GARCH Estimation and Discrete Stock Prices

Henrik Amilon*
Department of Economics
Lund University
P.O. Box 7082
SE-220 07 Lund
Sweden
Phone: +46(0)462227909
Fax: +46(0)462224118
E-mail: henrik.amilon@nek.lu.se

Abstract

The continuous-state GARCH model is misspecified if applied to returns calculated from discrete price series. This paper proposes modifications of the above model for handling such cases. The focus is on the AR-GARCH framework, but the same ideas could be used for other stochastic processes as well. Using Swedish stock price data and a stochastic optimization algorithm, simulated annealing, I compare the parameter estimates and asymptotic standard errors from the approximative model and the extended models. I find small deviations between the models for longer time series and small tick sizes, but larger differences for shorter series and for larger tick size to price ratios, mainly in the conditional variance parameter estimates. None of the models provide continuous residuals. By constructing generalized residuals, I show how valid residual diagnostic and specification tests can be performed in some cases.

Key words: EM estimation, compass rose, stock return modeling, latent variables, generalized residuals
JEL Codes: C35, C51, C52

*The author is grateful to Tobias Rydén for indispensable help on the statistical inference section in the paper. Also, thanks to Hossein Asgharian for providing the data, and to Birger Nilsson for helpful discussion on simulated annealing.

1 Introduction

An implicit assumption in modeling financial asset returns is that the dependent variable is continuous. This is reasonable if the aim is to model stock index returns, for example, but may be inappropriate in modeling returns from an individual asset, like a stock, since asset prices are quoted in discrete units or ticks. For some assets, tick sizes might be small enough to justify a continuous-state model, but for other securities, such as low-priced stocks, the effects of discreteness are not necessarily negligible. It should also be clear that discreteness is less of a problem for coarser-sampled data, but becomes more relevant for high frequency data or, ultimately, transaction data, since the price changes between two consecutive trading points are then rarely greater than a few ticks.

The breakdown of the continuous-state modeling in analyzing transaction data has lead some researchers to use statistical tools developed for cases where the dependent variable is discrete, as the price change in the ordered probit model of Hausman et al. (1992). Another line of work is found in Rydberg and Shephard (1998), who recognize that the conditional distribution of price changes can be quite complicated. In order to make their analysis tractable and interpretable, they decompose the conditional distribution in three parts: activity, direction, and size of the price moves. Each part is then constructed using autologistic or Poisson based generalized linear autoregressive moving average (GLARMA) models, which are described in Shephard (1994). To some extent related work, although in a multinomial setup, is given in the paper of Engle and Russel (1998).

The above references analyze price changes, and thus assume that the conditional distribution of price changes is stationary. This may be appropriate when dealing with short time periods of high frequency data, where the prices move around approximately the same price level, but the assumption is likely to fail when investigating series with price fluctuations around different price levels. In such cases it is more reasonable to assume that the conditional distribution of returns is stationary, as is done here.

For intradaily data, the effects of discreteness are clearly visible in a scatter plot of returns versus lagged returns. Recently, Crack and Ledoit (1996) found that a geometrical pattern also is present in scatter plots of daily data. This compass rose shaped pattern is a direct consequence of the discreteness of prices since the stock return will, by necessity, fall on a grid, given a certain price. As showed by Szpiro (1998), when the price evolves over time, the superposition of the grids at each price level will result in a scatterplot shaped as a compass rose. In a series of simulations, Amilon and Byström (1999) found that continuous-state modeling of stock returns is likely to have little impact on parameter estimates and some specification testing, provided the tick size is reasonably small compared to the price level. For heavier rounded

prices, or equivalently for lower priced stocks, the effects are more serious. Amilon and Byström discovered distortions, particularly in the distributions of the nonparametric portmanteau test of Brock et al (1996), for tick size to price ratios around 1/50, a by no means unusually large ratio in many modern stock markets.

In this paper, I extend the continuous-state GARCH model to handle cases where the return series are calculated from discrete prices by modeling the dependent variable as an unobservable stochastic variable with certain observed outcomes. Two alternative specifications are considered. In the first, the stochastic process is driven by the observable returns, while the other is generated from the expectations of the unobservable data conditioned on all available (observable) information. The first model is inspired by the work of Hausman et al. (1992) and its log-likelihood function is easily derived, while the other is more complicated and calls for approximations. Following the ideas in Morgan and Trevor (1999), the log-likelihood function in this case is obtained by replacing the latent variables by their conditional expectations. Morgan and Trevor argue that such a model can be motivated by the EM algorithm of Dempster et al. (1977). The latter framework has a closer connection to an underlying equilibrium price model which is hidden by the discrete prices[1], but it does not necessarily mean that it is a better statistical model. It it depends on whether the driving forces are the unobservable or observable returns, and if the EM approximation is justified or not.

The parameter estimates and standard errors from the extended models are compared to the misspecified continuous-state model. To circumvent the problem of local maxima, I use a global optimization algorithm called simulated annealing. In this way, I reduce the possibility of any differences in the estimates being due to the potential existence of local maxima, an issue noticed in Glosten et al. (1993), and more thoroughly treated in Doornik and Ooms (2000). The focus is on the GARCH framework of Bollerslev (1986), as this is often used in modeling financial asset returns, see e.g. Bollerslev et al. (1992), although the same ideas could be used for other stochastic processes as well. Caution must be taken in specification tests, since none of the models provide continuous residuals. In the spirit of Gourieroux et al. (1985, 1987), I construct generalized residuals and show how valid specification tests can be performed.

In section 2, I describe the models. Section 3 presents the results of the different estimations, while section 4 contains the effects of discreteness on specification testing. A summary and concluding remarks are found in section 5.

[1] This was suggested to me by Neil Shephard, who also brought the work of Morgan and Trevor (1999) to my attention.

2 The Models

Let $\tilde{P}_t$ be the unobserved equilibrium price of a stock at time t, D_t its dividend, and $\tilde{r}_t$ the unobserved return of the stock, that is $\tilde{r}_t = \ln((\tilde{P}_t + D_t)/\tilde{P}_{t-1})$ with logarithmic returns, and $\tilde{r}_t = (\tilde{P}_t + D_t)/\tilde{P}_{t-1} - 1$ with percentage returns. The widely used continuous-state GARCH(1,1) model then takes the form

$$\begin{aligned} \tilde{r}_t &= \tilde{m}_t + \tilde{\varepsilon}_t, \qquad\qquad \tilde{\varepsilon}_t \mid I_{t-1} \backsim IID(0, \sigma_t^2) \\ \sigma_t^2 &= \gamma_0 + \gamma_1 \tilde{\varepsilon}_{t-1}^2 + \gamma_2 \sigma_{t-1}^2, \end{aligned} \tag{1}$$

where the conditional mean process in the following equals $\tilde{m}_t = \beta_0 + \beta_1 \tilde{r}_{t-1} + \beta_2 \tilde{r}_{t-2}$, and I_t is the unobserved information set at time t, that is $I_t = \left\{\tilde{P}_t, D_t, \tilde{P}_{t-1}, ...\right\}$. The log-likelihood function for this model is given by

$$\ln \tilde{L}\left(\tilde{r}_1, ..., \tilde{r}_T \mid I_0\right) = \sum_{t=1}^{T} \ln \tilde{l}\left(\tilde{r}_t | I_{t-1}\right) = -\frac{1}{2} \sum_{t=1}^{T} \left(\ln\left(2\pi\right) + \ln\left(\sigma_t^2\right) + \frac{\tilde{\varepsilon}_t^2}{\sigma_t^2} \right). \tag{2}$$

This model can, of course, not be estimated since $\tilde{P}_t$, and hence $\tilde{r}_t$, is not revealed to us. One simple approach would be to ignore that prices are discrete and estimate

$$\begin{aligned} r_t &= m_t + \varepsilon_t, \qquad\qquad \varepsilon_t \mid \Psi_{t-1} \backsim IID(0, \sigma_t^2) \\ \sigma_t^2 &= \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 \sigma_{t-1}^2, \end{aligned} \tag{3}$$

where r_t is computed from the observed prices, $\{P_s\}_1^T$, $m_t = \beta_0 + \beta_1 r_{t-1} + \beta_2 r_{t-2}$, and $\Psi_t = \left\{P_t, D_t, P_{t-1}, ...\right\}$ is the observed information set. This model is obviously misspecified, since ε_t cannot be continuous if r_t is discretely distributed around the conditional mean. A potential candidate for a correct model would be the following:

$$\begin{aligned} r_t^* &= m_t + \varepsilon_t^*, \qquad\qquad \varepsilon_t^* \mid \Psi_{t-1} \backsim IID(0, \sigma_t^2) \\ \sigma_t^2 &= \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 \sigma_{t-1}^2, \end{aligned} \tag{4}$$

where r_t^* is an unobservable, or latent, continuous random variable. Note that the expression for the conditional variance still is as in (3), that is $\varepsilon_t = r_t - m_t$.

The basic idea, which is illustrated in figure 1, is that we do not know r_t^*, but given the prices and the tick size, we can partition the state space of r_t^* in such a way that we know the boundaries, α_{1t} and α_{2t}, comprising r_t^*. Using a *rounding* procedure is one way of determining the boundaries. Say, for instance, that $P_{t-1} = 100, P_t = 101, D_t = 0$, and the tick size is $h = 1$. The observed percentage return is then 1%, but we can only conclude that r_t^* is between

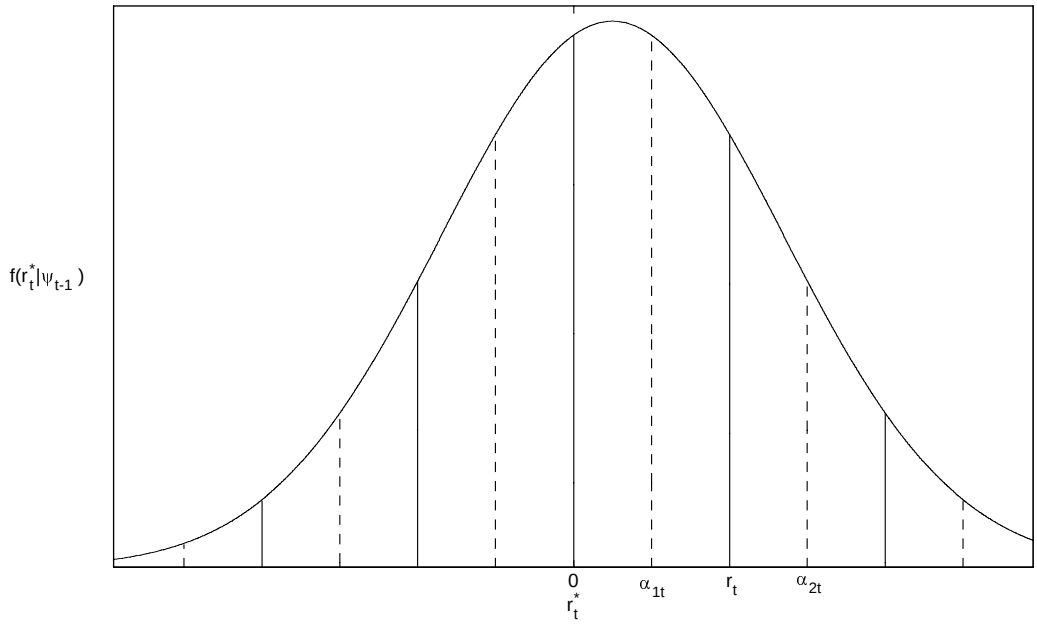

Figure 1: The figure shows the conditional distribution of r_t^*. The observable states are marked with solid lines. The dashed lines partition the conditional state space of r_t^*. The boundaries comprising the realized return r_t are denoted α_{1t} and α_{2t}.

$\alpha_{1t} = (101 - 0.5 - 100)/100 = 0.5\%$, and $\alpha_{2t} = (101 + 0.5 - 100)/100 = 1.5\%$. More generally, modeling percentage returns, the boundaries are given by

$$\alpha_{1t} = \frac{P_t + D_t - h/2 - P_{t-1}}{P_{t-1}} \qquad \text{and} \qquad \alpha_{2t} = \frac{P_t + D_t + h/2 - P_{t-1}}{P_{t-1}}, \tag{5}$$

and for the more frequently used logarithmic returns by

$$\alpha_{1t} = \ln\left(\frac{P_t + D_t - h/2}{P_{t-1}}\right) \qquad \text{and} \qquad \alpha_{2t} = \ln\left(\frac{P_t + D_t + h/2}{P_{t-1}}\right). \tag{6}$$

The conditional distribution of r_t can now be expressed by assuming a suitable distribution of ε_t^*. With Gaussian error terms it is given by

$$\Pr(r_t \mid \Psi_{t-1}) = \Pr(\alpha_{1t} \leq m_t + \varepsilon_t^* < \alpha_{2t} \mid \Psi_{t-1}) = \Phi\left(\frac{\alpha_{2t} - m_t}{\sigma_t}\right) - \Phi\left(\frac{\alpha_{1t} - m_t}{\sigma_t}\right), \tag{7}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Once $\Pr(r_t \mid \Psi_{t-1})$ has been specified, the log-likelihood function for a sample of T observations will be

$$\ln L(r_1, ..., r_T \mid \Psi_0) = \sum_{t=1}^{T} \ln l\left(r_t | \Psi_{t-1}\right) = \sum_{t=1}^{T} \ln\left(\Phi\left(\frac{\alpha_{2t} - m_t}{\sigma_t}\right) - \Phi\left(\frac{\alpha_{1t} - m_t}{\sigma_t}\right)\right).$$

We can actually introduce an additional degree of freedom. The rounding procedure would be correct if the relationship between the virtual (logarithmic) return process and the price process is defined as $r_t^* \equiv \ln\left((\tilde{P}_t + D_t)/P_{t-1}\right)$, but this is an unnecessary restriction. What we seek is the mapping between the conditional distribution of r_t^* and that of r_t, where the aim is to partition the conditional state space of r_t^* in such a way that when the price process is between two boundaries, r_t is observed. In other words, the boundaries are not required to be restricted as above, but they must be non-overlapping. To be more specific, we can introduce yet another parameter, H, such that the expressions in (6) are replaced by

$$\alpha_{1t} = \ln\left(\frac{P_t + D_t - (h - H)}{P_{t-1}}\right) \qquad \text{and} \qquad \alpha_{2t} = \ln\left(\frac{P_t + D_t + H}{P_{t-1}}\right), \tag{8}$$

and similarly for the boundaries in (5). If $H = h/2$, (6) and (8) are thus identical. This model follows closely the ideas in Hausman et al. (1992).

One may argue that if one believes that there exists an underlying equilibrium price model which is truncated, then the connection between (1) and (4) is rather loose. We can, however, as shown in section 4, test the model in (4) to one were r_{t-j}^* and $\varepsilon_{t-j}^{*2}, j \geq 1$, are included in the conditional mean and variance processes, respectively. By doing so, we can test the specification in (4) to the one in (1).

In order to address this issue more thoroughly, let us consider the following model:

$$\begin{aligned} \tilde{r}_t &= E\left[\tilde{m}_t|\Psi_{t-1}\right] + \tilde{\varepsilon}_t, \qquad \tilde{\varepsilon}_t \mid \Psi_{t-1} \backsim N(0, \sigma_t^2) \\ \sigma_t^2 &= \gamma_0 + \gamma_1 E\left[\tilde{\varepsilon}_{t-1}^2|\Psi_{t-1}\right] + \gamma_2 \sigma_{t-1}^2, \end{aligned} \tag{9}$$

To some extent, Morgan and Trevor (1999) estimate similar models in the related problem of censored GARCH returns[2]. Inspired by the EM algorithm of Dempster et al. (1977), they replace the unobservable elements of the log-likelihood function (2) by their expectations conditioned on the information available to traders:

$$\ln L = -\frac{1}{2}\sum_{t=1}^{T}\left(\ln(2\pi) + \ln\left(\sigma_t^2\right) + \frac{E\left[\tilde{\varepsilon}_t^2|\Psi_t\right]}{\sigma_t^2}\right). \tag{10}$$

As Morgan and Trevor carefully explain, if $\Psi_t = I_t$, maximizing (10) will produce maximum likelihood (ML) estimates of the model (1). Since $\Psi_t \neq I_t$, we will only obtain approximative

[2]Some comments on the Morgan and Trevor (1999) paper: In their analysis, the conditional mean is reduced to constant only. If this is not the case, their comparison of sizes of the observed and unobserved errors in table 1, p. 399, will no longer hold, and the rest of their analysis is likely to fail.

On the same page, they introduce another source of approximation. When they move σ_t outside the expectation operator, they implicitly assume that the conditional variance of the censored errors is given by (9), in contrast to the uncensored errors, where it is given by (1).

ML estimates, unless the true model is given by (9), in which case (10) could be seen as the exact log-likelihood function, at least for Gaussian errors.

In order to estimate the model, we need to calculate $E\left[\tilde{\varepsilon}_t|\Psi_t\right]$, which is the generalized error in Gourieroux et al. (1987), a topic I return to in the specification tests in section 4. With Gaussian error terms, it is the expectation of a variable with a doubly truncated normal distribution, and can conveniently be expressed as (see Johnson and Kotz (1970)):

$$E\left[\tilde{\varepsilon}_t|\Psi_t\right] = \sigma_t \frac{\varphi(c_{1t}) - \varphi(c_{2t})}{\Phi(c_{2t}) - \Phi(c_{1t})}, \tag{11}$$

with

$$c_{1t} = \frac{1}{\sigma_t}\left(\tilde{\alpha}_{1t} - E\left[\tilde{m}_t|\Psi_t\right]\right) \qquad \text{and} \qquad c_{2t} = \frac{1}{\sigma_t}\left(\tilde{\alpha}_{2t} - E\left[\tilde{m}_t|\Psi_t\right]\right), \tag{12}$$

where $\varphi(\cdot)$ is the standard normal probability density function. Conditional on all observable prices (including P_t), we know the lower and upper limits of $\tilde{r}_t$. Hence,

$$\tilde{\alpha}_{1t} = \ln\left(\frac{P_t + D_t - h/2}{P_{t-1} + h/2}\right) \qquad \text{and} \qquad \tilde{\alpha}_{2t} = \ln\left(\frac{P_t + D_t + h/2}{P_{t-1} - h/2}\right). \tag{13}$$

We also need to compute $E\left[\tilde{\varepsilon}_t^2|\Psi_t\right]$. The variance of a doubly truncated normally distributed variable is given in Johnson and Kotz (1970) as

$$Var\left[\tilde{\varepsilon}_t|\Psi_t\right] = \sigma_t^2\left(1 + \frac{c_{1t}\varphi\left(c_{1t}\right) - c_{2t}\varphi\left(c_{2t}\right)}{\Phi(c_{2t}) - \Phi(c_{1t})} - \left(\frac{\varphi(c_{1t}) - \varphi(c_{2t})}{\Phi(c_{2t}) - \Phi(c_{1t})}\right)^2\right),$$

yielding

$$E\left[\tilde{\varepsilon}_t^2|\Psi_t\right] = \sigma_t^2\left(1 + \frac{c_{1t}\varphi\left(c_{1t}\right) - c_{2t}\varphi\left(c_{2t}\right)}{\Phi(c_{2t}) - \Phi(c_{1t})}\right). \tag{14}$$

With these conditional expectations at hand, we compute σ_t^2 from (9), $\tilde{\alpha}_{1t}$ and $\tilde{\alpha}_{2t}$ from (13), $E\left[\tilde{m}_t|\Psi_t\right] = \beta_0 + \beta_1 E\left[\tilde{r}_{t-1}|\Psi_{t-1}\right] + \beta_2 E\left[\tilde{r}_{t-2}|\Psi_{t-2}\right]$, c_{1t} and c_{2t} from (12), $E\left[\tilde{\varepsilon}_t|\Psi_t\right]$ and $E\left[\tilde{\varepsilon}_t^2|\Psi_t\right]$ from (11) and (14), maximize (10), and finally update $E\left[\tilde{r}_t|\Psi_t\right] = E\left[\tilde{m}_t|\Psi_t\right] + E\left[\tilde{\varepsilon}_t|\Psi_t\right]$. In the algorithm, $E\left[\tilde{r}_1|\Psi_1\right] = r_1$ and $E\left[\tilde{r}_2|\Psi_2\right] = r_2$, while the initial value for σ_3^2 is estimated as a parameter.

3 The Estimates

How serious is a neglect of the effects of discreteness? To check this, I use the daily prices of four major stocks traded at the Stockholm Stock Exchange over different periods of time[3].

[3]These stocks are chosen for convenience only. No stock splits occurred during the respective time periods.

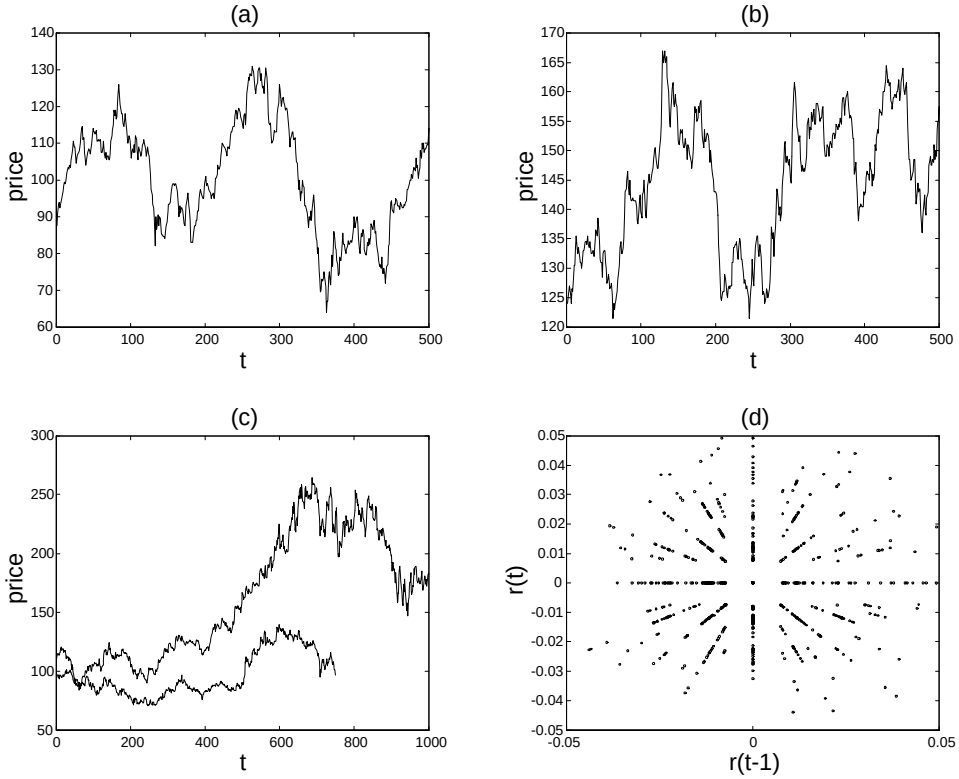

Figure 2: The price series for: (a) ABB, (b) SKF, and (c) AC (1001 observations) and TRE. A scatterplot of r_t vs. r_{t-1} is shown in (d) for TREr.

The stocks and sample periods are ABB A (970421-990422, 501 observations), SKF B (950102-961227, 501 observations), Trelleborg B (950102-971230, 751 observations), and Atlas Copco A (950102-981230, 1001 observations). All price series have a tick size of 0.5. To capture the effects of a higher tick size to price ratio, which is more representative for some lower priced stocks, I construct an artificial price series by dividing the Trelleborg B prices by 2, and round them to the nearest integer. We thus have four price series with tick sizes of 0.5, and one with $h = 1$, from which we compute logarithmic returns. Henceforth, I denote the series ABB, SKF, AC, TRE, and TREr. Figures 2a-c show the original price series, and figure 2d shows a scatterplot of r_t versus r_{t-1} for TREr. The compass rose shaped pattern of Crack and Ledoit (1996) is clearly visible.

The series are fitted to the misspecified model (3) and to the modified models (4) and (9). As mentioned above, the conditional mean process is AR(2). The lag length is chosen ad hoc,

but in the specification testing in section 4, I investigate the correctness of this choice.

In order to circumvent the problem of local maxima, see Glosten et al. (1993), and Doornik and Ooms (2000), all estimations are conducted using a stochastic optimization technique originating from thermodynamics, called simulated annealing. It is inspired by the incredible ability of nature to find the lowest energy configurations in phase transitions, for example when liquid metal cools and solidifies. It is, however, only after slow cooling (annealing) that the metal arrives at a low energy state. Otherwise, the system may end up in a configuration containing higher energy, that is, the analogy of being trapped in a local maxima. The key idea of simulated annealing, which Metropolis et al. (1953) were the first to transform into numerical purposes, is to imitate the way of nature to allow for transitions to higher energy states in the search for the minimal state. The decision rule for accepting higher energy states (or equivalently, lower values of the log-likelihood function), the Metropolis criteria, has been implemented in several continuous optimization algorithms, see Press et al. (1992) and Goffe et al. (1994) for further references. My choice, a modification of the downhill simplex method, is the one provided in Press et al[4]. The estimates from simulated annealing are then used as starting values in a conventional maximization algorithm, from which the final estimates as well as the asymptotic maximum likelihood covariance matrices are extracted[5].

The parameter estimates and the asymptotic standard errors from the different time series and model specifications are presented in table 1[6]. The notations 'model$_C$', 'model$_{D1}$', and 'model$_{D2}$' refer to models (3), (4) with a parameterized H, and (9) respectively. The estimates of β_1 and β_2, as well as their standard errors, are quite similar across the different models. The difference in the parameter estimates of the conditional variance process are more notable, especially for the shorter series ABB and SKF. We see that $\hat{\gamma}_0$ (ABB) and $\hat{\gamma}_1$ (SKF) are much smaller for model$_{D1}$ compared to the estimates of the other two models. For these series, $\hat{H}$ is very different from $h/2$, although only significantly for ABB. Looking at TRE ($h = 0.5$), the

[4]Basically, three parameters are important in the simulated annealing algorithm. The initial temperature, $T_{start} = 10$, the geometric temperature reduction, $r_T = 0.95$, and the maximum number of function evaluations at each temperature, $n_{iter} = 1000\times$ the number of parameters to estimate.

[5]The conventional algorithm is either the BHHH or the Newton-Raphson method in GAUSS. The covariance matrix is computed as the inverse of the Hessian. For the series TRE and TREr, the Hessian is not positive definite for the extended models. The covariance matrix in these cases are computed as the inverse of the cross-product of the first derivatives. To facilitate the comparisons, all covariance matrices for the three models for TRE and TREr are computed in this way. These problems may be overcome by calculating the Hessian from first derivatives only.

[6]The estimates of the initial values in the conditional variance processes are of less importance and therefore not presented.

Table 1: Maximum likelihood estimates and log-likelihood values of the different models. The notations 'model$_C$', 'model$_{D1}$', and 'model$_{D2}$' refer to models (3), (4) with a parameterized H, and (9) respectively. Small numbers are asymptotic standard errors.

Stocks	Models	$\beta_0 \times 10^3$	β_1	β_2	$\gamma_0 \times 10^4$	γ_1	γ_2	H	$\ln L$
	model$_C$	1.898 0.896	0.097 0.049	0.083 0.047	0.328 0.133	0.159 0.040	0.784 0.051		1204
ABB	model$_{D1}$	9.782 3.030	0.087 0.045	0.075 0.045	0.179 0.107	0.129 0.035	0.817 0.043	1.131 0.324	-1427
	model$_{D2}$	1.868 0.879	0.098 0.049	0.082 0.048	0.291 0.125	0.169 0.042	0.775 0.051	0.250	1207
	model$_C$	0.416 0.594	0.165 0.047	0.013 0.047	0.553 0.319	0.064 0.041	0.630 0.191		1444
SKF	model$_{D1}$	-13.84 9.121	0.177 0.046	0.020 0.047	0.513 0.348	0.020 0.022	0.676 0.199	-1.787 1.285	-1377
	model$_{D2}$	0.409 0.582	0.170 0.048	0.015 0.049	0.470 0.347	0.064 0.043	0.653 0.227	0.250	1445
	model$_C$	0.681 0.504	0.145 0.033	0.006 0.033	0.092 0.049	0.073 0.022	0.899 0.034		2645
AC	model$_{D1}$	-0.123 0.342	0.145 0.033	0.008 0.032	0.094 0.052	0.074 0.023	0.896 0.036	0.131 0.085	-3070
	model$_{D2}$	0.682 0.435	0.148 0.033	0.006 0.012	0.072 0.055	0.073 0.026	0.902 0.042	0.250	2651
	model$_C$	0.150 0.660	0.010 0.042	-0.030 0.045	0.609 0.513	0.122 0.034	0.695 0.182		1957
TRE	model$_{D1}$	-0.249 3.607	0.010 0.042	-0.030 0.045	0.620 0.509	0.124 0.035	0.689 0.182	0.211 0.337	-1976
	model$_{D2}$	0.184 0.615	0.013 0.039	-0.033 0.046	0.595 0.404	0.140 0.038	0.658 0.163	0.250	1966
	model$_C$	0.144 0.682	-0.017 0.042	-0.031 0.045	0.526 0.598	0.093 0.030	0.755 0.188		1939
TREr	model$_{D1}$	-1.417 3.711	-0.015 0.043	-0.032 0.045	0.542 0.507	0.102 0.032	0.711 0.179	0.425 0.175	-959.5
	model$_{D2}$	0.031 0.717	0.009 0.042	-0.010 0.050	0.460 1.194	0.013 0.007	0.675 0.792	0.500	2142

estimates and standard errors for model$_{D2}$ differ somewhat more in comparison to model$_C$ and model$_{D1}$. The same observation is noted for TREr ($h = 1$), but now also with considerably larger standard errors. In fact, none of the estimates in model$_{D2}$ are significantly different from zero at the 5% level. However, the log-likelihood value for model$_{D2}$ is substantially larger than for model$_C$.

For one parameter, β_0, the estimates differ much more for model$_{D1}$. This is expected, bearing figure 1 and expression (7) for the conditional probabilities of the observed returns in mind. If the boundaries are shifted downwards, that is if $H < h/2$, the conditional probability of observing r_t will alter, increasing if r_t is positive and decreasing if r_t is negative. If this change in probabilities is not supported by the data, it must be offset by a change in the conditional mean, $X_t\boldsymbol{\beta}$. Any, or all, of the $\boldsymbol{\beta}$-parameters may change, but if the dynamical relationship (β_1 and β_2) is unaffected, β_0 must change. A positive correlation between H and β_0 can thus be

expected, which can be seen in table 1. If $\widehat{H} < h/2$, $\hat{\beta}_0$ is decreased and vice versa if $\widehat{H} > h/2$.

It seems, from the data used here, as if the effects of discreteness in stock returns influence the parameter estimates and the standard errors, particularly for shorter time series and high tick sizes. It is mainly the conditional variance parameters that are affected, and may therefore have important consequences in financial areas such as volatility forecasting and portfolio optimization.

4 Diagnostics and Specification Testing

A most useful tool in regression diagnostics is the analysis of model residuals. The residuals from a correctly specified model should resemble IID random variables. In this case, the "residuals" from model (3) are, by necessity, of a discrete nature, and the residuals from models (4) and (9) are not computable, since we do not know $\tilde{r}_t$ or the latent variable r_t^*. As suggested by Gourieroux et al. (1985, 1987) we can, however, define a quantity called the generalized residual $\hat{\varepsilon}_t$, which is the best prediction of the model error given the information set, the maximum likelihood estimates, and the observation at time t. For model$_{D1}$, we have:

$$\hat{\varepsilon}_t = E[\varepsilon_t^* \mid r_t, \Psi_{t-1}; \hat{\boldsymbol{\theta}}],$$

where $\hat{\boldsymbol{\theta}} = \{\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \widehat{H}\}$ are the maximum likelihood estimates. From section 2 we know that with Gaussian error terms, $\hat{\varepsilon}_t$ is the expectation of a doubly truncated normally distributed variable, yielding:

$$\begin{aligned} \hat{\varepsilon}_t &= \hat{\sigma}_t \frac{\varphi(\hat{c}_{1t}) - \varphi(\hat{c}_{2t})}{\Phi(\hat{c}_{2t}) - \Phi(\hat{c}_{1t})}, & \hat{\sigma}_t &= \sqrt{\hat{\gamma}_0 + \hat{\gamma}_1 e_{t-1}^2 + \hat{\gamma}_2 \hat{\sigma}_{t-1}^2}, \\ \hat{c}_{1t} &= \frac{1}{\hat{\sigma}_t}\left(\widehat{\alpha}_{1t} - X_t\hat{\boldsymbol{\beta}}\right), & \hat{c}_{2t} &= \frac{1}{\hat{\sigma}_t}\left(\widehat{\alpha}_{2t} - X_t\hat{\boldsymbol{\beta}}\right), \end{aligned}$$

where $\varphi(\cdot)$ is the standard normal probability density function, and $e_t = r_t - X_t\hat{\boldsymbol{\beta}}$ are the observed residuals.

Unfortunately, the generalized residuals cannot normally be used as ordinary residuals in different diagnostic tests, see Gourieroux et al. (1985). Valid specification tests, often used in binary choice models, are Likelihood ratio tests, Wald tests, and Lagrange multiplier or score tests. The last tests have the advantage of being carried out by using the estimates from the restricted model only, which may sometimes save the computational effort, and is crucial in testing for lagged unobservable variables. Gourieroux et al. (1987) show how the score test can be expressed in terms of the generalized residuals, and present numerous examples of its use.

It is of particular interest to test for serial correlation from lagged endogenous variables. Suppose that the mean process in (4) is given by

$$r_t^* = \delta r_{t-j}^* + X_t\boldsymbol{\beta} + \varepsilon_t^*, \qquad j = 1, 2, ..., \mid \delta \mid < 1.$$

The logarithm of the conditional density of the latent variable is

$$\ln l^* \left(r_t^* | \Psi_{t-1}; \boldsymbol{\theta}\right) = -\frac{1}{2} \ln \left(\sigma_t^2\right) - \frac{1}{2\sigma_t^2} \left(r_t^* - \delta r_{t-j}^* - X_t\boldsymbol{\beta}\right)^2.$$

Under the null hypothesis that $\delta = 0$, the latent score increment is

$$\hat{S}_t^* = \left. \frac{\partial \ln l^* \left(r_t^* | \Psi_{t-1}; \boldsymbol{\theta}\right)}{\partial \delta} \right|_{\delta=0, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{r_{t-j}^* \varepsilon_t^*}{\hat{\sigma}_t^2}. \tag{15}$$

Amongst others, Louis (1982) and Gourieroux et al. (1985) show that the observable score is the expectation of the latent score, given the observable variables. In Appendix A, I show how this also implies that the observable score increment is the expectation of (15), given the observables:

$$\hat{S}_t = E\left[\hat{S}_t^* | \left\{r_u\right\}_{u=1}^T\right] = E\left[\hat{S}_t^* | \left\{r_u\right\}_{u=1}^t\right] = \frac{\hat{r}_{t-j}\hat{\varepsilon}_t}{\hat{\sigma}_t^2},$$

where $\hat{r}_t = E[r_t^* \mid r_t, \Psi_{t-1}; \hat{\boldsymbol{\theta}}] = X_t\boldsymbol{\beta} + \hat{\varepsilon}_t$, since r_{t-j}^* and ε_t^* are independent, given the observable data. Since the score increments are martingale increments and hence uncorrelated, the variance of the score is $Var\left[\sum_{t=j+1}^T \hat{S}_t\right] = \sum_{t=j+1}^T E\left[\hat{S}_t^2\right]$. A computable score statistic is then:

$$\hat{\xi}_j = \frac{\left(\sum_{t=j+1}^T \hat{S}_t\right)^2}{\sum_{t=j+1}^T \hat{S}_t^2} = \frac{\left(\sum_{t=j+1}^T \hat{r}_{t-j}\hat{\varepsilon}_t/\hat{\sigma}_t^2\right)^2}{\sum_{t=j+1}^T \left(\hat{r}_{t-j}\hat{\varepsilon}_t/\hat{\sigma}_t^2\right)^2}, \tag{16}$$

which is asymptotically $\chi^2(1)$ under the null hypothesis[7].

We can go one step further and test for the presence of lagged squared error terms in the variance process in (4), which may be assumed to be given by

$$\sigma_t^2 = \delta \varepsilon_{t-k}^{*2} + \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 \sigma_{t-1}^2, \qquad k = 1, 2, ...$$

Under the null hypothesis that $\delta = 0$, the latent score increment is now

$$\hat{S}_t^* = \left. \left(-\frac{1}{2}\frac{\partial \sigma_t^2/\partial \delta}{\sigma_t^2} + \frac{1}{2}\frac{\partial \sigma_t^2/\partial \delta}{\sigma_t^4}\varepsilon_t^{*2}\right) \right|_{\delta=0, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

[7] In a somewhat similar framework, also with a time-varying conditional variance, Hausman et al. (1992) derive this score statistic. In their expression, they do not divide by $\hat{\sigma}_t^2$, which is odd, since $\hat{\sigma}_t^2$ only vanishes if it is a constant.

Since $\partial\sigma_t^2/\partial\delta = \varepsilon_{t-k}^{*2} + \gamma_2 \partial\sigma_{t-1}^2/\partial\delta = \sum_{i=0}^{t-k-1} \gamma_2^i \varepsilon_{t-k-i}^{*2}$, it follows that

$$\hat{S}_t^* = \frac{1}{2\hat{\sigma}_t^2}\left(\frac{\varepsilon_t^{*2}}{\hat{\sigma}_t^2} - 1\right) \sum_{i=0}^{t-k-1} \hat{\gamma}_2^i \varepsilon_{t-k-i}^{*2}. \tag{17}$$

Given the observables, ε_t^{*2} and ε_{t-k-i}^{*2} are independent, so what remains is an expression for the conditional expectation of ε_t^{*2}, which from (14) is given by

$$\widehat{\varepsilon_t^2} = E\left[\varepsilon_t^{*2} \mid r_t, \Psi_{t-1}; \hat{\boldsymbol{\theta}}\right] = \hat{\sigma}_t^2\left(1 + \frac{\hat{c}_{1t}\varphi\left(\hat{c}_{1t}\right) - \hat{c}_{2t}\varphi\left(\hat{c}_{2t}\right)}{\Phi(\hat{c}_{2t}) - \Phi(\hat{c}_{1t})}\right).$$

Since $\hat{S}_t = E\left[\hat{S}_t^* | \left\{r_u\right\}_{u=1}^t\right]$, taking the expectation of (17), the score statistic is

$$\hat{\eta}_k = \frac{\left(\sum_{t=k+1}^T \hat{S}_t\right)^2}{\sum_{t=k+1}^T \hat{S}_t^2} \tag{18}$$

with

$$\hat{S}_t = \frac{1}{2\hat{\sigma}_t^2} \frac{c_{1t}\phi\left(c_{1t}\right) - c_{2t}\phi\left(c_{2t}\right)}{\Phi(c_{2t}) - \Phi(c_{1t})} \sum_{i=0}^{t-k-1} \hat{\gamma}_2^i \widehat{\varepsilon_{t-k-i}^2}. \tag{19}$$

These score statistics are also asymptotically $\chi^2(1)$ under the null hypothesis that $\delta = 0$. Using the generalized residuals in this way, we can test our model against the model we would have if actually observing $\{r_s^*\}_1^T$, although not jointly. If we could test jointly the significance of r_{t-1}^*, r_{t-2}^*, and ε_{t-1}^{*2} in model$_{D1}$, we would be able to say if model$_{D1}$ is different from the model in (1).

In table 2, I report the score statistics $\hat{\xi}_j$ for $j = 1, 2, ..., 5$, and $\hat{\eta}_k$ for $k = 1, 2, 3$ together with the corresponding p-values for model$_{D1}$. Since I have included two lags, $\hat{\xi}_1$ and $\hat{\xi}_2$ are expected to be small. Only $\hat{\xi}_5$ for AC are significant at the 5% level. Including this lag in the mean process might improve the specification for AC. None of the η-statistics are significant at the 5% significance level. The largest is $\hat{\eta}_1$ for SKF with a p-value of 0.21.

I have been assuming Gaussian error terms, but see no reason why any suitable distribution could not be used. The moments of the truncated distribution may then not be explicitly given, and must be computed numerically.

Basically, the same steps can be taken to construct score tests for model$_{D2}$. In this case the lagged endogenous variables, $\delta\tilde{r}_{t-j}$ or $\delta\tilde{\varepsilon}_{t-k}^2$, enter all conditional expectations. Because of the recursive structure of the model, both in the conditional mean and the conditional variance processes, the calculations will be practically impossible to perform. A compromise would be a model like (9), but where the conditional mean process is given by

$$r_t^* = \beta_0 + \beta_1 r_{t-1} + \beta_2 r_{t-2} + \varepsilon_t^*.$$

Table 2: Score test statistics $\hat{\xi}_j$ for $j = 1, ..., 5$ as defined in (16), and $\hat{\eta}_k$ for $k = 1, 2, 3$ as defined in (18) and (19), for model$_{D1}$. Small numbers are asymptotic p-values.

Stocks	$\hat{\xi}_1$	$\hat{\xi}_2$	$\hat{\xi}_3$	$\hat{\xi}_4$	$\hat{\xi}_5$	$\hat{\eta}_1$	$\hat{\eta}_2$	$\hat{\eta}_3$
ABB	0.100 0.752	0.331 0.565	0.451 0.502	0.046 0.831	0.098 0.754	0.004 0.949	0.193 0.660	0.272 0.602
SKF	0.003 0.959	0.003 0.956	0.065 0.799	0.436 0.509	0.001 0.971	1.560 0.211	0.188 0.665	0.275 0.600
AC	0.057 0.811	0.002 0.966	2.465 0.116	0.548 0.459	4.017 0.045	0.048 0.827	0.000 0.993	0.004 0.947
TRE	0.156 0.693	0.605 0.437	0.305 0.581	2.980 0.084	0.149 0.700	0.081 0.776	0.268 0.604	0.082 0.775
TREr	0.214 0.644	0.274 0.600	0.231 0.631	3.301 0.069	0.530 0.467	0.239 0.625	0.428 0.513	0.321 0.571

The calculations would still be difficult, but not overwhelmingly so.

5 Summary and Conclusions

This paper presents two extensions of the GARCH-framework in order to handle situations where the dependent variable is conditionally discrete, such as would be the case if one wished to model returns from individual stock prices. Using price data from some Swedish stocks, I find that the parameter estimates and the asymptotic standard errors sometimes differ when comparing the modified models to the classical, but misspecified, continuous-state GARCH model. In particular, this is the case for the conditional variance parameters, and when the models are fitted to shorter price series, or series with a tick size to price ratio more similar to those found in low priced stocks. This may have important real effects in areas where the specification of the second moment is of significance, for example in volatility forecasting and portfolio optimization. By using simulated annealing in my estimations, I have reduced the possibility of differences of the models being due to the existence of local maxima.

Caution must be taken in diagnostic and specification testing when the dependent variable is latent. The observed residuals are not continuous, and may distort diagnostic tests applied to these residuals. If the tick size is large, the null distribution of correlation integral based statistics, such as the BDS test (Brock et al., 1996), will certainly be biased, as documented by Amilon and Byström (1999). Following the lines suggested by Gourieroux et al. (1985, 1987), and applied by Hausman et al. (1992), I construct generalized residuals which can be used in score tests for omitted lagged endogenous variables, for example. Unfortunately, only for one of the models, the calculations are practically feasible.

To sum up, the cost of using the extended models and a valid diagnostic test procedure instead of ignoring discreteness is small, and consists, in principle, of collecting price data instead of return data, and keeping dates of stock splits in mind. To me, this seems to be a reasonable

sacrifice.

References

Amilon, H., and Byström, H. (1999), "The Compass Rose Pattern of the Stock Market: How Does it Affect Parameter Estimates, Forecasts, and Statistical Tests?", Working Paper, Lund University, Lund.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics*, 31, 307-327.

Bollerslev, T., Chou, R., and Kroner, K. (1992), "ARCH Modeling in Finance", *Journal of Econometrics*, 52, 5-59.

Brock, W., Dechert, W., Scheinkmann, J., and LeBaron, B. (1996), "A Test for Independence Based on the Correlation Dimension", *Econometric Reviews*, 15(3), 197-235.

Crack, T., and Ledoit, O. (1996), "Robust Structure Without Predictability: The "Compass Rose" Pattern of the Stock Market", *Journal of Finance*, June, 751-762.

Dempster, A., Laird, N., and Rubin, D. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society,* Ser. B, 39, 1-22.

Doornik, J., and Ooms, M. (2000), "Multimodality in the GARCH Likelihood", Working Paper, Nuffield College, Oxford.

Glosten, L., Jagannathan, R., and Runkle, D. (1993), "Relationship Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *Journal of Finance*, 48, 1779-1801.

Goffe, W., Ferrier, G., and Rogers, J. (1994), "Global Optimization of Statistical Functions with Simulated Annealing", *Journal of Econometrics*, 60, 65-99.

Gourieroux, C., Monfort, A., Renault, E., and Trognon, A. (1987), "Generalised Residuals", *Journal of Econometrics*, 34, 5-32.

Gourieroux, C., Monfort, A., and Trognon, A. (1985), "A General Approach to Serial Correlation", *Econometric Theory*, 1, 315-340.

Hausman, J., Lo, A., and MacKinlay, A. (1992), "An Ordered Probit Analysis of Transaction Stock Prices", *Journal of Financial Economics*, 31, 319-379.

Johnson, N., and Kotz, S. (1970), *Distributions in statistics: continuous univariate distributions-1*, Houghton Mifflin, Boston.

Louis, T. (1982), "Finding the Observed Information Matrix When Using the EM Algorithm", *Journal of the Royal Statistical Society,* Ser. B, 44, 226-233.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953), "Equations of State Calculations by Fast Computing Machines", *Journal of Chemical Physics*, 21, 1087-1092.

Morgan, I., and Trevor, R. (1999), "Limit Moves as Censored Observations of Equilibrium Futures Price in GARCH Processes", *Journal of Business and Economic Statistics*, 17(4), 397-408.

Press, W., Flannery, B., Teukolsky, S., and Vetterling, W. (1992), *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, New York, NY.

Rydberg, T., and Shephard, N. (1998), "Dynamics of Trade-by-Trade Price Movements: Decomposition and Models", Working Paper, Nuffield College, Oxford.

Shephard, N. (1994), "Autoregressive Based Generalized Linear Models", Unpublished Paper, Nuffield College, Oxford.

Szpiro, G. (1998), "Tick Size, the Compass Rose and Market Nanostructure", *Journal of Banking and Finance*, 22, 1559-1569.

A Appendix

For convenience, let D_λ be the notation for the partial derivative with respect to a parameter λ, given the maximum likelihood estimates. The observable score increment can be written as the difference of the following observable scores:

$$\hat{S}_t = D \ln l(r_t|r_1,...,r_{t-1};\lambda) = D_\lambda \ln l(r_1,...,r_t;\lambda) - D_\lambda \ln l(r_1,...,r_{t-1};\lambda)$$

From Louis (1982) and Gourieroux et al. (1985) we know that the observable score is the expectation of the latent score, given the observables. Hence

$$\hat{S}_t = E\left[D_\lambda \ln l^*(r_1^*,...,r_t^*;\lambda)|r_1,...,r_t\right] - E\left[D_\lambda \ln l^*(r_1^*,...,r_{t-1}^*;\lambda)|r_1,...,r_{t-1}\right] = E\left[\sum_{k=1}^{t} \hat{S}_k^*|r_1,...,r_t\right] - E\left[\sum_{k=1}^{t-1} \hat{S}_k^*|r_1,...,r_{t-1}\right]$$

Because of the Markov property of $\hat{S}_k^*$, that is $E\left[\hat{S}_k^*|r_1,...,r_k\right] = E\left[\hat{S}_k^*|r_1,...,r_T\right]$ with $T \geq k$, we have

$$\hat{S}_t = \sum_{k=1}^{t} E\left[\hat{S}_k^*|r_1,...,r_k\right] - \sum_{k=1}^{t-1} E\left[\hat{S}_k^*|r_1,...,r_k\right] = E\left[\hat{S}_t^*|r_1,...,r_t\right].$$

Hence, the observable score increment is equal to the expectation of the latent score increment, given the observable data.