The Stata Journal (2021)
21, Number 2, pp. 295–319 DOI: 10.1177/1536867X211025797

"What good is a volatility model?" A reexamination after 20 years

Christopher F. Baum
Department of Economics
Boston College
Chestnut Hill, MA
baum@bc.edu

Stan Hurn
School of Economics and Finance
Queensland University of Technology
Brisbane, Australia
s.hurn@qut.edu.au

Abstract. This article is primarily a replication study of Engle and Patton (2001, *Quantitative Finance* 1: 237–245), but it also serves as a demonstration of the time-series features introduced into Stata over the past two decades. The dataset used in the original study is extended from the end date of the original sample on 22 August 2000 to 1 August 2017 to examine the robustness of the models.

Keywords: st0637, volatility, GARCH, time series, reproducible research

1 Introduction

The aim of this project is to reproduce Engle and Patton (2001) "What good is a volatility model" 20 years after it was first published in *Quantitative Finance*. The data used in the original article (hereafter referred to as EP) consisting of the Dow Jones Industrial Average Index and the three-month U.S. Treasury Bill rate for the period 23 August 1988 to 22 August 2000 are available for download.[1] The sample is later extended to include data up to 1 August 2017. This classic article is a nice introduction to volatility modeling for students of financial econometrics and represents a good target for reproducible research. It is also a vehicle to demonstrate some of the time-series features introduced in Stata over these two decades.

In a seminal article that is regarded as the starting point of the discipline of financial econometrics, Engle (1982) introduced the concept of autoregressive conditional heteroskedasticity (ARCH) to model a time-varying variance using a simple linear model. A generalization of the model due to Bollerslev (1986) is known as generalized autoregressive conditional heteroskedasticity (GARCH). In its simplest form, the model is given by

$$
\begin{aligned}
y_t &= \mu + u_t \\
u_t &\sim N(0, h_t) \\
h_t &= \omega + \alpha u_{t-1}^2 + \beta h_{t-1}
\end{aligned} \tag{1}
$$

The fundamental property of the model in (1), known as the GARCH(1,1) model, is that the conditional variance h_t is time varying with an autoregressive component, h_{t-1}, and

1. See http://public.econ.duke.edu/~ap172/.

a component driven by unexpected events proxied by the squared disturbance in the previous period, u_{t-1}^2. In this model, the parameter β, where $0 \leq \beta < 1$, determines how past shocks affect the conditional variance at time t. The initial impact of the previous shock on h_t is α. It is this basic model and a few variations to it that EP estimate in their article.

For illustrative purposes, we do not include additional terms in the mean equation for y_t other than the constant term, μ; additional variables can be easily included. Similarly, the assumption of only one lag on both the squared error term, u_t^2, and the conditional variance, h_{t-1}, is only for ease of exposition. Additional terms in $u_{t-2}^2, u_{t-3}^2, \ldots$ could be added to the variance equation as well as additional autoregressive terms $h_{t-2}, h_{t-3}, \ldots$, leading to a GARCH(p,q) model.

The rest of this article is structured as follows. In section 2, we review the data used by EP and highlight the characteristics of financial returns that give rise to GARCH modeling. In sections 3 to 7, we reproduce and explore the results reported by EP in section 3 of their article. Finally, in section 8, we extend the EP dataset from 22 August 2000 to 1 August 2017. The original models stand up well to this extension of the sample period, despite now including episodes of severe turbulence in the stock market.

2 Summary of the data

The daily data are observed only on days when the Dow Jones Index trades. Simply using the dates provided to `tsset` the data will yield a time series with gaps. This means that referring to the lag of the trading date will always use yesterday's date (which may be a missing value) instead of the date of the previous trading day. There are two fixes. The first quick fix is simply to use the observation numbers as the time variable. This device ensures no gaps in the series, but it is a stop-gap approach that does not allow reference to calendar dates in analyzing the data and presenting results. A far better way of dealing with the problem involves creating user-defined business dates.

Designated as `%tb` dates, a business-daily calendar omits all dates on which there is no trading. In the current data, the date variable is called `datevec`; it is a daily date variable with missing values for all nontrading days. It would also be possible to use data in which the nontrading dates do not appear as observations. The code to make a business calendar named `buscal.stbcal` is as follows. After creating the calendar, Stata recognizes the new format `%tbbuscal`, and the Stata variable `bcaldatevec` is used to `tsset` the data.

```
. use englepatton
. generate t = _n
. tsset t
. bcal create buscal, from(datevec)
. bcal load buscal
. generate bcaldatevec = bofd("buscal",datevec)
. assert !missing(bcaldatevec) if !missing(datevec)
. format %tbbuscal bcaldatevec
. tsset bcaldatevec
```

```
. generate double ldj = 100*log(dj_ind)
. generate double djrets = D.ldj
```

The percentage log returns on the Dow Jones Index (`djrets`) are computed as

$$r_t = 100 \times (\log p_t - \log p_{t-1})$$

The summary statistics reported in EP table 1 are reproduced as follows:

```
. tabstat djrets, statistics(mean variance skewness kurtosis) columns(v)
```

stats	djrets
mean	.0550373
variance	.8253352
skewness	-.5266243
kurtosis	9.047384

As this table shows, the index had a small positive average return of about one-twentieth of one percent per day. The daily variance was 0.8253, implying an average annualized volatility of 14.42%. The annualized volatility is computed as $\sqrt{252\,\sigma^2}$, where 252 is the median number of equity trading days per year in the United States and σ^2 is the unconditional variance of the returns. The returns distribution is substantially negatively skewed, and the kurtosis coefficient indicates that the returns distribution has thicker tails than would be found in a Gaussian distribution, which has a kurtosis coefficient of 3. These “fat tails” are commonly found in high-frequency financial time series.

Figures 1 and 2 reproduce the daily index and returns, respectively. These figures illustrate many of the stylized facts about volatility alluded to in section 2 of EP.

Figure 1. The Dow Jones Industrial Index, 23 August 1988 to 22 August 2000

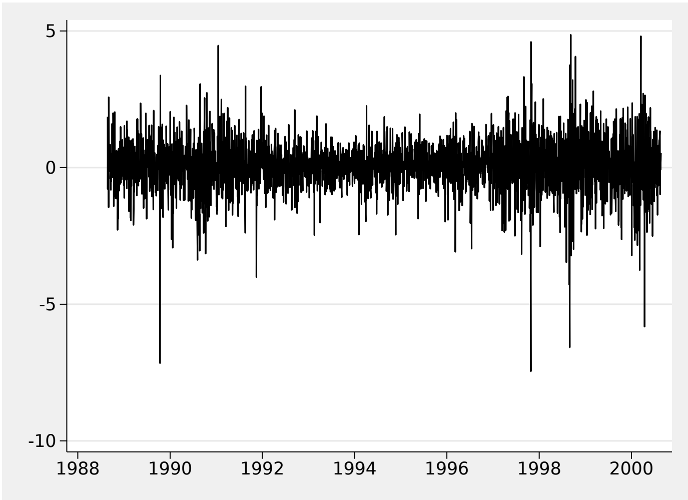

Figure 2. Returns on the Dow Jones Industrial Index, 23 August 1988 to 22 August 2000

1. From figure 1, it is apparent that the variance of the index changes over time as its growth is accompanied by ever-increasing swings.

2. Figure 2 displays volatility clustering in which periods of turbulence and periods of tranquility tend to cluster in time. The implication of such clustering is that volatility shocks today will influence the expectation of volatility many periods in the future.

3. Volatility is mean reverting. Mean reversion in volatility is generally interpreted as implying a normal level of volatility to which volatility will eventually return. Long-run forecasts of volatility should all converge to this same normal level of volatility, no matter when they are made. Thus, the volatility plot in figure 2 shows no trend.

4. Many proposed volatility models impose the assumption that the conditional volatility of the asset is affected symmetrically by positive and negative innovations. In the ARCH(1) and GARCH(1,1) models, for example, the variance is affected only by the square of the lagged innovation, disregarding the sign of that innovation. For equity returns, it is particularly unlikely that positive and negative shocks—"good news" and "bad news"—have the same impact on volatility. In figure 2, many negative returns are substantially larger than the largest positive returns. Assuming that these negative innovations are linked to bad news, it is reasonable to conjecture that bad news has a greater influence on volatility than does good news of a similar size.

Figure 3 presents the correlograms of the returns and the squared returns series, respectively. It is apparent from the correlogram of the returns that there is very little linear dependence in the series. This result is one of the important predictions of the celebrated efficient markets hypothesis (Fama 1970). Briefly, the efficient markets hypothesis states that current stock prices incorporate all relevant information so that all subsequent price changes represent random departures from previous prices. In an efficient market, therefore, the series of returns should show no time dependence. This result is in stark contrast to the correlogram of squared returns, where much stronger dependence is evident. This plot suggests that squared returns—and volatility—may be predictable.

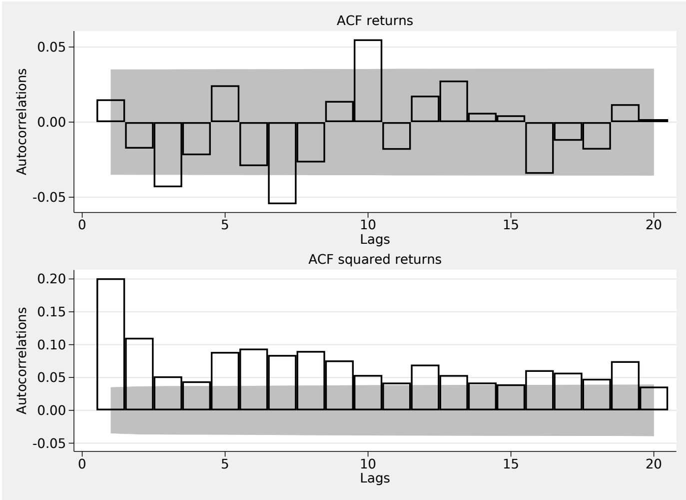

Figure 3. Correlograms of returns and squared returns

3 A volatility model

The parameters of the GARCH(1,1) model in (1) are estimated by maximum likelihood in Stata using the `arch` command. EP base their estimation of the assumption of normally distributed errors, as in (1), so that the log likelihood function for observation t is given by

$$l_t = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log h_t - \frac{1}{2}\frac{u_t^2}{h_t} \tag{2}$$

The starting value for the conditional variance h_{t-1} may be set in a few ways. The Stata default for `arch` is to use the unconditional variance, and this is the method chosen here. Given the values of the skewness and kurtosis coefficients reported previously, the assumption of Gaussian errors is not likely to be supported by the data. The estimates based on the normal log likelihood are therefore known as quasi–maximum likelihood estimates. It turns out that, in the GARCH model, the parameter estimates are still consistent but care must be taken when computing their standard errors. Most econometric packages now routinely support estimation of GARCH models based on different distributional assumptions. In Stata, the `distribution()` option of the `arch` command also supports the t distribution and the generalized error distribution, which allow estimation of the tail thickness of the error distribution.

```
. arch djrets, arch(1) garch(1) vce(robust) vsquish nolog cformat(%6.4f)
ARCH family regression

Sample: 24aug1988 - 22aug2000                   Number of obs     =      3,130
Distribution: Gaussian                          Wald chi2(.)      =          .
Log pseudolikelihood = -3920.313                Prob > chi2       =          .

------------------------------------------------------------------------------
             |             Semirobust
      djrets |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
djrets       |
       _cons |     0.0596     0.0149     4.01   0.000       0.0304      0.0887
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |     0.0371     0.0149     2.50   0.012       0.0080      0.0663
       garch |
         L1. |     0.9546     0.0192    49.64   0.000       0.9169      0.9923
       _cons |     0.0072     0.0047     1.53   0.126      -0.0020      0.0164
------------------------------------------------------------------------------

. estimates store mod1
. predict double h, var
. generate ha = sqrt(252*h)
. predict double resids, res
(1 missing value generated)
. generate stres = resids/sqrt(h)
(1 missing value generated)
. generate stres2 = stres^2
(1 missing value generated)
```

There are two issues of note with these results. The first is that the coefficient estimates obtained here do not quite match those reported in table 2 of EP. The EP estimates of $\widehat{\alpha} = 0.0399$ and $\widehat{\beta} = 0.9505$ are quite similar to those reported here, but this observation masks an important difference. In footnote 4 on page 242 of EP, the authors point out that a t test rejects the null hypothesis that $\widehat{\alpha} + \widehat{\beta} \geq 1$, known as integrated generalized autoregressive conditional heteroskedasticity (IGARCH). The confidence interval for the current estimates provided by the Stata **nlcom** command indicates that this is not the case for the estimates reported here.[2] EP does not report the value of the log-likelihood function at the optimum, thus making it difficult to ascertain which set of estimates are to be preferred. The value of the log-likelihood function obtained by Stata is −3920.313.

2. See section 4 for this calculation.

The second issue of note relates to the standard errors. The standard errors reported in table 2 of EP are significantly smaller than those reported by Stata. This observation is important and masks an issue that is sometimes glossed over. The maximum likelihood estimates of the parameters of the GARCH model, based on the assumption of Gaussian errors, are consistent even if the true distribution of the innovations is not Gaussian. However, the usual standard errors of the estimators are not consistent when the assumption of Gaussian errors is violated. If the parameters of the model are collected into the vector $\boldsymbol{\theta}$, then standard errors can be estimated consistently using the so-called sandwich estimator,

$$\text{VCE}(\theta) = T^{-1} H^{-1}(\boldsymbol{\theta})\, J(\boldsymbol{\theta}) H^{-1}(\boldsymbol{\theta}) \tag{3}$$

where $H(\boldsymbol{\theta})$ is the second derivative of the log-likelihood function and $J(\boldsymbol{\theta})$ is the outer product of the gradients matrix, respectively given by

$$H(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 l_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}, \qquad J(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial l_t}{\partial\boldsymbol{\theta}}\frac{\partial l_t}{\partial\boldsymbol{\theta}'}$$

When using the `vce(robust)` option, Stata's `arch` command reports standard errors based on implementing (3), a task that requires computing both the first and the second derivatives of the log-likelihood function. Bollerslev and Wooldridge (1992) provide a way of expressing $H(\boldsymbol{\theta})$ in terms of first derivatives only. When implemented, the standard errors are known as Bollerslev–Wooldridge standard errors. From (2), the first and second derivatives of the log-likelihood function at time t are given by

$$g_t = -\frac{1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial\boldsymbol{\theta}}\left(1 - \frac{u_t^2}{h_t}\right)$$

$$h_t = -\frac{1}{2}\left\{-\frac{1}{h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}\left(1 - \frac{u_t^2}{h_t}\right) + \frac{1}{h_t}\frac{\partial^2 h_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\left(1 - \frac{u_t^2}{h_t}\right) + \frac{1}{h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}\frac{u_t^2}{h_t}\right\}$$

The conditional expectation of the first derivative taken at $t-1$ is

$$E_{t-1}(g_t) = -\frac{1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial\boldsymbol{\theta}}\left\{1 - E_{t-1}\left(\frac{u_t^2}{h_t}\right)\right\} = 0$$

because the variance of standardized residual u_t^2/h_t is 1 in expectation. The second derivative now takes the simple form

$$E_{t-1}(h_t) = E_{t-1}\left(-\frac{1}{2}\frac{1}{h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}\frac{u_t^2}{h_t}\right)$$

requiring only the first derivatives. A consistent estimate of the matrix $H(\boldsymbol{\theta})$ is

$$H(\boldsymbol{\theta}) = E\left\{E_{t-1}(h_t)\right\} = -\frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}\frac{u_t^2}{h_t}$$

The discrepancy between the standard errors is then probably due to the difference between the Bollerslev–Wooldridge approach that uses first derivatives and the full sandwich estimator used by Stata.[3]

EP states that the choice of a GARCH(1,1) model is based on the Schwarz information criterion (SIC) after fitting GARCH(p, q) models and searching over $p \in [1, 5]$ and $q \in [1, 2]$. The results of a similar search in Stata suggest that a GARCH(2,2) model gives the lowest SIC, which is then estimated.

```
. quietly {
.   noisily display  "p" _col(12)  "q" _col(20) "SIC"
.   noisily display  _dup(30) "-"
.      forvalues p = 1/5 {
.         forvalues q = 1/2 {
.            arch djrets, arch(1/`p´) garch(1/`q´)
.            estat ic
.            mat stats = r(S)
.            noisily display `p´ _col(12) `q´  _col(20) stats[1,6]
.         }
.      }
. }
p          q       SIC
------------------------------
1          1       7872.822
1          2       7880.0442
2          1       7877.3055
2          2       7860.8244
3          1       7873.1162
3          2       7880.0294
4          1       7879.8376
4          2       7884.7481
5          1       7887.8475
5          2       7900.1665
. arch djrets, arch(1/2) garch(1/2) nolog vsquish cformat(%6.4f)
ARCH family regression

Sample: 24aug1988 - 22aug2000                     Number of obs   =      3,130
Distribution: Gaussian                            Wald chi2(.)    =          .
Log likelihood = -3906.266                        Prob > chi2     =          .
```

djrets	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
djrets						
_cons	0.0629	0.0146	4.31	0.000	0.0343	0.0915
ARCH						
arch						
L1.	0.0588	0.0059	10.04	0.000	0.0473	0.0702
L2.	-0.0581	0.0058	-9.96	0.000	-0.0695	-0.0466
garch						
L1.	1.8664	0.0196	95.28	0.000	1.8280	1.9048
L2.	-0.8671	0.0194	-44.71	0.000	-0.9052	-0.8291
_cons	0.0001	0.0000	1.53	0.126	-0.0000	0.0001

3. For further details of the Bollerslev–Wooldridge approach, see Martin, Hurn, and Harris (2013, chap. 20).

The use of simple information criteria in the selection of GARCH models is known to be problematic (Brooks and Burke 2003). Without knowing exactly how EP computed the SIC, it is not possible to further explore the reasons for the discrepancy.[4] Looking at the parameter estimates of the GARCH(2,2) model, however, it seems that although the specification gives a better SIC it does not look particularly sensible, in that the absolute values of the second-order terms are close in magnitude to the first-order terms. Here, therefore, as in most empirical applications, the GARCH(1,1) specification or some variant of a GARCH(1,1) model is a safe option. In this regard, it is also important to consider the work of Hansen and Lunde (2005), who find that the forecasts of conditional variance obtained from this simple model are always difficult to beat.

The final issue EP deal with in this subsection is whether or not the model has captured all of the persistence in the squared residuals. They suggest examining the correlogram of the standardized squared residuals. If the model's specification is adequate, the standardized squared residuals should be serially uncorrelated.

```
. corrgram stres2, lag(20)
                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.0286   0.0286   2.565  0.1093          |                   |
2        0.0193   0.0185  3.7338  0.1546          |                   |
3       -0.0119  -0.0130  4.1786  0.2428          |                   |
4       -0.0076  -0.0073  4.3607  0.3594          |                   |
5       -0.0055  -0.0046  4.4553  0.4859          |                   |
6        0.0020   0.0024  4.4678  0.6136          |                   |
7       -0.0058  -0.0059  4.5743  0.7118          |                   |
8        0.0183   0.0184  5.6263  0.6890          |                   |
9        0.0021   0.0013  5.6398  0.7754          |                   |
10      -0.0071  -0.0081  5.8002  0.8318          |                   |
11      -0.0107  -0.0100  6.1612  0.8624          |                   |
12      -0.0064  -0.0053  6.2895  0.9008          |                   |
13      -0.0148  -0.0141  6.9792  0.9032          |                   |
14       0.0097   0.0103  7.2774  0.9235          |                   |
15      -0.0180  -0.0182  8.2997  0.9112          |                   |
16      -0.0000  -0.0002  8.2997  0.9394          |                   |
17      -0.0070  -0.0064   8.452  0.9559          |                   |
18      -0.0002  -0.0000  8.4521  0.9711          |                   |
19       0.0075   0.0080   8.629  0.9791          |                   |
20      -0.0159  -0.0168  9.4274  0.9774          |                   |
```

The Ljung–Box Q statistic at the twentieth lag of the standardized squared residuals is 9.4274, which is slightly different from the 8.9545 reported by EP. This slight difference is to be expected given that the parameter estimates and hence the standardized residuals differ slightly, but the overall conclusion holds: the standardized squared residuals are indeed serially uncorrelated.

4. EP used an early version of EViews to produce their results. The discrepancy is most likely due to differences in the computation of the SIC with respect to the treatment of constants in the log-likelihood function and the number of observations.

4 Mean reversion and persistence in volatility

The results for the GARCH(1,1) model indicate that the volatility of returns is very persistent, with $\widehat{\alpha} + \widehat{\beta} = 0.9917$. EP find that the sum of these coefficients is 0.9904. One way of measuring the persistence of the process is in terms of the half-life (HL) of volatility, which is defined as the time taken for the volatility to move halfway back toward its unconditional mean following an impulse. Formally, HL is that smallest k for which

$$|h_{t+k|t} - \overline{\sigma}^2| = \frac{1}{2}|h_{t+1|t} - \overline{\sigma}^2| \tag{4}$$

where the long-run level to which volatility reverts is given by

$$\overline{\sigma}^2 = \frac{\omega}{1 - \alpha - \beta} \tag{5}$$

A representation of the k-step-ahead mean-adjusted forecasting equation is given by (see, for example, Zivot [2009] for details)

$$h_{t+k|t} - \overline{\sigma}^2 = (\alpha + \beta)^{k-1}(h_{t+1|t} - \overline{\sigma}^2) \tag{6}$$

Substituting (6) into the definition of HL in (4) gives

$$(\alpha + \beta)^{k-1}|h_{t+1|t} - \overline{\sigma}^2| = \frac{1}{2}|h_{t+1|t} - \overline{\sigma}^2|$$

After simplifying and taking logs, a simple expression for the HL, k, is

$$k \approx \frac{\log(1/2)}{\log(\alpha + \beta)}$$

The EP parameter estimates indicate an HL of 73 trading days, whereas the results reported here suggest an HL of about 84 trading days.

Notice that, from (6), it is apparent that as $k \to \infty$, the volatility forecast tends to $\overline{\sigma}^2$ provided that $\alpha + \beta < 1$. In other words, for the conditional variance to be stationary, the sum $\widehat{\alpha} + \widehat{\beta}$ must be less than 1. If the sum is 1, then the process is known as an IGARCH process (Engle and Bollerslev 1986). Although EP find that the sum is significantly less than 1, the same is not true of the results reported here.

```
. estimates restore mod1
(results mod1 are active now)
. estimates replay mod1
  (output omitted)
. nlcom _b[ARCH:L.arch]+_b[ARCH:L.garch]
       _nl_1:  _b[ARCH:L.arch]+_b[ARCH:L.garch]
```

djrets	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
_nl_1	.9917178	.0062132	159.61	0.000	.9795401	1.003896

```
. test _b[ARCH:L.arch]+_b[ARCH:L.garch] = 1
 ( 1)  [ARCH]L.arch + [ARCH]L.garch = 1
           chi2(  1) =    1.78
         Prob > chi2 =    0.1825
```

Although the unconditional variance of an IGARCH(1,1) process does not exist, Lumsdaine (1996) shows that standard asymptotically based inference procedures are generally valid even in the presence of IGARCH effects.[5]

The unconditional mean of the GARCH(1,1) process in (5) when calculated for the Dow Jones over the sample period turns out to be 0.8542, which implies that the mean annualized volatility over the sample was 14.77%.

```
. nlcom sqrt(252*(_b[ARCH:_cons]/(1-_b[ARCH:L.arch]-_b[ARCH:L.garch])))
       _nl_1:  sqrt(252*(_b[ARCH:_cons]/(1-_b[ARCH:L.arch]-_b[ARCH:L.garch])))
```

djrets	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_nl_1	14.77221	2.035058	7.26	0.000	10.78356	18.76085

This estimate is slightly different from the 14.67% reported by EP, but this is to be expected given the slight discrepancies in parameter estimates. A plot of the annualized conditional volatility estimates over the sample period is given in figure 4. The conditional volatility is very similar to that plotted by EP. In fact, to the naked eye, the plots are identical notwithstanding the slight differences in parameter estimates.

5. Often, the apparent existence of a unit root as in the IGARCH model may be attributable to regime shift in the level of the unconditional variance (Diebold 1986).

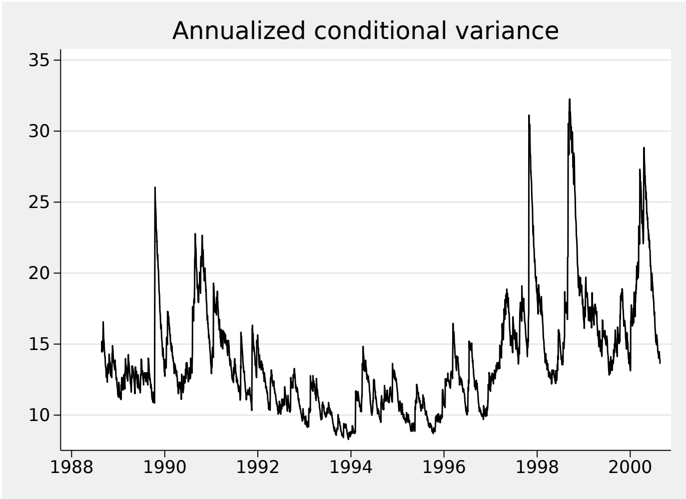

Figure 4. Estimated conditional volatility using a GARCH(1,1) model, August 1988–August 2000

The mean-reverting behavior of conditional volatility is evident in the patterns of dynamic forecasts of volatility. Following EP, dynamic forecasts of annualized daily volatility are produced starting at 23 August 1995 and 22 August 1997, respectively. The first of these forecasts was made at a date with unusually low volatility, and so the forecasts of volatility increase gradually to the unconditional level. The second forecast was made during a period of high volatility. The forecasts of volatility decrease slowly toward the unconditional level of volatility. Figure 5 demonstrates this pattern clearly.[6]

6. In their discussion, EP transpose the high and low periods. It should also be noted that 23 August 1997, mentioned as the start date of the second forecast in EP, is not actually a trading day.

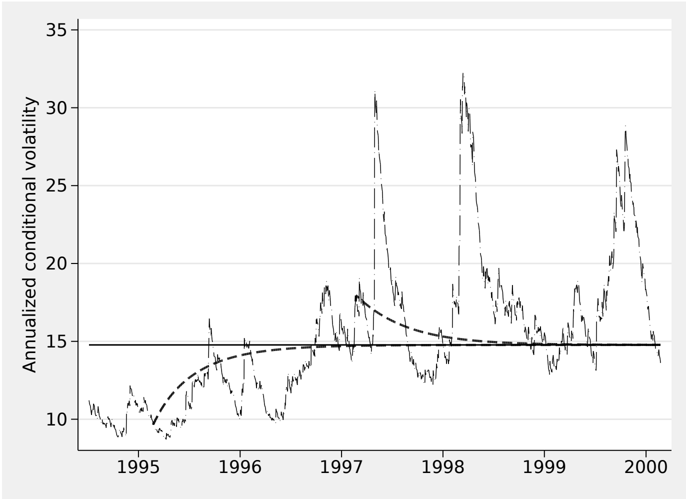

Figure 5. Forecasts of daily return volatility using the GARCH(1,1) model

An alternative way of visualizing the mean reversion of volatility is in terms of figure 6 in EP. Our figure 6 below is based on Stata estimates of the GARCH(1,1) parameters and shows some differences with EP. In particular, the reversion to the mean in EP is not completed even within 200 days. In our figure, the adjustment is completed by about 150 days. The respective HL estimates based on the GARCH models are 73 days (EP) and 84 days (current estimate). Given the size of these half-lives, it seems more appropriate that the adjustment would be complete well before 200 days.

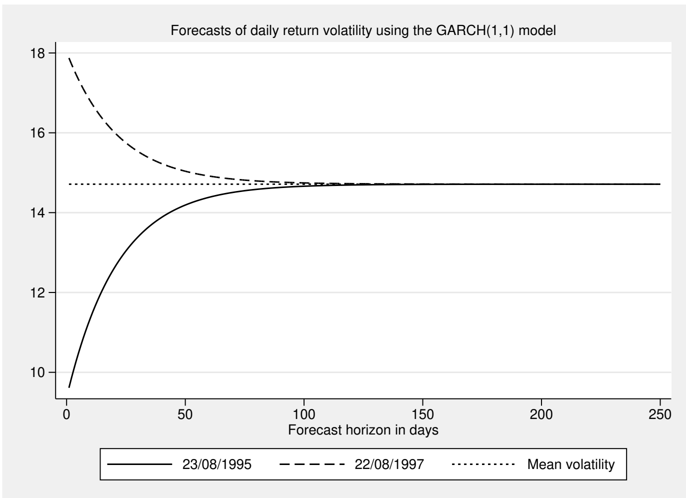

Figure 6. Illustrating mean reversion in the forecasts of daily return volatility using the GARCH(1,1) model

EP suggest examining the volatility of volatility by observing the behavior of the k-period-ahead forecast volatility for different choices of k. In figure 7 below, which is similar to figure 7 of EP, forecasts are presented for horizons of one week (5 days), one quarter (62 days), and one year (252 days). It is expected that the movements in volatility forecasts will become more muted as the horizon increases. At one year ahead, the volatility forecasts should approach the estimated mean obtained from the GARCH(1,1) model of 14.77%. These forecasts are constructed using (6) with the appropriate index k.

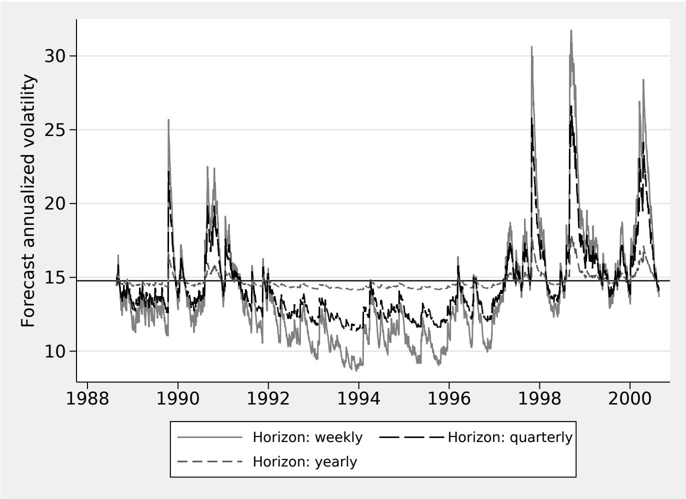

Figure 7. Forecast annualized volatilities for different horizons obtained from the GARCH(1,1) model. The solid horizontal line is the unconditional estimate of annualized volatility obtained from the fitted model of 14.77%.

Just as in the original EP article, it is immediately apparent that the movements at shorter horizons are larger than the movements at longer horizons. This pattern is an implication of the mean reversion in volatility.

5 An asymmetric volatility model

Based on the behavior of returns in figure 2, it was conjectured that the sign of the "news", represented by the prior period's residual, might influence the magnitude of the response in volatility. We can parameterize this concept in many ways, one of which is the threshold GARCH (or TARCH) model. This model was proposed by Glosten, Jagannathan, and Runkle (1993) and Zakoian (1994), motivated by the exponential GARCH model of Nelson (1991).

In Stata, the `tarch()` specification for the conditional variance is

$$h_t = \omega + \alpha u_{t-1}^2 + \phi u_{t-1}^2 I(u_{t-1} > 0) + \beta_2 h_{t-1}$$

where $I(\cdot)$ is the indicator function that takes the value 1 if $(\cdot)$ is true and 0 otherwise. This implies that the coefficients on the news will differ depending on whether news is good or bad:

$$\text{effect of news on variance} \; = \begin{cases} \alpha + \phi & u_{t-1} > 0 \quad \text{good news} \\ \alpha & u_{t-1} \leq 0 \quad \text{bad news} \end{cases}$$

The presence of the leverage effect in Stata’s TARCH model requires that the coefficient ϕ is negative so that bad news has a greater impact on volatility than good news. Asymmetric effects will be present if the estimated ϕ is statistically distinguishable from 0. This specification is the opposite of that used by EP who define the indicator function as $I(u_{t-1} > 0)$. To allow for non-Gaussian errors, we fit the model with a t distribution.

```
. arch djrets, arch(1) garch(1) tarch(1) vce(robust) distribution(t) nolog
> vsquish cformat(%6.4f)
initial values not feasible
(note:  default initial values infeasible; starting ARCH/ARMA estimates from 0)
ARCH family regression
Sample: 24aug1988 - 22aug2000                   Number of obs   =      3,130
Distribution: t                                 Wald chi2(.)    =          .
Log pseudolikelihood = -3765.317                Prob > chi2     =          .

------------------------------------------------------------------------------
             |             Semirobust
      djrets |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
djrets       |
       _cons |     0.0652      0.0127     5.15   0.000       0.0404      0.0901
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |     0.0637      0.0227     2.81   0.005       0.0192      0.1082
       tarch |
         L1. |    -0.0455      0.0224    -2.03   0.042      -0.0895     -0.0016
       garch |
         L1. |     0.9473      0.0179    52.86   0.000       0.9122      0.9825
       _cons |     0.0088      0.0052     1.70   0.090      -0.0014      0.0189
-------------+----------------------------------------------------------------
     /lndfm2 |     1.1984      0.1686     7.11   0.000       0.8680      1.5289
-------------+----------------------------------------------------------------
          df |     5.3149      0.5589                        4.3821      6.6130
------------------------------------------------------------------------------
```

These results confirm the conclusion of EP that the sign of the news has a significant influence on the volatility of returns. The estimate of ϕ is negative and significant, with the effect on volatility summarized as follows:

$$\text{effect of news on variance} \; = \begin{cases} 0.0637 - 0.0455 = 0.0182 & u_{t-1} > 0 \\ 0.0637 & u_{t-1} \leq 0 \end{cases}$$

In other words, bad news at time $t-1$ increases the volatility at time t by 3.5 times as much as good news of the same magnitude. This is a similar effect to that found by EP whose reported leverage effect is about four times greater for bad news. The estimated degrees of freedom of 5.3 strongly rejects Gaussian errors.

6 A model with exogenous volatility regressors

Exogenous regressors are dealt with in Stata by using the het(*varlist*) option of the arch command. Stata adopts a slightly different approach from other econometric packages by specifying that the constant and the exogenous regressors enter the conditional variance equation in exponentiated form. For a single exogenous variable x_t, the conditional variance equation is

$$h_t = \exp(\omega + \gamma x_t) + \alpha u_{t-1}^2 + \beta h_{t-1}$$

This specification allows the x_t variable to take on any values on the real line, while ensuring that the parenthesized expression is strictly positive.

EP used the lagged level of the three-month U.S. Treasury Bill rate as an exogenous regressor in their model of returns, arguing that the Treasury Bill rate is correlated with the cost of borrowing to firms and thus may carry information that is relevant to the volatility of returns. Estimation of the model yields the following results:

```
. arch djrets, arch(1) garch(1) het(m_tbill) vce(robust) vsquish nolog
> cformat(%6.4f)
ARCH family regression -- multiplicative heteroskedasticity
Sample: 24aug1988 - 22aug2000                     Number of obs   =      3,130
Distribution: Gaussian                            Wald chi2(.)    =          .
Log pseudolikelihood = -3909.176                  Prob > chi2     =          .

------------------------------------------------------------------------------
             |              Semirobust
      djrets |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
djrets       |
       _cons |     0.0605     0.0144     4.19   0.000       0.0322      0.0888
-------------+----------------------------------------------------------------
HET          |
     m_tbill |     0.1976     0.0921     2.15   0.032       0.0171      0.3781
       _cons |    -5.4878     0.7639    -7.18   0.000      -6.9849     -3.9907
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |     0.0429     0.0171     2.50   0.012       0.0093      0.0765
       garch |
         L1. |     0.9420     0.0236    39.90   0.000       0.8957      0.9883
------------------------------------------------------------------------------
```

The impact of the lagged Treasury Bill rate is significant but not quite as significant as the EP results suggest. The downside of the estimation of the model in this exponentiated form is that it makes direct comparison with EP difficult. Using the ml command in Stata (see Gould, Pitblado, and Poi [2010] for details), the GARCH model in the form estimated by EP is easily programmed. Using the unconditional variance as the starting value for the conditional variance, the results obtained are as follows.

```
. missings dropobs djrets, force
  (output omitted)
. summarize djrets, detail
  (output omitted)
. global mu0 = r(mean)
. quietly summarize djrets
. global start = r(Var)
. display $start
.82533518
. program garchx
  1. args lnf omega alpha beta psi constant
  2. tempvar err ht
  3. quietly generate double `err´ = djrets-`constant´
  4. quietly generate double `ht´= $start in 1
  5. quietly replace `ht´ = `omega´ + `alpha´*L.`err´^2 + `beta´*L.`ht´
> + `psi´*L.m_tbill in 2/l
  6. quietly replace `lnf´ = -0.5*log(2*_pi) - 0.5*log(`ht´) - 0.5*`err´^2/`ht´
  7. end
. ml model lf garchx  /omega /alpha /beta /psi /constant, vce(robust)
. ml init /constant=$mu0 /omega=0.0010 /alpha=0.03 /beta=0.9 /psi=0.0031
. ml search
  (output omitted)
. ml max, nolog cformat(%6.4f)
                                                Number of obs     =      3,130
                                                Wald chi2(5)      =  186461.80
Log pseudolikelihood = -3908.2469               Prob > chi2       =     0.0000
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/omega	-0.0013	0.0058	-0.22	0.827	-0.0127	0.0101
/alpha	0.0438	0.0176	2.49	0.013	0.0094	0.0783
/beta	0.9391	0.0251	37.39	0.000	0.8899	0.9883
/psi	0.0029	0.0022	1.30	0.193	-0.0015	0.0072
/constant	0.0604	0.0144	4.18	0.000	0.0321	0.0887

The positive sign on ψ, the lagged Treasury Bill coefficient, indicates that higher interest rates are generally associated with higher levels of volatility of equity returns. This result is taken to confirm those reported by Glosten, Jagannathan, and Runkle (1993), who also find that the Treasury Bill rate is positively related to equity return volatility. The problem, however, is that the coefficient estimate of ψ is insignificant. The problem seems to stem from the standard errors: the coefficients are similar to those reported by EP but the robust standard errors are much larger. Reestimating and using standard errors from the outer product of gradients matrix yields results very similar to EP.

```
. ml model lf garchx  /omega /alpha /beta /psi /constant, vce(opg)
. ml init /constant=$mu0 /omega=0.0010 /alpha=0.03 /beta=0.9 /psi=0.0031
. ml search
  (output omitted)
```

```
. ml max, nolog cformat(%6.4f)
                                                 Number of obs     =      3,130
                                                 Wald chi2(5)      = 1572040.50
Log likelihood = -3908.2469                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /omega |    -0.0013     0.0015    -0.87   0.386      -0.0041      0.0016
      /alpha |     0.0438     0.0038    11.40   0.000       0.0363      0.0514
       /beta |     0.9391     0.0062   150.92   0.000       0.9269      0.9513
        /psi |     0.0029     0.0005     6.32   0.000       0.0020      0.0038
   /constant |     0.0604     0.0144     4.19   0.000       0.0321      0.0886
------------------------------------------------------------------------------
```

It seems, therefore, that the standard errors reported in table 5 of EP are not robust.

On reflection, to counter the argument that Stata's convention for dealing with exogenous variables is not as transparent as a simple linear form, there are at least two advantages to the exponentiated form of the `het()` model.

1. The contribution of the exogenous regressors is constrained to be positive. There is, therefore, no instance in which a particular combination of the value of the exogenous variable and its coefficient can cause a negative variance to occur.

2. Imposing this restriction has teased out a significant coefficient on the exogenous regressor when using robust standard errors, a result that is elusive if the nonexponentiated form is used.

7 Aggregation of volatility models

Volatility clustering and non-Gaussian behavior in financial returns is typically seen in weekly, daily, or intraday data. In the final subsection of their empirical example, EP provide evidence consistent with the theoretical result that the empirical results obtained are dependent on the sampling frequency. However, as shown in Drost and Nijman (1993), for GARCH models there is no simple aggregation principle that links the parameters of the model at one sampling frequency to the parameters at another frequency. This means that if a GARCH model is correctly specified for one frequency of data, then it will be misspecified for data with different time scales.

EP fit the simple GARCH(1,1) model on the data, sampled at different frequencies, and compute the HL for each of the models. The results are presented in table 1. While the results indicate that the sampling frequency affects the results in terms of coefficient estimates and HL, they also show that the original estimates presented in EP are quite different from those presented here.

Table 1. Estimates of volatility models over varying horizons

Variable	Daily	2-day	3-day	4-day	Weekly
Constant	0.05955	0.11445	0.16660	0.21348	0.26258
ω	0.00717	0.01375	0.02312	0.02417	0.02072
α	0.03714	0.04193	0.05001	0.04260	0.04039
β	0.95458	0.94982	0.94155	0.95166	0.95733
EP HL	73	68	183	508	365
HL	84	168	246	482	1517

Clearly, these are substantial differences, and while their statistical significance has not been assessed, there is some question as to why the original EP HL estimates are not monotonically increasing; in theory, the persistence of conditional volatility increases with the sampling frequency.

8 Updating the data

To examine how well the volatility models have stood the test of time, the daily dataset for the Dow Jones Index and the U.S. Treasury Bill used in EP are updated to include data to 1 August 2017. The summary statistics for the extended data are as follows.

```
. use englepatton_updated, clear
. generate t = _n
. tsset t
  (output omitted)
. generate double ldj    = 100*log(dj_ind)
. generate double djrets = D.ldj
(1 missing value generated)
. tabstat djrets, statistics(mean variance skewness kurtosis) columns(v)

   stats |    djrets
---------+----------
    mean |  .0317653
variance |  1.075583
skewness | -.2046746
kurtosis |   11.8587
--------------------
```

The small positive average return on the Dow Jones is now even smaller, and the variance is larger. The daily variance of 1.0756 implies an average annualized volatility of 16.46%, which is substantially larger than the 14.42% recorded previously. The returns exhibit slightly less negative skewness but substantially more kurtosis. These changes in summary statistics are consistent with the period of dramatic turbulence experienced during the global financial crisis of 2007–2009.

Table 2 reports the parameter estimates of the three main volatility models fit by EP: the GARCH(1,1), the TARCH(1,1), and the GARCH(1,1)-X with the three-month U.S. Treasury Bill rate used as an exogenous regressor in the conditional variance equation. Robust standard errors based on the Huber/White/sandwich estimator are also reported.

Table 2. Estimates of the GARCH(1,1) models fit by EP using the extended data, 23 August 1988 to 1 August 2017. Robust standard errors are in parentheses.

	GARCH(1,1)	TARCH(1,1)	GARCH(1,1)-X
Constant	0.0548	0.0318	0.0549
	(0.0091)	(0.0087)	(0.0091)
$\omega^{\ddagger}$	0.0147	0.0180	−4.4891
	(0.0037)	(0.0042)	(0.2918)
α	0.0776	0.1302	0.0792
	(0.0106)	(0.0187)	(0.0108)
β	0.9077	0.9110	0.9049
	(0.0122)	(0.0125)	(0.0127)
ϕ		−0.1240	
		(0.0188)	
γ			0.0998
			(0.0546)
log likelihood	73	68	183

‡ The ω coefficient in the GARCH(1,1)-X model enters the conditional variance in exponentiated form.

Overall, the models stand up to estimation on this extended sample remarkably well. Several points of interest evident in these estimates are worth mentioning. Turning first to the GARCH(1,1) model, the persistence of the conditional variance is slightly reduced, with the sum of the ARCH and GARCH coefficients now equal to 0.9854, as opposed to 0.9917 in the original sample. This result reflects the extreme swings of volatility experienced during the crisis period. Contrary to the results reported for the original sample, the sum of α and β in the GARCH model is now significantly less than 1.

```
. estimates restore mod1
(results mod1 are active now)
. nlcom _b[ARCH:L.arch]+_b[ARCH:L.garch]
       _nl_1:  _b[ARCH:L.arch]+_b[ARCH:L.garch]

------------------------------------------------------------------------------
      djrets |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |   .9853579   .0045235   217.83   0.000      .976492    .9942238
------------------------------------------------------------------------------
```

```
. test _b[ARCH:L.arch]+_b[ARCH:L.garch] = 1
 ( 1)  [ARCH]L.arch + [ARCH]L.garch = 1
           chi2(  1) =   10.48
         Prob > chi2 =    0.0012
```

The unconditional mean of the GARCH(1,1) process when calculated for the updated Dow Jones returns data is 15.89%.

```
. nlcom sqrt(252*(_b[ARCH:_cons]/(1-_b[ARCH:L.arch]-_b[ARCH:L.garch])))
       _nl_1:  sqrt(252*(_b[ARCH:_cons]/(1-_b[ARCH:L.arch]-_b[ARCH:L.garch])))
```

djrets	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_nl_1	15.89499	1.418793	11.20	0.000	13.11421	18.67577

This increase in the value of the unconditional mean is also as expected given the effect of the crisis on return volatility.

The second point of interest is that the leverage parameter ϕ is once again negative and statistically significant. Furthermore, the size of the effect is approximately doubled from -0.0609 to -0.1240. The preponderance of bad news during the extended sample period appears to have magnified the leverage effect.

Finally, the estimate of the coefficient γ on the Treasury Bill rate in the GARCH(1,1)-X model is now found to be insignificant, even in the exponentiated specification adopted by Stata. This accords with intuition, because for a large part of this extended sample, short-term interest rates were at or near the 0 lower bound.

A plot of the annualized conditional volatility estimates over the sample period is given in figure 8. It is interesting to note how the peak of the conditional variance during the global financial crisis makes the previous peaks during the earlier sample around the dot-com bubble look rather modest.

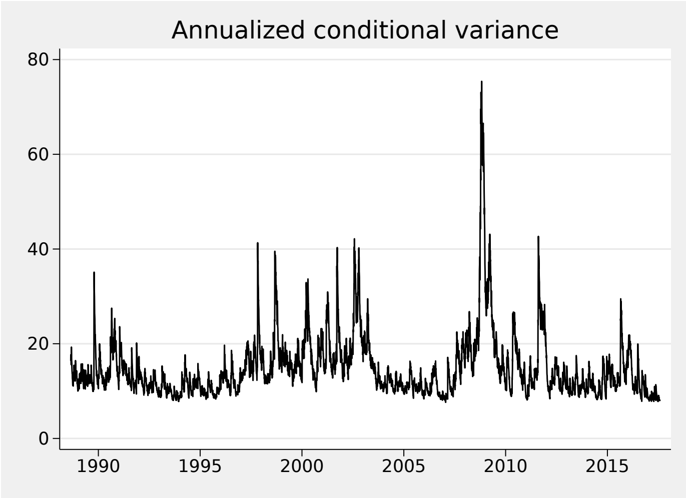

Figure 8. Estimated conditional volatility using a GARCH(1,1) model on the extended dataset, 23 August 1988–1 August 2017

Although no forecasting exercise is undertaken, the conditional variance is strongly mean reverting with an estimated HL of 47 days. This estimate is almost half of the estimate for the earlier sample and is indicative of a more powerful dynamic process for the conditional variance.

9 Conclusion

The aim of the original EP article was to characterize a volatility model in terms of its ability to forecast volatility and also to capture the stylized empirical facts about conditional volatility. Their article succeeds in doing this and also provides an accessible and useful introduction to volatility modeling. In terms of reproducibility, the results reported by EP stand up well to scrutiny and bring out some differences that prompt thought, particularly with respect to computing standard errors in GARCH models. Interestingly, the GARCH(1,1) model fit on updated data is very similar in terms of coefficient estimates, although the conditional variance process appears to be substantially less persistent when estimated over the longer sample. The model performs well in capturing the volatility around the global financial crisis and the turbulence in the markets during that period.

10 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 21-2
. net get st0637          (to install ancillary files)
```

11 References

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–327. https://doi.org/10.1016/0304-4076(86)90063-1.

Bollerslev, T., and J. M. Wooldridge. 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11: 143–172. https://doi.org/10.1080/07474939208800229.

Brooks, C., and S. P. Burke. 2003. Information criteria for GARCH model selection. *European Journal of Finance* 9: 557–580. https://doi.org/10.1080/1351847021000029188.

Diebold, F. X. 1986. Modeling the persistence of conditional variances: A comment. *Econometric Reviews* 5: 51–56. https://doi.org/10.1080/07474938608800096.

Drost, F. C., and T. E. Nijman. 1993. Temporal aggregation of GARCH processes. *Econometrica* 61: 909–927. https://doi.org/10.2307/2951767.

Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50: 987–1007. https://doi.org/10.2307/1912773.

Engle, R. F., and T. Bollerslev. 1986. Modelling the persistence of conditional variances. *Econometric Reviews* 5: 1–50. https://doi.org/10.1080/07474938608800095.

Engle, R. F., and A. J. Patton. 2001. What good is a volatility model? *Quantitative Finance* 1: 237–245. https://doi.org/10.1088/1469-7688/1/2/305.

Fama, E. F. 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25: 383–417. https://doi.org/10.2307/2325486.

Glosten, L. R., R. Jagannathan, and D. E. Runkle. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48: 1779–1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x.

Gould, W., J. Pitblado, and B. Poi. 2010. *Maximum Likelihood Estimation with Stata*. 4th ed. College Station, TX: Stata Press.

Hansen, P. R., and A. Lunde. 2005. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* 20: 873–889. https://doi.org/10.1002/jae.800.

Lumsdaine, R. L. 1996. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64: 575–596. https://doi.org/10.2307/2171862.

Martin, V., S. Hurn, and D. Harris. 2013. *Econometric Modelling with Time Series: Specification, Estimation and Testing*. New York: Cambridge University Press.

Nelson, D. B. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59: 347–370. https://doi.org/10.2307/2938260.

Zakoian, J.-M. 1994. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18: 931–955. https://doi.org/10.1016/0165-1889(94)90039-6.

Zivot, E. 2009. Practical issues in the analysis of univariate GARCH models. In *Handbook of Financial Time Series*, ed. T. G. Andersen, R. A. Davis, J.-P. Kreiß, and T. Mikosch, 113–155. Berlin: Springer. https://doi.org/10.1007/978-3-540-71297-8_5.

About the authors

Christopher F. Baum is a professor of economics and social work at Boston College and a DIW research fellow at the German Institute for Economic Research. He is the author of two Stata Press books and maintains the Statistical Software Components archive of community-contributed software.

Stan Hurn is a professor of econometrics in the School of Economics and Finance at Queensland University of Technology.