ARMA and ARCH models

Jack Adeney

January 20, 2026

PROOF OF CONCEPT

Take daily percentage change data, where r_t represents the percentage change at day t. Assume that there are T observations. The first step is to fit an ARMA(p,q) model in order to forecast changes in the expected value of daily returns. Next, fit an ARCH(d) model to forecast conditional heteroskedasticity in residual variance. Step 3, use this information to calculate a probability of the daily return being positive. Finally, optimize bounds on probabilities to maximize profit.

Step 1.

An ARMA(p,q) model is defined as follows:

$$r_t = \epsilon_t + \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} - \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

In other words, today's return is a linear combination of past returns and past 'exogenous shocks'. In other words, information that is not contained in the return data. We assume that $\epsilon_t \sim N(0, \sigma_t^2)$, meaning that we can use maximum likelihood in order to solve for ϕ and θ parameters. The log likelihood function is given as follows:

$$l(\phi, \theta; r) = -\frac{T}{2} \ln(2\pi\sigma^2) + \frac{1}{2} \sum_{i=1}^{T} \frac{\epsilon_i^2}{\sigma^2}$$

Where

$$\epsilon_t = r_t - \phi_0 - \sum_{j=1}^{p} \phi_j r_{t-j} - \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

Clearly, ϵ_t must be calculated recursively in order to be identified. Set $\epsilon_0 = 0$ and continue from there, including all lags up to $t - p$ for AR arguments and $t - q$ for MA arguments where possible.

This maximum likelihood estimation will provide $\hat{\phi}$ and $\hat{\theta}$ values for the specified model. We can now calculate values of $\hat{\epsilon}_t$ and $\hat{r}_t$, where $\hat{r}_t = \hat{\phi}_0 + \sum_{t=1}^{p} \hat{\phi}_i r_{t-i} + \sum_{i=1}^{q} \hat{\theta}_i (r_t - \hat{r}_t)$.

Next, use the ARCH(d) model to forecast σ_t^2, which is now allowed to vary with time. The formula for this is

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{d} \alpha_i a_{t-i}^2$$

This can again be solved using maximum likelihood by assuming that $a_t = \sigma_t \epsilon_t$ and $\epsilon_t \sim N(0,1)$ for all t.

We now have estimated returns, $\hat{y}_t$, model residuals ϵ_t, and estimated shock variances σ_t^2 for every period.

Strategy: suppose that we only ever buy. When we do buy, we buy at open and sell at close. This is obviously a much more simplistic strategy than we may otherwise support, but it makes the analysis easier.

So, for each time period t, find $\Phi_{\sigma_t^2}(\hat{y}_t)$ where $\Phi_{\sigma_t^2}$ is the normal CDF with mean 0 and variance σ_t^2. This means that we have a probability of each return being larger than 0 (i.e profitable to invest). Now we have these probabilities for each period t.

It makes sense that the closer that probability is to 0.5, the more we should be wary of the models predictive accuracy. As a result, we are going to define bounds $b \geq 0.5$, such that we only invest on days where the probability of positive return is larger than b, and do nothing otherwise.

As b increases, we become more selective about when we trade but also make less trades. If b were 0.5 on the other hand, then we would be buying on any day that $\hat{r}_t > 0$. The strategy therefore is buy on any day where $\Phi_{\sigma_t^2}(\hat{r}_t) > b^*$ for some optimized b^*, and do nothing otherwise.