are important in many contexts: particularly financial models, in which movements in volatility are themselves important. Many researchers have found "ARCH effects" in higher-frequency financial data, and to the extent to which they are present, we may want to take advantage of them. We may test for the existence of ARCH effects in the residuals of a time series regression by using the command `estat archlm`. The null hypothesis is that of no ARCH effects; a rejection of the null implies the existence of significant ARCH effects, or persistence in the squared errors.

The ARCH model is inherently nonlinear. If we assume that the u_t are distributed Normally, we may use a maximum likelihood procedure such as that implemented in Stata's `arch` command to jointly estimate its mean and conditional variance equation.

The ARCH model has been extended to a generalized form which has proven to be much more appropriate in many contexts. In the simplest example, we may write

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \gamma_1 h_{t-1}$$

which is known as the GARCH(1,1) model since it involves a single lag of both the ARCH term and the conditional variance term. We must impose the additional constraint that $\gamma_1 > 0$ to ensure a positive variance.

We may also have a so-called ARCH-in-mean model, in which the h_t term itself enters the regression equation. This sort of model would be relevant if we had a theory that suggests that the level of a variable might depend on its variance, which may be very plausible in financial markets contexts or in terms of, say, inflation, where we often presume that the level of inflation may be linked to inflation volatility. In such instances we may want to specify a ARCH- or GARCH-in-mean model and consider interactions of this sort in the conditional mean (level) equation.

Alternative GARCH specifications

A huge literature on alternative GARCH specifications exists; many of these models are preprogrammed in Stata's `arch` command, and references for their analytical derivation are given in the Stata manual.

One of particular interest is Nelson's (1991) exponential GARCH, or EGARCH. He proposed:

$$\log h_t = \eta + \sum_{j=1}^{\infty} \lambda_j \left(\left| \epsilon_{t-j} \right| - E \left| \epsilon_{t-j} \right| + \theta \epsilon_{t-j} \right)$$

which is then parameterized as a rational lag of two finite–order polynomials, just as in Bollerslev's GARCH.

Advantages of the EGARCH specification include the positive nature of h_t irregardless of the estimated parameters, and the asymmetric nature of the impact of innovations: with $\theta \neq 0$, a positive shock will have a different effect on volatility than will a negative shock, mirroring findings in equity market research about the impact of "bad news" and "good news" on market volatility. For instance, a simple EGARCH(1,1) model will provide a variance equation such as

$$\log h_t = -\delta_0 + \delta_1 z_{t-1} + \delta_2 \left| z_{t-1} - \sqrt{2/\pi} \right| + \delta_3 \log h_{t-1}$$

where $z_t = \epsilon_t / \sigma_t$, which is distributed as $N(0, 1)$.

Nelson's model is only one of several extensions of GARCH that allow for asymmetry, or consider nonlinearities in the process generating the conditional variance: for instance, the threshold ARCH model of Zakoian (1990) and the Glosten et al. model (1993).

Stata 12 provides a suite of commands to estimate time series models in the ARCH (Autoregressive Conditional Heteroskedasticity) family. The command `arch` is used to estimate single-equation models. Its options allow the specification of over a dozen models from the literature, including ARCH, GARCH, ARCH-in-mean, GARCH with ARMA errors, EGARCH (exponential GARCH), TARCH (threshold ARCH), GJR (Glosten et al., 1993), SAARCH (simple asymmetric ARCH), PARCH (power ARCH), NARCH (nonlinear ARCH), APARCH (asymmetric power ARCH) and NPARCH (nonlinear power ARCH).

Errors may be specified as Gaussian, t, or GED (generalized error distribution).

To estimate an ARCH model, you give the `arch` *varname* command, followed by (optionally) the independent variables in the mean equation and the options indicating the type of model. For instance, to fit a GARCH(1,1) to the mean regression of `cpi` on `wage`,

```
arch cpi wage, arch(1) garch(1)
```

It is important to note that a GARCH(2,1) model would be specified with the option `arch(1/2)`. If the option was given as `arch(2)`, only the second-order term would be included in the conditional variance equation.

A test for ARCH effects in a linear regression can be conducted with the `estat archlm` command. Using Stata's `urate` dataset of monthly unemployment rates for several US states:

```
. webuse urates, clear
. qui reg D.tenn LD.tenn
. estat archlm, lags(3)
LM test for autoregressive conditional heteroskedasticity (ARCH)
```

lags(p)	chi2	df	Prob > chi2
3	11.195	3	0.0107

H0: no ARCH effects   *vs.*   H1: ARCH(p) disturbance

The LM test indicates the presence of significant ARCH effects.

We estimate a GARCH(1,1) model:

```
. arch D.tenn LD.tenn, arch(1) garch(1) nolog vsquish
```

ARCH family regression

Sample: 1978m3 – 2003m12   Number of obs   =   310
Distribution: Gaussian   Wald chi2(1)   =   9.39
Log likelihood = 127.4172   Prob > chi2   =   0.0022

D.tenn	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf. Interval]	
tenn						
tenn						
LD.	.2129528	.0694996	3.06	0.002	.076736	.3491695
_cons	−.0155809	.0085746	−1.82	0.069	−.0323868	.0012251
ARCH						
arch						
L1.	.1929262	.0675544	2.86	0.004	.0605219	.3253305
garch						
L1.	.7138542	.0923551	7.73	0.000	.5328415	.894867
_cons	.0028566	.0016481	1.73	0.083	−.0003736	.0060868

Following estimation, we may use `predict` with the `variance` option to produce the conditional variance series.

Conditional variance from GARCH(1,1)

We may also fit a model with additional variables in the mean equation:

```
. arch D.tenn LD.tenn LD.indiana LD.arkansas, arch(1) garch(1) nolog vsquish
ARCH family regression
Sample: 1978m3 - 2003m12                          Number of obs   =        310
Distribution: Gaussian                            Wald chi2(3)    =      41.31
Log likelihood =  135.1611                        Prob > chi2     =     0.0000
```
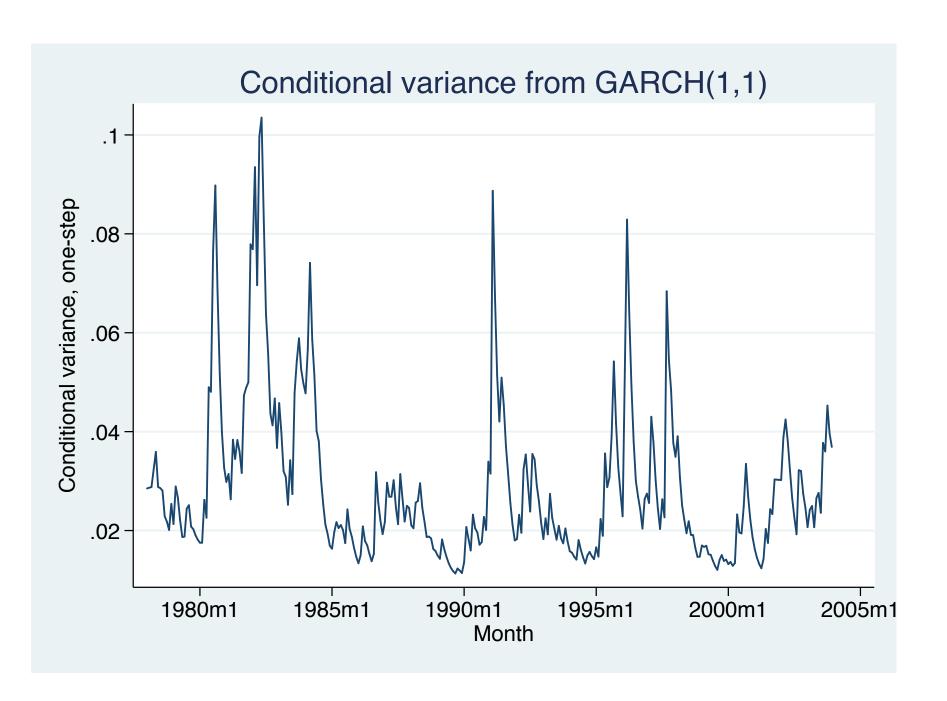
D.tenn	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf. Interval]	
tenn						
tenn LD.	.1459972	.0723994	2.02	0.044	.004097	.2878974
indiana LD.	.1751591	.047494	3.69	0.000	.0820727	.2682455
arkansas LD.	.1170958	.0757688	1.55	0.122	-.0314083	.2655999
_cons	-.0078106	.0087075	-0.90	0.370	-.0248769	.0092558
ARCH						
arch L1.	.1627143	.0712808	2.28	0.022	.0230064	.3024221
garch L1.	.6793291	.1388493	4.89	0.000	.4071896	.9514687
_cons	.0042064	.0026923	1.56	0.118	-.0010704	.0094832

Following estimation, we may test hypotheses on the coefficients of the conditional variance equation: for instance, that they sum to unity, indicating integrated GARCH:

```
. test [ARCH]L.arch + [ARCH]L.garch == 1
( 1)   [ARCH]L.arch + [ARCH]L.garch = 1
          chi2(  1) =     2.30
        Prob > chi2 =     0.1297
```

In this case, that hypothesis cannot be rejected at 90%.

Multiple-equation GARCH models

Multivariate GARCH models allow the conditional covariance matrix of the dependent variables to follow a flexible dynamic structure and allow the conditional mean to follow a vector autoregressive (VAR) structure.

The general MGARCH model can be written as

$$
\begin{aligned}
y_t &= Cx_t + \varepsilon_t \\
\varepsilon_t &= H_t^{1/2} \nu_t
\end{aligned}
$$

where y_t is a m-vector of dependent variables, C is a $m \times k$ parameter matrix, x_t is a k-vector of explanatory variables, possibly including lags of y_t, $H_t^{1/2}$ is the Cholesky factor of the time-varying conditional covariance matrix H_t, and ν_t is a m-vector of zero-mean, unit-variance i.i.d. innovations.

In this general framework, H_t is a matrix generalization of univariate GARCH models. For example, a general MGARCH(1,1)) model may be written as:

$$\text{vech}(H_t) = s + A\,\text{vech}(\varepsilon_{t-1}\varepsilon'_{t-1}) + B\,\text{vech}(H_{t-1})$$

where the *vech*$(\cdot)$ function returns a vector containing the unique elements of its matrix argument. The various parameterizations of MGARCH provide alternative restrictions on H, the conditional covariance matrix, which must be positive definite for all t.

Implementation

Stata's `mgarch` command estimates multivariate GARCH models, allowing both the conditional mean and conditional covariance matrix to be dynamic. Four commonly used parameterizations are supported:

- the diagonal vech (DVECH) model
- the constant conditional correlation (CCC) model
- the dynamic conditional correlation (DCC) model
- the varying conditional correlation (VCC) model

Alternative parameterizations differ in terms of flexibility, allowing for more complex H processes, and parsimony, allowing the model to be specified with fewer parameters.

The oldest and simplest parameterization is the *diagonal vech* (DVECH) of Bollerslev, Engle, Wooldridge (*JPE*, 1988), which restricts the A and B matrices to be diagonal. The number of parameters grows rapidly with the size of the model. For instance, there are $3m(m+1)/2$ parameters in a $DVECH(1, 1)$ with m series.

Despite the large number of parameters, the diagonal structure implies that each conditional variance and covariance depends only on its own past, and not on past values of other elements. For a DVECH(1,1),

$$h_{ij,t} = s_{ij} + a_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1} + b_{ij}h_{ij,t-1}$$

Conditional correlation models

Conditional correlation (CC) models use nonlinear combinations of univariate GARCH models to represent the conditional covariances in H. They often have less difficulty with satisfying the restrictions on the estimated H, and their number of parameters grows more slowly than in the DVECH specification.

In CC models, H_t is decomposed into a matrix of conditional correlations R_t and a diagonal matrix of conditional variances, D_t:

$$H_t = D_t^{1/2} R_t D_t^{1/2}$$

implying that $h_{ij,t} = \rho_{ij,t} \sigma_{i,t} \sigma_{j,t}$, where $\sigma_{i,t}$ is modeled as a univariate GARCH process. The CC models differ in how they parameterize R_t.

The constant CC model of Bollerslev (*REStat*, 1990) specifies the correlation matrix as time invariant:

$$h_{ij,t} = \rho_{ij}\sqrt{h_{ii,t}h_{jj,t}}$$

where the diagonal elements follow univariate GARCH processes, and ρ_{ij} is a time-invariant weight.

Engle's (*JBES*, 2002) extension, the dynamic CC model, allows the conditional correlations (technically, quasicorrelations) to follow a GARCH(1,1)-like process:

$$h_{ij,t} = \rho_{ij,t}\sqrt{h_{ii,t}h_{jj,t}}$$

where now the ρ parameters follow a dynamic process.

Tse and Tsui's (*JBES*, 2002) variant, the varying CC model, expresses the conditional correlations using a time-invariant component, a measure of recent correlations among the residuals, and last period's values. It differs from the DCC model in terms of the dynamic process followed by the ρ parameters.

In Stata, the four MGARCH specifications are invoked with the `mgarch` command, with a first argument being the model specification: `dvech`, `ccc`, `dcc` or `vcc`.

To illustrate, we use Stata's `stocks` dataset, and model daily Toyota and Honda equity returns as AR(1) processes with the `ccc` and `dcc` specifications.

The estimated Toyota mean and conditional variance equations:

```
. webuse stocks, clear
(Data from Yahoo! Finance)

. mgarch ccc (toyota honda = L.toyota L.honda), arch(1) garch(1) nolog vsquish

Constant conditional correlation MGARCH model
```

```
Sample: 1 - 2015                                   Number of obs   =      2014
Distribution: Gaussian                             Wald chi2(4)    =      4.34
Log likelihood =  11602.61                         Prob > chi2     =    0.3620
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
toyota						
toyota						
L1.	-.03374	.032697	-1.03	0.302	-.097825	.030345
honda						
L1.	-.005188	.0288975	-0.18	0.858	-.0618261	.0514502
_cons	.0004523	.0003094	1.46	0.144	-.0001542	.0010587
ARCH_toyota						
arch						
L1.	.0661046	.0095018	6.96	0.000	.0474814	.0847279
garch						
L1.	.916793	.0117942	77.73	0.000	.8936769	.9399092
_cons	4.50e-06	1.19e-06	3.78	0.000	2.17e-06	6.83e-06

...

The estimated Honda mean and conditional variance equations, and correlation estimate:

honda						
toyota						
L1.	−.0066352	.0343028	−0.19	0.847	−.0738675	.0605971
honda						
L1.	−.0332976	.0316213	−1.05	0.292	−.0952743	.028679
_cons	.0006128	.0003394	1.81	0.071	−.0000524	.0012781
ARCH_honda						
arch						
L1.	.0498417	.0080311	6.21	0.000	.0341009	.0655824
garch						
L1.	.9321435	.0111601	83.52	0.000	.9102701	.9540168
_cons	5.26e−06	1.41e−06	3.73	0.000	2.50e−06	8.02e−06
Correlation						
toyota						
honda	.7176095	.0108477	66.15	0.000	.6963483	.7388707

In this CCC specification, the sizable correlation indicates the interaction between the two equations' error processes.

In the DCC model, the diagonal elements of H_t are modeled as univariate GARCH models. The off-diagonal elements are modeled as nonlinear functions of the diagonal terms:

$$h_{ij,t} = \rho_{ij,t}\sqrt{h_{ii,t}h_{jj,t}}$$

where $\rho_{ij,t}$ follows a dynamic process, rather than being constrained to be constant as in the CCC specification.

Two additional parameters, λ_1 and λ_2, are adjustment parameters that govern the evolution of the conditional quasicorrelations. They must be positive and sum to less than one. A test for the sum of these parameters equalling zero tests the DCC model against the special case of the CCC model.

The DCC model may be written as

$$
\begin{aligned}
y_t &= Cx_t + \epsilon_t \\
\epsilon_t &= H_t^{1/2}\nu_t \\
H_t &= D_t^{1/2}R_tD_t^{1/2} \\
R_t &= diag(Q_t)^{-1/2}Q_t\,diag(Q_t)^{-1/2} \\
Q_t &= (1 - \lambda_1 - \lambda_2)R + \lambda_1\tilde{\epsilon}_{t-1}\tilde{\epsilon}'_{t-1} + \lambda_2 Q_{t-1}
\end{aligned}
$$

where D_t is a diagonal matrix of conditional variances,
R_t is a matrix of conditional quasicorrelations,
and $\tilde{\epsilon}_t$ is a vector of standardized residuals, $D_t^{-1/2}\epsilon_t$.
R is a weighted average of the unconditional VCE of the standardized residuals and the unconditional mean of Q_t.

With the DCC specification:

```
 . mgarch dcc (toyota honda = L.toyota L.honda), arch(1) garch(1) nolog vsquish
Dynamic conditional correlation MGARCH model

Sample: 1 - 2015                                   Number of obs   =       2014
Distribution: Gaussian                             Wald chi2(4)    =       4.81
Log likelihood =  11624.54                         Prob > chi2     =     0.3074
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
toyota						
toyota						
L1.	−.0346653	.0319267	−1.09	0.278	−.0972404	.0279098
honda						
L1.	−.0069742	.0284872	−0.24	0.807	−.0628081	.0488597
_cons	.000373	.0003108	1.20	0.230	−.0002362	.0009821
ARCH_toyota						
arch						
L1.	.0629146	.0093309	6.74	0.000	.0446263	.0812029
garch						
L1.	.9208039	.0116908	78.76	0.000	.8978904	.9437175
_cons	4.32e−06	1.16e−06	3.72	0.000	2.04e−06	6.60e−06

...

honda						
toyota						
L1.	.0030378	.0339118	0.09	0.929	-.0634281	.0695036
honda						
L1.	-.0367691	.0316091	-1.16	0.245	-.0987219	.0251836
_cons	.0005624	.000341	1.65	0.099	-.0001059	.0012307
ARCH_honda						
arch						
L1.	.0536899	.008511	6.31	0.000	.0370087	.0703711
garch						
L1.	.928433	.0115932	80.08	0.000	.9057107	.9511554
_cons	5.43e-06	1.44e-06	3.77	0.000	2.61e-06	8.26e-06
Correlation						
toyota						
honda	.7264858	.0132659	54.76	0.000	.7004852	.7524864
Adjustment						
lambda1	.0528653	.014217	3.72	0.000	.0250005	.0807301
lambda2	.746622	.0746374	10.00	0.000	.6003354	.8929085

In both the CCC and DCC specifications, the mean equations indicate that lagged daily returns of both stocks are not significant determinants of current returns, as is implied by efficient markets theory.

There are very significant GARCH effects in both specifications. A sizable correlation parameter appears, as it did in the CCC specification. The magnitudes of the `lambda` parameters indicate that the evolution of the conditional covariances depends more on their past values than on lagged residuals' innovations.

The VCC model of Tse and Tsui can be written as

$$
\begin{aligned}
y_t &= Cx_t + \epsilon_t \\
\epsilon_t &= H_t^{1/2} \nu_t \\
H_t &= D_t^{1/2} R_t D_t^{1/2} \\
R_t &= (1 - \lambda_1 - \lambda_2)R + \lambda_1 \Psi_{t-1} + \lambda_2 R_{t-1}
\end{aligned}
$$

where D_t is a diagonal matrix of conditional variances,
R_t is a matrix of conditional correlations,
R is the matrix of means to which the dynamic process reverts, and
Ψ_t is the rolling estimator of the covariance matrix of the standardized residuals $\tilde{\epsilon}_t$.

We illustrate the VCC model with two companies' shares, assumed to have no mean equation per previous findings, but with their ARCH and GARCH parameters constrained to be equal.

```
. constraint 1 _b[ARCH_toyota:L.arch] = _b[ARCH_nissan:L.arch]

. constraint 2 _b[ARCH_toyota:L.garch] = _b[ARCH_nissan:L.garch]

. mgarch vcc (toyota nissan =, noconstant), arch(1) garch(1) constraints(1 2) n
> olog vsquish

Varying conditional correlation MGARCH model

Sample: 1 – 2015                                  Number of obs   =      2015
Distribution: Gaussian                            Wald chi2(.)    =        .
Log likelihood =  11282.46                        Prob > chi2     =        .
 ( 1)  [ARCH_toyota]L.arch – [ARCH_nissan]L.arch = 0
 ( 2)  [ARCH_toyota]L.garch – [ARCH_nissan]L.garch = 0
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
ARCH_toyota						
arch						
L1.	.0797459	.0101634	7.85	0.000	.059826	.0996659
garch						
L1.	.9063808	.0118211	76.67	0.000	.883212	.9295497
_cons	4.24e-06	1.10e-06	3.85	0.000	2.08e-06	6.40e-06
ARCH_nissan						
arch						
L1.	.0797459	.0101634	7.85	0.000	.059826	.0996659
garch						
L1.	.9063808	.0118211	76.67	0.000	.883212	.9295497
_cons	5.91e-06	1.47e-06	4.03	0.000	3.03e-06	8.79e-06

...

```
. . .
Correlation
  toyota
     nissan         .6720056     .0162585     41.33    0.000      .6401394     .7038718
──────────────────────────────────────────────────────────────────────────────────────
Adjustment
    lambda1         .0343012     .0128097      2.68    0.007      .0091945     .0594078
    lambda2         .7945548      .101067      7.86    0.000       .596467     .9926425
──────────────────────────────────────────────────────────────────────────────────────
```

The validity of the constraints could be established with a likelihood ratio test against the unconstrained model.

We can produce predictions of the three series in the conditional VCE, ex post and ex ante. Notice that the ex ante predictions (beyond the sample period, ending in day 2015) quickly converge in the absence of additional information, as these are dynamic forecasts.

```
. tsappend, add(50)
. predict H*, variance dynamic(2016)
. lab var H_toyota_toyota CV_Toy
. lab var H_nissan_nissan CV_Nis
. lab var H_nissan_toyota CCov_Toy_Nis
. lab var t "Trading Day"
. tsline H* in 1800/l, leg(rows(1)) xline(2015) ylab(,angle(0))
```