

Chapter 1

LEARNING AND CONDITIONAL HETEROSCEDASTICITY IN ASSET RETURNS

*Bruce Mizrach**

Department of Economics, Rutgers University,
303b New Jersey Hall, New Brunswick, NJ 08901

1.1 Introduction

Empirical researchers have uncovered three robust stylized facts about financial time series: (1) unconditional leptokurtosis; (2) serially correlated heteroscedasticity; (3) convergence to normality with temporal aggregation; These observations appear to be robust to time period¹, choice of asset², and country.³

The autoregressive conditionally heteroscedastic (*ARCH*) regression model, developed by Engel (1982), and generalized (*GARCH*) by Bollerslev (1986), is consistent with all three stylized facts and has proved a popular model for financial modeling. In the *GARCH* models, the volatile episodes that characterize the tails of the return distribution are clustered. Large changes in the absolute value of returns tend to follow other large changes.

There is an acronymous literature of extensions to the basic model. In the *ARCH-in-mean* or *ARCH-M* model developed by Engle, Lillien and Robins (1987), the conditional variance enters the conditional mean. Nelson's (1991) *exponential GARCH* or *EGARCH* models the log of the conditional variance. Because volatility appears to be quite persistent, Engle and Bollerslev (1986) have considered an *integrated GARCH* or *IGARCH* model. High frequency data appears to be fractionally integrated process which Baillie, Bollerslev

*E-mail address: mizrach@econ.rutgers.edu

¹Pagan and Schwert (1990) report volatility persistence in stock prices in data sets from the mid-19th century.

²These properties seems to hold for bonds, see e.g. Engle, Lillien and Robins (1987), exchange rates, Hsieh (1988), commodities, Mandelbrot (1963), derivative securities, Engle and Mustafa (1992), as well as stocks. The references are merely representative. A comprehensive survey may be found in Bollerslev, Chou and Kroner (1992).

³Lin, Engle and Ito (1994), for example, show that volatility persists throughout the day, moving across international borders.

and Mikkelsen call *FIGARCH*. Bera, Higgins and Lee (1992) have introduced an *augmented ARCH* model with time varying coefficients that they call *AARCH*. This model's statistical properties are quite similar to the model I develop here.

The applied literature is simply enormous and still growing. The survey article by Bollerslev, Chou and Kroner (1992) lists several hundred references to almost every conceivable asset.

Despite the widespread use of the *GARCH* model, the specification of the conditional heteroscedasticity is essentially ad hoc. As Diebold (1988) notes, the literature is lacking a truly economic theory explaining the persistence in volatility.

This paper's contribution is to develop a model of asset pricing and learning where *GARCH* disturbances evolve naturally out of the decision problem of economic agents. Agents try to learn about the parameters of a stochastic process. Their evolving beliefs are incorporated into asset returns, leading to time variation in the parameters. I show that one representation for the data generating mechanism is a Markov model with *GARCH* disturbances.

The learning model suggests an extended *GARCH* specification in which variables from the conditional mean explain time variation in the coefficients of the conditional variance. I propose a time-varying *GARCH* specification which nests the standard model. In an empirical example with the Italian Lira-German Deutschmark exchange rate, I reject the standard framework in favor of the learning model.

The paper is organized in the following manner. Section 1 develops Bollerslev's generalization of the *ARCH* model. A review of the time series properties of the model highlights the three stylized facts. Section 2 describes the model of asset pricing and learning. The covariance structure of this model is described in Section 3. Finite sample properties are explored in Section 4 in four different designs including one with structural change. The exchange rate example is presented in Section 5. Section 6 concludes and suggests some directions for future research.

1.2 GARCH in the Linear Regression Model

I work throughout the paper with the linear regression model

$$y_t = x_t \beta + \varepsilon_t. \quad (1.1)$$

y_t is an observation on an endogenous variable which may be thought of as the fundamental, x_t is a $1 \times k$ vector of explanatory variables, β is a $k \times 1$ vector of unknown parameters, and ε_t is a disturbance term distributed $N(0, \sigma_\varepsilon^2)$. Define $Y_t = [y_1 \cdots y_t]'$, $X_t = [x_1 \cdots x_t]'$, and $\xi_t = [\varepsilon_1 \cdots \varepsilon_t]'$.

Engle extended the model 1.1 to allow for an explicitly time varying conditional variance h_t

$$\varepsilon_t = \sqrt{h_t} \eta_t, \quad \eta_t \sim N(0, 1). \quad (1.2)$$

Engle called it the *ARCH* model because the conditional variance was expressed as a function of the lagged innovations to the conditional mean

$$h_t = a_0 + \sum_{i=1}^q a_i \varepsilon_{t-i}^2. \quad (1.3)$$

Following Bollerslev (1986), I will call 1.3 the $ARCH(q)$ model.

Bollerslev generalized Engle's model to allow the conditional variance to depend on lagged conditional variances as well. The $GARCH(p, q)$ model is written as

$$h_t = a_0 + \sum_{i=1}^q a_i \varepsilon_{t-i}^2 + \sum_{j=1}^p b_j h_{t-j}. \quad (1.4)$$

An equivalent, perhaps more intuitive formulation for 1.3, as in Bollerslev (1988), arises by substituting $v_t \equiv (\varepsilon_t^2 - h_t) = h_t(\eta_t^2 - 1)$ at each ε_t^2 ,

$$\varepsilon_t^2 = a_0 + \sum_{i=1}^{\max[p, q]} (a_i + b_i) \varepsilon_{t-i}^2 - \sum_{j=1}^p b_j v_{t-j} + v_t, \quad (1.5)$$

where $a_i = 0$ for $i > q$ and $b_i = 0$ for $i > p$. In this form, we can see that the $GARCH(p, q)$ model is an $ARMA(m, p)$ in the squared disturbances, where $m = \max[p, q]$.

1.2.1 GARCH and the stylized facts for asset returns

1.2.1.1 Leptokurtosis

Calculation of the kurtosis requires the second and fourth unconditional moments. The unconditional variance of the $GARCH(p, q)$ model is given in Bollerslev (1986) as

$$\sigma_\varepsilon^2 = a_0 / (1 - \sum_{i=1}^q a_i - \sum_{j=1}^p b_j). \quad (1.6)$$

Let $A(L)$ be the q -dimensional polynomial in the lag operator for the squared innovations and $B(L)$ be the p -dimensional polynomial for the lagged conditional variances. A sufficient condition for the existence of the second moment is $A(1) + B(1) < 1$.

The fourth moment requires even stronger parameter restrictions. For the $ARCH(q)$ case, Milhoj (1985) derives necessary and sufficient conditions. Pre-multiplying 1.5 by ε_{t-j}^2 and taking expectations yields the j th autocovariance,

$$\gamma_j \equiv cov(\varepsilon_t^2, \varepsilon_{t-j}^2) = \sum_{i=1}^q a_i \gamma_{j-i}. \quad (1.7)$$

Dividing through by γ_0 , he obtains the analog of the Yule-Walker equations,

$$\rho_j \equiv \gamma_j / \gamma_0 = \sum_{i=1}^q a_i \rho_{j-i}, \quad (1.8)$$

where the a 's can now be interpreted as the q partial autocorrelations of the $GARCH$ process,

$$a_j = corr(\varepsilon_t^2, \varepsilon_{t-j}^2 \mid \varepsilon_{t-1}^2, \dots, \varepsilon_{t-j-1}^2). \quad (1.9)$$

The standard $ARMA$ diagnostics apply here. The partial autocorrelations are zero for $j > q$.

Milhoj stacks the equations 1.9 in matrix form as

$$A = (I - \Psi)\rho \quad (1.10)$$

where $A = [a_1 \ \dots \ a_q]'$, I is the $q \times q$ identity matrix, $\rho = [\rho_1 \ \dots \ \rho_q]'$ and Ψ is the $q \times q$ matrix, $\Psi_{ij} = a_{i+j} + a_{i-j}$, with $a_j = 0$ for $j \leq 0$ and $j > q$. He shows that a necessary and sufficient condition for the existence of the fourth moment⁴ is

$$3A'(I - \Psi)^{-1}A < 1 \quad (1.11)$$

⁴Note that many stationary $GARCH$ models will not have a fourth moment.

For the $ARCH(1)$, this is simply

$$3a_1^2 < 1.$$

For the $ARCH(2)$, 1.11 can be written as

$$3 \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} 1-a_2 & 0 \\ a_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

which, upon inverting yields,

$$a_2 + 3a_1^2 + 3a_2^2 + 3a_1^2a_2 - 3a_2^3$$

The kurtosis⁵ is

$$\kappa = 2 \frac{3A'(I - \Psi)^{-1}A}{1 - 3\Phi'(I - \Psi)^{-1}\Phi} + 1.$$

Given condition 1.11, this implies the kurtosis is positive.

For the $GARCH(1,1)$ process, Bollerslev (1986) has derived a necessary and sufficient condition for the existence of the $2r$ th moment,

$$\mu(a_1, b_1, r) = \sum_{j=0}^r \binom{r}{j} c_j a_1^j b_1^{r-j} < 1, \quad (1.12)$$

where $c_0 = 1$, $c_j = \prod_{i=1}^j (2i - 1)$. For the fourth moment, he obtains from 1.12,

$$\mu_4 = \frac{3a_0^2(1 + a_1 + b_1)}{(1 - a_1 - b_1)(1 - b_1^2 - 2a_1b_1 - 3a_1^2)}$$

Given the existence of the second moment, the first term in parentheses is less than 1, making a necessary and sufficient condition⁶ for the existence of the fourth moment that

$$3a_1^2 + 2a_1b_1 + b_1^2 < 1. \quad (1.13)$$

The kurtosis is then

$$\kappa = \frac{6a_1^2}{(1 - b_1^2 - 2a_1b_1 - 3a_1^2)} \quad (1.14)$$

which is positive if 1.13 is satisfied.

The leptokurtosis in asset returns is consistent with a number of other statistical models. Our next stylized fact will help rule out one important class.

⁵The coefficient of kurtosis κ is defined as $(E[\varepsilon_t^4] - 3(\sigma_\varepsilon^2)^2)/(\sigma_\varepsilon^2)^2$. If this is positive, it indicates that the distribution is "fat-tailed" relative to the normal.

⁶There are no general formulas for the existence of the fourth moment in other $GARCH$ processes. Bollerslev has derived conditions for the $GARCH(1,2)$: $a_2 + 3a_1^2 + 3a_2^2 + b_1^2 + 2a_1b_1 - 3a_2^3 + 3a_1^2a_2 + 6a_1a_2b_1 + a_2b_1^2 < 1$; and for the $GARCH(2,1)$: $b_2 + 3a_1^2 + b_1^2 + b_2^2 + 2a_1b_1 - b_2^3 - a_1^2b_2 + 2a_1b_1b_2 + b_1^2b_2 < 1$. These sums all will appear in the denominator of 1.14, making them leptokurtic as well.

1.2.1.2 Normality with temporal aggregation

Consider a time series $\{y_t\}_{t=1}^T$ generated by 1.1 and 1.2. Denote by Y_t^s the s -period temporal aggregate

$$Y_t^s = \sum_{i=1}^s y_{t+i-1}.$$

For instance, if y_t was the change in the daily exchange rate, Y_t^5 would be the weekly change.

A statistical representation for asset returns must account for our third stylized fact, that as s increases, asset returns converge to normality. One of the strong objections to the Paretian densities⁷ is that they are stable under aggregation. If the tails are thick in the daily changes, they will also be thick in the weekly aggregated series.

The convergence to normality in the *GARCH* process is a fairly straightforward implication of central limit theory. The temporal aggregate sums over a very large number of draws of y_t as s grows large. The only difficulty in proving convergence is to account for the dependence in the data. A central limit theory for dependent observations has been worked out under fairly weak assumptions: White (1984) shows that stationarity and ergodicity are sufficient.

Diebold (1988) has proven the convergence using White's theorem. He shows that if y_t is an $AR(p)$, $GARCH(p, q)$ process, the aggregated series $\{Y_t^s\}_{t=1}^n$, $n = \text{int}[T/s]$, has an unconditional normal distribution as $n \rightarrow \infty$.⁸

1.3 A Model of Asset Pricing and Learning

A representative agent determines the return, r_t , to a security based on his beliefs about the fundamental,

$$r_t = E[y_t]. \quad (1.15)$$

The agent is assumed to know the structure of the model but not its parameters. He updates his beliefs recursively using a least squares algorithm,

$$E[y_t] = x_t \hat{\beta}_t,$$

where

$$\hat{\beta}_t = (\tilde{X}_t' \tilde{X}_t)^{-1} \tilde{X}_t' \tilde{Y}_t = \beta + (\tilde{X}_t' \tilde{X}_t)^{-1} \tilde{X}_t' \tilde{\xi}_t \equiv \beta + \tilde{M}_t \tilde{\xi}_t. \quad (1.16)$$

I assume that the agent observes k pre-sample values of the fundamental and explanatory variables so that his beliefs start at time $t = 1$. I use the tilde to denote matrices that are augmented with these pre-sample values, e.g. $\tilde{X}_t = [x_{-k+1} \cdots x_0 \cdots x_t]$,

Consider now the perspective of an econometrician trying to analyze the relationship between security returns and the fundamental. He does not directly observe the agent's

⁷For discussion on this family in the economics literature, see Mandelbrot (1963) or Fama and Roll (1968).

⁸There is considerable discussion in the literature over the non-normality of the conditional distribution. Many studies find that v_t is not normally distributed. Bollerslev (1987), for example, proposes a model where v is Student- t .

beliefs, but he does observe a time series of security returns,

$$R_T = \begin{bmatrix} r_1 \\ \vdots \\ r_T \end{bmatrix} = \begin{bmatrix} x_1 \beta \\ \vdots \\ x_T \beta \end{bmatrix} + \begin{bmatrix} x_1 \tilde{M}_1 \tilde{\xi}_1 \\ \vdots \\ x_T \tilde{M}_T \tilde{\xi}_T \end{bmatrix} \equiv X_T \beta + V_T. \quad (1.17)$$

Now suppose the econometrician regressed the return series on the same vector of explanatory variables used by the agent,

$$\hat{r}_t = x_t \bar{\beta}_T, \quad (1.18)$$

where

$$\bar{\beta}_T = (X_T' X_T)^{-1} X_T' R_T = \beta + M_T V_T \quad (1.19)$$

Both the agent and the econometrician have unbiased expectations

$$E[\hat{\beta}_t] = E[\bar{\beta}_T] = \beta,$$

but the covariance structure is quite different from the standard Markov model. In particular, I will show that the residuals are conditionally heteroscedastic.

1.4 The Covariance Structure of the Residuals

I first describe the covariance matrix of this process. In the second part, I focus on the conditional variance.

1.4.1 The covariance matrix

The residuals of 1.15 less 1.18 are given by

$$r_t - \hat{r}_t \equiv z_t = x_t (\tilde{M}_t \tilde{\xi}_t - M_T V_T) \equiv z_{1t} - z_{2t}. \quad (1.20)$$

There will be four terms that make up each element of the covariance matrix $\Omega \equiv E[Z_T Z_T']$. For each term below, let $t = 1, \dots, T$, and $j = 0, \dots, t-1$. The first component is

$$\begin{aligned} E[z_{1t} z_{1t-j}'] &= x_t \tilde{M}_t E[\tilde{\xi}_t \tilde{\xi}_{t-j}'] \tilde{M}_{t-j}' x_{t-j}', \\ &= x_t \tilde{M}_t \sigma_\varepsilon^2 I_{t,t-j} \tilde{M}_{t-j}' x_{t-j}', \\ &= \sigma_\varepsilon^2 x_t (\tilde{X}_t' \tilde{X}_t)^{-1} x_{t-j}'. \end{aligned} \quad (1.21)$$

$I_{t,t-j}$ is a $t \times t-j$ identity matrix.

The second term is given by

$$\begin{aligned}
E[z_{1t}z'_{2t-j}] &= \begin{bmatrix} x_t \tilde{M}_t E[\tilde{\xi}_t \tilde{\xi}'_1] \tilde{M}_1 x'_1 \\ \vdots \\ x_t \tilde{M}_t E[\tilde{\xi}_t \tilde{\xi}'_T] \tilde{M}_T x'_T \end{bmatrix} M'_T x'_{t-j}, & (1.22) \\
&= \sigma_\varepsilon^2 x_t \begin{bmatrix} (\tilde{X}'_t \tilde{X}_t)^{-1} x'_1 \\ \vdots \\ (\tilde{X}'_t \tilde{X}_t)^{-1} x'_t \\ \vdots \\ (\tilde{X}'_T \tilde{X}_T)^{-1} x'_T \end{bmatrix} M'_T x'_{t-j} \equiv \sigma_\varepsilon^2 x_t W_t M'_T x'_{t-j}.
\end{aligned}$$

The second line uses the same substitution as in 1.21. $\tilde{M}_t I_{t,k} \tilde{M}_k$ reduces to $(\tilde{X}'_t \tilde{X}_t)^{-1}$ until $k > t$ when it then starts to cancel to the left.

The third term is

$$E[z_{2t}z'_{1t-j}] = \sigma_\varepsilon^2 x_t M_T \begin{bmatrix} x_1 (\tilde{X}'_{t-j} \tilde{X}_{t-j})^{-1} \\ \vdots \\ x_{t-j} (\tilde{X}'_{t-j} \tilde{X}_{t-j})^{-1} \\ \vdots \\ x_T (\tilde{X}'_T \tilde{X}_T)^{-1} \end{bmatrix}' = \sigma_\varepsilon^2 x_t M_T W'_{t-j}. \quad (1.23)$$

In this case, the elements $E[V_T V_T' \tilde{\xi}'_{t-j} \tilde{M}_{t-j}]$ leave an $(\tilde{X}'_{t-j} \tilde{X}_{t-j})^{-1}$ until the left index exceeds $t - j$.

The final term is a bit more involved because the cross product of $V_T V_T'$ produces a $(T \times T)$ matrix,

$$E[z_{2t}z'_{2t-j}] = \sigma_\varepsilon^2 x_t M_T \begin{bmatrix} x_1 (\tilde{X}'_1 \tilde{X}_1)^{-1} x'_1 & \cdots & x_1 (\tilde{X}'_1 \tilde{X}_1)^{-1} x'_T \\ \vdots & \ddots & \vdots \\ x_T (\tilde{X}'_T \tilde{X}_T)^{-1} x'_1 & \cdots & x_T (\tilde{X}'_T \tilde{X}_T)^{-1} x'_T \end{bmatrix} M'_T x'_{t-j}. \quad (1.24)$$

Overall, this covariance matrix is quite complex. The off-diagonal elements $z_t z_{t-j}$ are all non-zero, indicating that the residuals are serially correlated.⁹

Now consider the diagonal terms, $E[z_t^2]$, for $t = 1, \dots, T$. 1.21-1.24 simplify for $j = 0$. 1.22 and 1.23 are identical, and because the residuals are orthogonal to the x' s, $E[x_t z_t] = 0$, the fourth term cancels one of them. I am left with,

$$\gamma_z(t) \equiv E[z_t^2] = E[z_{1t}^2 - z_{1t}z'_{2t}] = \sigma_\varepsilon^2 [x_t (\tilde{X}'_t \tilde{X}_t)^{-1} x'_t - x_t W_t M'_T x'_t]. \quad (1.25)$$

From 1.25, it is clear that the variances will be time varying and related to the explanatory variables. In the next section, I will show that the squared residuals are an *ARMA* process.

⁹The presence of serial correlation poses some difficulties for *GARCH* hypothesis testing. See e.g. Bera, Higgins and Lee (1992) for an alternative approach to the standard Lagrange multiplier test of Engle (1982).

1.4.2 A GARCH representation for the conditional variances

I will analyze in this section the autocorrelation and partial autocorrelation functions. To show that the model has a GARCH representation, I will show that they are both non-zero at all lags.

Let's begin with the autocovariance function for z_t^2 ,

$$\gamma_{zz}(t, t-j) \equiv E[(z_t^2 - E[z_t^2])(z_{t-j}^2 - E[z_{t-j}^2])].$$

Since ε_t is Gaussian, the terms are just cross products of γ_z ,

$$\begin{aligned} \gamma_{zz}(t, t-j) &= E[z_{1t}^2 z_{1t-j}^2 - 2z_{1t} z_{1t-j} + z_{1t}^2 z_{2t-j}^2 - 2z_{1t} z_{2t} z_{1t-j}^2 \\ &\quad + 4z_{1t} z_{2t} z_{1t-j} z_{2t-j} - 2z_{1t} z_{2t} z_{t-j}^2 + z_{2t}^2 z_{1t-j}^2 - 2z_{2t}^2 z_{1t-j} z_{2t-j} + z_{2t}^2 z_{2t-j}^2]. \end{aligned} \quad (1.26)$$

Note that for large T at any j , $E[z_{2t-j}] \approx 0$. To understand this intuitively, consider the case where x is a scalar,

$$z_{2t-j} = x_{t-j} \left(\frac{\sum_{r=1}^T x_r^2 (\sum_{s=1}^r x_s \varepsilon_s / \sum_{s=1}^r x_s^2)}{\sum_{r=1}^T x_r^2} \right). \quad (1.27)$$

The inner double sum converges quickly. Consequently,

$$\gamma_{zz}(t, t-j) \approx E[z_{1t}^2 z_{1t-j}^2] = E[x_t \tilde{M}_t \tilde{\xi}_t \tilde{\xi}_t' \tilde{M}_t' x_t' x_{t-j} \tilde{M}_{t-j} \tilde{\xi}_{t-j} \tilde{\xi}_{t-j}' \tilde{M}_{t-j}' x_{t-j}'].$$

This will be non-zero if the stochastic part of this term is non-zero,

$$E[\tilde{\xi}_t \tilde{\xi}_t' I_{t,t-j} \tilde{\xi}_{t-j} \tilde{\xi}_{t-j}'] \equiv \Sigma_{t-j}, \quad (1.28)$$

where Σ_{t-j} is a symmetric $(t-j) \times (t-j)$ matrix whose r th diagonal element is

$$E[\varepsilon_r^2 \sum_{s=1}^{t-j} \varepsilon_s^2] = \mu_4 + \sum_{s \neq r}^{t-j} (\sigma_\varepsilon^2)^2. \quad (1.29)$$

Because the inner sum runs from 1 to $t-j$, this expectation is non-zero for *any* j , and consequently, so is $\gamma_{zz}(t, t-j)$.

To assess the significance of the partial autocorrelation coefficients, consider the regression of z_t^2 on z_{t-1}^2 and call this coefficient ϕ_1 . The residuals from this regression,

$$z_t^2 - c - \phi_1 z_{t-1}^2,$$

will include the difference of the partial sums up to time $t-1$,

$$\Delta(\sum_{j=-k+1}^t x_j \varepsilon_j)^2 = x_t \varepsilon_t \sum_{j=-k+1}^{t-1} x_j \varepsilon_j.$$

These will be correlated with any z_{t-j}^2 because the remaining portion still sums disturbances from the first time period.

Intuitively, the learning process induces parameter variation in the conditional mean. An econometrician who fails to model this structure leaves information in the residuals. An equivalent representation, this section demonstrates, is a Markov model with *GARCH* disturbances.

Technically, we still need to show that the autocorrelations and partial autocorrelations decay sufficiently quickly for the process to be stationary. This unfortunately must be left for future research. In this paper, Monte Carlo simulations will have to suffice.

Table 1.1: Average autocorrelation coefficient (*AC*) and partial autocorrelation coefficient (*PAC*) at each lag in 250 replications. Simulation results for equation $y_t = 0.05 + \varepsilon_t$, $\varepsilon_t \sim N(0, 1)$.

Lag	$T = 25$		$T = 100$		$T = 250$	
	AC	PAC	AC	PAC	AC	PAC
1	0.397	0.397	0.643	0.643	0.724	0.724
2	0.156	-0.083	0.458	-0.009	0.569	0.078
3	0.063	0.039	0.356	0.071	0.477	0.078
4	0.012	-0.059	0.282	-0.036	0.410	-0.021
5	-0.032	-0.012	0.226	0.032	0.356	0.048
6	-0.045	-0.034	0.187	-0.004	0.320	0.009
7	-0.054	-0.015	0.157	0.029	0.289	0.042
8	-0.066	-0.060	0.132	-0.006	0.264	-0.001
9	-0.074	-0.024	0.112	0.022	0.243	0.032
10	-0.086	-0.064	0.092	-0.011	0.224	-0.004

1.5 Finite Sample Properties

I consider four finite sample exercises. The first design is a simple scalar case where the only regressor is a constant term. In this exercise, I hope to show the persistence coming solely from the disturbances. In the second case, I consider a multivariate regression model where the x_t 's are independent. Since the heteroscedasticity varies with the explanatory variables, I want to exclude the effects of any dependence coming from the regressors. In the third simulation, I consider an $AR(p)$ model. Finally, in a more realistic model, I consider structural change in the parameters.

In each simulation, I draw samples of $T + 5$ observations on the disturbance term and regressors to generate the y_t 's, essentially assuming the first 5 are only visible to the agent. I consider three sample sizes of length $T = 25, 100$ and 250 . I then do 250 replications of each. I report sample averages of the first ten autocorrelations and partial autocorrelations of z_t^2 in Tables 1 through 4.

1.5.1 Scalar regression model

In this first example, $y_t = 0.05 + \varepsilon_t$, x_t is a scalar constant, so $k = 1$, and $x_t = 1$ for all t . ε is $N(0, 1)$ and independent and identically distributed. The results are in Table 1.

There is clear evidence of an AR process even in the small sample. The partial autocorrelations are large only at the first lag. An econometrician might readily conclude that this is an $AR(1)$ in ε_t^2 or an $ARCH(1)$ process. It appears that some structure in the independent variables is needed to generate a high order $ARCH$ or $GARCH$. This motivates our next simulation.

Table 1.2: Average autocorrelation coefficient (*AC*) and partial autocorrelation coefficient (*PAC*) at each lag in 250 replications.. Simulation results for equation $y_t = 0.25 + 0.5x_{1t} - 0.25x_{2t} + \varepsilon_t$, x_t and $\varepsilon_t \sim N(0, 1)$.

Lag	$T = 25$		$T = 100$		$T = 250$	
	AC	PAC	AC	PAC	AC	PAC
1	0.107	0.107	0.246	0.246	0.292	0.292
2	0.084	0.040	0.206	0.117	0.255	0.159
3	0.051	0.045	0.164	0.113	0.235	0.148
4	0.018	-0.034	0.163	0.062	0.213	0.070
5	0.008	0.003	0.129	0.061	0.189	0.067
6	-0.013	-0.029	0.125	0.029	0.167	0.033
7	-0.022	-0.022	0.104	0.034	0.157	0.049
8	-0.021	-0.036	0.089	0.000	0.140	0.012
9	-0.028	-0.011	0.083	0.026	0.141	0.041
10	-0.046	-0.044	0.076	0.008	0.134	0.013

1.5.2 Multiple independent regressors

I use a three variable model, $y_t = 0.25 + 0.5x_1 - 0.25x_2 + \varepsilon_t$. Each x and the ε are independent draws from an $N(0, 1)$. I again drop 5 pre-sample values for each of the three sample sizes. Results are reported in Table 2.

By the mid-sized sample, $T = 100$, there is clear evidence of a higher order *ARMA* process. In the large sample, there is the nice smooth decay in both the *AC* and *PAC* that the theory predicts.

1.5.3 An $AR(p)$ model

I next consider the *AR*(2) model, $y_t = x_t = 0.05 + 0.5x_{t-1} - 0.25x_{t-2} + \varepsilon_t$. The ε 's are again $N(0, 1)$, and I set the initial values for the x 's to zero, still dropping the first 5 observations. Results for this third exercise are in Table 3. The covariance structure suggests that dependence in the x 's should induce additional dependence in the squared disturbances. This can be detected in this simulation. At lags 1 and 2, the *AC* and *PAC* are larger than in Table 2 for all three sample sizes. These coefficients again suggest a high order *ARMA* process.

1.5.4 Structural change

If agent's beliefs were finite only on a single model, volatility would tend to die out. In the third finite sample exercise with $T = 100$, the unconditional variance in the first half of the sample is 35% larger on average than in the second half. The conditional heteroscedasticity remains significant in the sample as a whole though because small errors in the second half of the series are following other small errors.

In financial market data, we see repeated bursts of volatility. In this framework, I model these volatility outbreaks as changes in the coefficients of the fundamental. In this last

Table 1.3: Average autocorrelation coefficient (*AC*) and partial autocorrelation coefficient (*PAC*) at each lag in 250 replications. Simulation results for equation $y_t = x_t = 0.05 + 0.5x_{t-1} - 0.25x_{t-2} + \varepsilon_t$, $\varepsilon_t \sim N(0, 1)$.

Lag	$T = 25$		$T = 100$		$T = 250$	
	AC	PAC	AC	PAC	AC	PAC
1	0.156	0.156	0.286	0.286	0.333	0.333
2	0.139	0.066	0.252	0.139	0.295	0.168
3	0.065	0.059	0.176	0.103	0.219	0.107
4	0.042	-0.031	0.161	0.030	0.205	0.046
5	0.010	-0.001	0.147	0.067	0.192	0.079
6	-0.009	-0.033	0.117	0.014	0.163	0.027
7	-0.015	-0.011	0.089	0.011	0.137	0.027
8	-0.025	-0.045	0.078	-0.009	0.127	0.005
9	-0.039	-0.020	0.067	0.023	0.117	0.029
10	-0.043	-0.041	0.063	0.004	0.111	0.013

Table 1.4: Average autocorrelation coefficient (*AC*) and partial autocorrelation coefficient (*PAC*) at each lag in 250 replications. Simulation results for *AR*(2) with 1% chance of shift in β_2 .

Lag	$T = 25$		$T = 100$		$T = 250$	
	AC	PAC	AC	PAC	AC	PAC
1	0.157	0.157	0.287	0.287	0.333	0.333
2	0.138	0.063	0.251	0.140	0.297	0.173
3	0.063	0.058	0.180	0.109	0.226	0.114
4	0.043	-0.029	0.162	0.030	0.200	0.035
5	0.010	-0.003	0.148	0.068	0.193	0.079
6	-0.009	-0.011	0.119	0.012	0.174	0.039
7	-0.015	-0.047	0.092	0.015	0.148	0.036
8	-0.025	-0.036	0.078	-0.011	0.132	0.006
9	-0.038	-0.019	0.069	0.023	0.123	0.033
10	-0.043	-0.040	0.061	-0.001	0.111	0.001

exercise, there is a 1% chance, each period, that the first AR parameter will shift up by 0.1. I constrain the parameter when necessary to keep the process stationary.

The results are in Table 4. The AC and PAC are basically the same as in Table 3 even though the unconditional volatility is now equal in both halves of the sample.

1.6 An Empirical Example

These simulation results do not constitute any direct evidence in favor of the learning model. In this section, I propose a flexible parameterization for the conditional heteroscedasticity. This functional form captures the parameter variation in the conditional variance and nests the standard model. To avoid any problems with negative variances, I use logs.

Let $h_t = E[\varepsilon_t^2 | \varepsilon_t^2, h_{t-1}, x_t, \dots]$, x_t be a scalar, and write the functional form as

$$\log(h_t) = a_0 + \left(\sum_{i=1}^q [a_{1,i} + \sum_{r=0}^s a_{2+r,i} x_t x_{t-r}] \right) \varepsilon_{t-i}^2 + \left(\sum_{j=1}^p [b_{1,j} + \sum_{r=0}^s b_{2+r,j} x_t x_{t-r}] \right) \log(h_{t-j}). \quad (1.30)$$

The model lets the coefficients on the squared disturbances and the conditional variances vary with cross products of the x 's. Also, if $\sum a_{2+r,i} = \sum b_{2+r,j} = 0$, it reduces to the standard model.

Table 1.5: Estimates of Log of the Conditional Variance for IL/DM Exchange Rate. Coefficient estimates for a $GARCH(1,1)$ and augmented $GARCH(1,1)$ estimation. Newey-West standard errors are reported. The likelihood ratio statistic is distributed $\chi^2(4)$.

Explanatory Variable	Model 1	Model 2
<i>Constant</i>	-0.261	-0.183
(<i>t</i> -stat)	(3.26)	(4.32)
ε_{t-1}^2	0.176	0.195
	(4.32)	(1.91)
$x_t^2 \varepsilon_{t-1}^2$		-0.004
		(0.34)
$x_t x_{t-1} \varepsilon_{t-1}^2$		0.036
		(0.77)
h_t	0.853	0.917
	(14.74)	(27.03)
$x_t^2 h_t$		-0.017
		(0.11)
$x_t x_{t-1} h_t$		-0.244
		(2.66)
Log likelihood	86.127	116.690
Likelihood ratio		61.126

I obtained a series of 510 daily Italian Lira-German Deutschemark exchange rates for 1992-93. This period includes the withdrawal of the Lira from the European exchange rate grid. The leptokurtosis is over 22, and previous work, Mizrach (1995), suggested the possibility of a $GARCH$ model.

I use log differences for returns, and the conditional mean is $AR(1)$. I then jointly estimated a log- $GARCH(1,1)$ for h_t . Results are in the first column of Table 5. Overall, the model appears quite satisfactory with significant¹⁰ AR and MA parameters. Nonetheless, I wanted to see if the learning model parameterization might locate additional structure.

The second column of Table 5 reports estimation for $s = 1$. The model appears to offer a good deal better fit. The lagged terms $x_t x_{t-1}$ are significant in both the squared disturbances and the lagged conditional variances. The likelihood improves by 25%, and the $\chi^2(4)$ likelihood ratio statistic is over 60. I can easily reject the standard model in favor of the learning specification.

1.7 Conclusion

The literature on financial market volatility has made great strides in the statistical modeling of conditional variances. A variety of parameterizations for volatility exist. This paper has not completely avoided offering yet another. Focusing on the source of volatility should help us choose among candidate models. Further research on the impact of learning on volatility should prove fruitful at least in this respect.

A single empirical example was also provided in favor of the learning specification. Additional research will be needed to establish whether this model is as empirically robust as its precursors.

References

- [1] Baillie, R., T. Bollerslev and H.O. Mikkelsen (1993), "Fractionally integrated generalized autoregressive conditional heteroscedasticity," Kellogg School Working Paper #168.
- [2] Bera, A., M.L. Higgins, and S. Lee (1992), "Interaction between autocorrelation and conditional heteroscedasticity: a random coefficient approach," *Journal of Business and Economic Statistics* 10, 133-142.
- [3] Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics* 31, 307-27.
- [4] Bollerslev, T. (1987), "A conditionally heteroskedastic time series model for speculative prices and rates of return," *Review of Economics and Statistics* 69, 542-47.
- [5] Bollerslev, T. (1988), "On the correlation structure for the generalized Autoregressive conditional heteroskedastic process," *Journal of Time Series Analysis* 9, 121-31.
- [6] Bollerslev, T., R. Chou and K. Kroner (1992), "ARCH modeling in finance: a review of theory and empirical evidence," *Journal of Econometrics* 52, 5-59.

¹⁰I use Newey-West HAC standard errors with 4 lags.

- [7] Diebold, F. (1988), *ARCH Modeling of Exchange Rates*, New York: Springer-Verlag.
- [8] Engle, R., "Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation," *Econometrica* 50, 987-1007.
- [9] Engle, R., D. Lillien and R. Robins (1987), "Estimating time varying risk premia in the term structure: the ARCH-M model," *Econometrica* 55, 391-407.
- [10] Engle, R. and C. Mustafa (1992), "Implied ARCH models from Option Prices," *Journal of Econometrics* 52, 289-311.
- [11] Fama, E. and R. Roll (1968), "Some properties of symmetric stable distributions,:" *Journal of the American Statistical Association*, 817-35.
- [12] Hsieh, D. (1988), "The statistical properties of daily foreign exchange rates: 1974-1983," *Journal of International Economics* 17, 173-84.
- [13] Lin, W-L., R. Engle, and T. Ito (1994), "Do bulls and bears move across borders? international transmission of stock returns and volatility," *Review of Financial Studies* 7, 507-38.
- [14] Mandelbrot, B. (1963), "The variation of certain speculative prices," *Journal of Business* 36, 394-419.
- [15] Milhoj, A. (1985), "The moment structure of ARCH processes," *The Scandinavian Journal of Statistics* 12, 281-92.
- [16] Mizrach, B. (1995), "Target Zone Models with Stochastic Realignments: An Econometric Evaluation," *Journal of International Money and Finance* 14, 641-57.
- [17] Nelson, D. (1991), "Conditional heteroscedasticity in asset returns: a new approach," *Econometrica* 59, 347-70.
- [18] Pagan, A. and Schwert, "Alternative models for conditional stock volatility," *Journal of Econometrics* 45, 267-90.