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

---

¹E-mail address: mizrach@econ.rutgers.edu


³These 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 Deutschemark 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

\[
y_t = x_t \beta + \epsilon_t. \quad (1.1)
\]

\(y_t\) 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, \(\beta\) is a \(k \times 1\) vector of unknown parameters, and \(\epsilon_t\) is a disturbance term distributed \(N(0, \sigma^2_{\epsilon})\). Define \(Y_t = [y_1 \cdots y_t]'\), \(X_t = [x_1 \cdots x_t]'\), and \(\xi_t = [\epsilon_1 \cdots \epsilon_t]'\).

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

\[
\epsilon_t = \sqrt{h_t} \eta_t, \quad \eta_t \sim N(0, 1). \quad (1.2)
\]

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

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

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

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

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

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

$$\sigma^2 = a_0/(1 - \sum_{i=1}^q a_i - \sum_{j=1}^p b_i). \tag{1.6}$$

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 $\varepsilon_{t-j}^2$ and taking expectations yields the $j$th autocovariance,

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

Dividing through by $\gamma_0$, he obtains the analog of the Yule-Walker equations,

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

where the $\rho$’s can now be interpreted as the $q$ partial autocorrelations of the $GARCH$ process,

$$a_j \equiv \text{corr}(\varepsilon_t^2, \varepsilon_{t-j}^2 | \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-j-1}^2). \tag{1.9}$$

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

Milhoj stacks the equations 1.9 in matrix form as

$$A = (I - \Psi)\rho \tag{1.10}$$

where $A = [a_1 \cdots a_q]^\prime$, $I$ is the $q \times q$ identity matrix, $\rho = [\rho_1 \cdots \rho_q]^\prime$ and $\Psi$ 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 moment\(^4\) is

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

\(^4\)Note 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

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

which, upon inverting yields,

\[ a_2 + 3a_1^2 + 3a_1^2 + 3a_1^2a_2 - 3a_1^3 \]

The kurtosis 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 2rth 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 \( c_0 = 1 \), \( c_j = \prod_{k} (2j - 1) \). 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 condition 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.

---

5 The coefficient of kurtosis \( \kappa \) is defined as \((E[e_t^4] - 3(\sigma_e^2)^2)/(\sigma_e^2)^2\). If this is positive, it indicates that the distribution is “fat-tailed” relative to the normal.

6 There are no general formulas for the existence of the fourth moment in other GARCH processes. Bollerslev has derived conditions for the GARCH(1, 2): \( a_2 + 3a_1^2 + 3a_1^2 + b_1^2 + 2a_1b_1 - 3a_1^2 + 3a_1^2a_2 + 6a_1\varepsilon_1b_1 + a_2b_2^2 < 1 \); and for the GARCH(2, 1): \( b_2 + 3a_1^2 + b_1^2 + b_1^2 + 2a_1b_1 - b_1^2 - a_1^2b_2 + 2a_1b_1b_2 + b_2^2b_2 < 1. \) 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 \( Y_s^t \) the \( s \)-period temporal aggregate
\[
Y_s^t = \sum_{t=1}^{s} y_t.
\]
For instance, if \( y_t \) was the change in the daily exchange rate, \( Y_5^t \) 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 densities\(^7\) 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 a \( GARCH(p,q) \) 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 \( \{ Y_s^t \}_{t=1}^n \), \( n = \text{int}[T/s] \), has an unconditional normal distribution as \( n \to \infty \).\(^8\)

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,
\[
r_t = E[y_t].
\]
(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
\[
\hat{\beta}_t = (\tilde{X}_t'\tilde{X}_t)^{-1}\tilde{X}_t'\tilde{Y}_t = \beta + (\tilde{X}_t'\tilde{X}_t)^{-1}\tilde{X}_t'\tilde{\xi}_t = \beta + \tilde{M}_t \tilde{\xi}_t.
\]
(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

\(^7\)For discussion on this family in the economics literature, see Mandelbrot (1963) or Fama and Roll (1968).
\(^8\)There is considerable discussion in the literature over the non-normality of the conditional distribution. Many studies find that \( \nu_t \) is not normally distributed. Bollerslev (1987), for example, proposes a model where \( \nu_t \) is Student- \( t \).
beliefs, but he does observe a time series of security returns,

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

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

\[ \hat{r}_t = x_t\hat{\beta}_T, \] (1.18)

where

\[ \hat{\beta}_T = (X_T'X_T)^{-1}X_T'R_T = \beta + M_TV_T \] (1.19)

Both the agent and the econometrician have unbiased expectations

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

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

\[ r_t - \hat{r}_t \equiv z_t = x_t(\widetilde{M}_t\widetilde{\xi}_t - M_TV_T) \equiv z_{1t} - z_{2t}. \] (1.20)

There will be four terms that make up each element of the covariance matrix \( \Omega \equiv E[Z_TZ_T'] \).

For each term below, let \( t = 1, \ldots, T \), and \( j = 0, \ldots, t - 1 \). The first component is

\[ E[z_{1t}z_{1t-j}'] = x_t\widetilde{M}_tE[\tilde{\xi}_{t-j}'\tilde{\xi}_{t-j}]\widetilde{M}_t'x_{t-j}, \] (1.21)

\[ = x_t\widetilde{M}_t\sigma^2_tI_{t-j}\widetilde{M}_t'x_{t-j}, \]

\[ = \sigma^2_tx_t(\tilde{X}_t'\tilde{X}_t)^{-1}x_{t-j}. \]

\( I_{t-j} \) is a \( t \times t-j \) identity matrix.
From 1.25, it is clear that the variances will be time varying and related to the explanatory variables. Overall, this covariance matrix is quite complex. The off-diagonal elements exceed when it then starts to cancel to the left. The second line uses the same substitution as in 1.21. \( \tilde{M}_t \) reduces to \((\tilde{X}_t')^{-1}\) until \( k > t \) when it then starts to cancel to the left.

The third term is

\[
E[z_t, z_{t-j}] = \sigma^2 \epsilon_t M_T \left[ \begin{array}{c}
- \tilde{X}_t' x_j \tilde{X}_t \end{array} \right] X_t \tilde{X}_t^{-1} \epsilon_t. 
\] (1.22)

In this case, the elements \( E[V_t V_T' \tilde{z}_t' \tilde{M}_t' M_T x_t, x_t] \) leave an \((\tilde{X}_t')^{-1}\) 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,

\[
E[z_t, z_{t-j}] = \sigma^2 \epsilon_t M_T \left[ \begin{array}{c}
- \tilde{X}_t' x_j \tilde{X}_t \end{array} \right] X_t \tilde{X}_t^{-1} \epsilon_t. 
\] (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 \( t = 1, \ldots, T \). 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,

\[
\gamma_t(t) = E[z_t^2] = E[z_t^2 - z_t z_{t-1}] = \sigma^2 \epsilon_t (\tilde{X}_t')^{-1} x_t'. \]

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.

---

\(^9\)The 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 \( \varepsilon_t \) is Gaussian, the terms are just cross products of \( \gamma_z \),

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

(1.26)

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

\[
z_{2T-j} = x_{T-j} \left( \frac{\sum_{t=1}^{T} x_t^2 (\sum_{t=1}^{T} x_t^2 / \sum_{t=1}^{T} x_t^2)}{\sum_{t=1}^{T} x_t^2} \right).
\]

(1.27)

The inner double sum converges quickly. Consequently,

\[
\gamma_{zz}(t, t-j) \approx E[z_t^2 z_{t-j}^2] = E[x_T \hat{M}_{T-j} \hat{\varepsilon}_j' \hat{M}_j x_{T-j} \hat{M}_{T-j} \hat{\varepsilon}_{T-j} \hat{M}_{T-j} x_{T-j}].
\]

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

\[
E[\hat{\varepsilon}_j' \sum_{t=1}^{T} \varepsilon_t^2] = \mu_4 + \sum_{t \neq j} (\sigma^2 \varepsilon)^2.
\]

(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 \( \phi_1 \). 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 \( t-1 \),

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

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 $y_t = 0.05 + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$.

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

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 $x$'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 $y$'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 $t$. $\varepsilon$ 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 $\varepsilon_t^2$ 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 $y_t = 0.25 + 0.5x_{1t} - 0.25x_{2t} + \varepsilon_t$ and $\varepsilon_t \sim N(0,1)$.

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

1.5.2 Multiple independent regressors

I use a three variable model, $y_t = 0.25 + 0.5x_{1t} - 0.25x_{2t} + \varepsilon_t$. Each $x$ and the $\varepsilon$ 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 $\varepsilon$’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 $y_t = x_t = 0.05 + 0.5x_{t-1} - 0.25x_{t-2} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$.

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

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 $\beta_2$.

<table>
<thead>
<tr>
<th>Lag</th>
<th>AC</th>
<th>PAC</th>
<th>AC</th>
<th>PAC</th>
<th>AC</th>
<th>PAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.157</td>
<td>0.157</td>
<td>0.287</td>
<td>0.287</td>
<td>0.333</td>
<td>0.333</td>
</tr>
<tr>
<td>2</td>
<td>0.138</td>
<td>0.063</td>
<td>0.251</td>
<td>0.140</td>
<td>0.297</td>
<td>0.173</td>
</tr>
<tr>
<td>3</td>
<td>0.063</td>
<td>0.058</td>
<td>0.180</td>
<td>0.109</td>
<td>0.226</td>
<td>0.114</td>
</tr>
<tr>
<td>4</td>
<td>0.043</td>
<td>-0.029</td>
<td>0.162</td>
<td>0.030</td>
<td>0.200</td>
<td>0.035</td>
</tr>
<tr>
<td>5</td>
<td>0.010</td>
<td>-0.003</td>
<td>0.148</td>
<td>0.068</td>
<td>0.193</td>
<td>0.079</td>
</tr>
<tr>
<td>6</td>
<td>-0.009</td>
<td>-0.011</td>
<td>0.119</td>
<td>0.012</td>
<td>0.174</td>
<td>0.039</td>
</tr>
<tr>
<td>7</td>
<td>-0.015</td>
<td>-0.047</td>
<td>0.092</td>
<td>0.015</td>
<td>0.148</td>
<td>0.036</td>
</tr>
<tr>
<td>8</td>
<td>-0.025</td>
<td>-0.036</td>
<td>0.078</td>
<td>-0.011</td>
<td>0.132</td>
<td>0.006</td>
</tr>
<tr>
<td>9</td>
<td>-0.038</td>
<td>-0.019</td>
<td>0.069</td>
<td>0.023</td>
<td>0.123</td>
<td>0.033</td>
</tr>
<tr>
<td>10</td>
<td>-0.043</td>
<td>-0.040</td>
<td>0.061</td>
<td>-0.001</td>
<td>0.111</td>
<td>0.001</td>
</tr>
</tbody>
</table>
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 $h_t = E[\varepsilon_t^2 | \varepsilon_{t-1}^2, h_{t-1}, x_t, \ldots, x_t]$, $x_t$ be a scalar, and write the functional form as

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

(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 $\Sigma a_{2+r,i} = \Sigma b_{2+r,j} = 0$, 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 $\chi^2(4)$.

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

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 significant 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


---

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


