Формулировка GARCH-процесса

Обозначим через σ_t^2 условную дисперсию ε_t по прошлой информации Ω_t:

$$\mathrm{Var}(\varepsilon_t \,|\, \Omega_t) = \sigma_t^2 .$$

В данном случае Ω_t — прошлая история процесса ε_t:

$$\Omega_t = \{\varepsilon_{t-j}\}_{j=1}^{\infty} .$$

Если процесс ε_t является GARCH(k, l), то он задается следующим рекуррентным соотношением для условной дисперсии ε_t:

$$\sigma_t^2 = \mu + \sum_{j=1}^{k} \gamma_j \varepsilon_{t-j}^2 + \sum_{j=1}^{l} \delta_j \sigma_{t-j}^2 . \qquad (1)$$

Предполагается, что

$$\mathrm{E}(\varepsilon_t \,|\, \Omega_t) = 0,$$

а также что нормированные величины ε_t независимы и одинаково распределены:

$$\xi_t = \frac{\varepsilon_t}{\sigma_t} \sim \mathrm{IID}.$$

Регрессия с GARCH-процессом в ошибке

Регрессия с GARCH-процессом в ошибке имеет вид

$$y_t = X_t \boldsymbol{\beta} + \varepsilon_t, \quad t = 1, ..., T.$$

Здесь y_t — зависимая переменная, X_t — вектор-строка ($1\times m$) независимых переменных, $\boldsymbol{\beta}$ — вектор-столбец ($m\times 1$) коэффициентов регрессии, ε_t — ошибка, представляющая собой GARCH-процесс.

В матричной записи

$$\boldsymbol{y} = \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Функция правдоподобия

Предположение об условной нормальности ε_t:

$$\varepsilon_t \mid \sigma_t^2 \sim \mathrm{N}(0, \sigma_t^2).$$

В предположении условной нормальности вклад в логарифмическую функцию правдоподобия t-го наблюдения равен

$$\ell_t = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma_t^2 - \frac{1}{2}\frac{\varepsilon_t^2}{\sigma_t^2}, \quad t = 1, ..., T.$$

Здесь ε_t и σ_t^2 следует рассматривать как функции от неизвестных параметров $\boldsymbol{\theta}$:

$$\varepsilon_t = \varepsilon_t(\boldsymbol{\theta}) \quad \text{и} \quad \sigma_t^2 = \sigma_t^2(\boldsymbol{\theta}) .$$

В вектор $\boldsymbol{\theta}$ должны войти параметры $\boldsymbol{\beta}$, μ, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$ и любые другие неизвестные параметры (см. ниже).

Логарифмическая функция правдоподобия равна

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\left(T \ln 2\pi + \sum_{t=1}^{T} \left[\ln \sigma_t^2(\boldsymbol{\theta}) + \frac{\varepsilon_t^2(\boldsymbol{\theta})}{\sigma_t^2(\boldsymbol{\theta})} \right] \right).$$

Оценки максимального правдоподобия по определению должны максимизировать логарифмическую функцию правдоподобия ℓ по неизвестным параметрам $\boldsymbol{\theta}$.

Оценки максимального правдоподобия являются корнями уравнения правдоподобия:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \mathbf{0}.$$

Начальные значения условной дисперсии

Для того, чтобы использовать рекуррентную формулу (1) для расчета условной дисперсии, используемой в функции правдоподобия, нужны начальные значения σ_t^2, $t = \min\{1, k+l-1\}$, ..., k. Один из способов — рассматривать эти начальные значения как параметры, которые требуется оценить. Обозначим эти параметры через η_t (= σ_t^2), тогда рекуррентная формула примет вид:

$$\sigma_t^2(\boldsymbol{\theta}) = \begin{cases} \eta_t, & t = \min\{1, k+1-l\}, ..., k \\ \mu + \sum_{j=1}^{k} \gamma_j \varepsilon_{t-j}^2(\boldsymbol{\theta}) + \sum_{j=1}^{l} \delta_j \sigma_{t-j}^2(\boldsymbol{\theta}), & t = k+1, ..., T. \end{cases} \tag{2}$$

Можно также использовать вместо начальных значений безусловную дисперсию:

$$\sigma_t^2 = \sigma^2,\ t = \min\{1, k+l-1\}, ..., k,$$

где

$$\sigma^2 = \frac{\mu}{1 - \sum_{j=1}^{k} \gamma_j - \sum_{j=1}^{l} \delta_j}.$$

При этом предполагается, что

$$1 - \sum_{j=1}^{k} \gamma_j - \sum_{j=1}^{l} \delta_j > 0,$$

то есть, что GARCH-процесс стационарен.

В дальнейшем рассмотрим только первый способ (с параметрами η_t). При этом в вектор $\boldsymbol{\theta}$ следует включить и η_t.

Если, GARCH-процесс стационарен, то с асимптотической точки зрения выбор начальных приближений не играет роли.

Метод оценивания

Для оценивания модели можно применить следующий общий итерационный метод:

$$\boldsymbol{\theta}^{r+1} = \boldsymbol{\theta}^r + \left(\hat{\Im}(\boldsymbol{\theta}^r)\right)^{-1} \mathbf{g}(\boldsymbol{\theta}^r), \tag{3}$$

где

$$\mathbf{g}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \text{ —}$$

градиент логарифмической функции правдоподобия (вектор-столбец), $\hat{\Im}(\boldsymbol{\theta})$ — оценка информационной матрицы, r — номер итерации.

Стационарная точка итераций ($\boldsymbol{\theta}^{r+1} = \boldsymbol{\theta}^r$) соответствует решению уравнения правдоподобия:

$$\mathbf{g}(\boldsymbol{\theta}^r) = \mathbf{0}.$$

Алгоритм (3) на практике применяют обычно в несколько модифицированном виде:

$$\boldsymbol{\theta}^{r+1} = \boldsymbol{\theta}^r + \lambda\left(\hat{\Im}(\boldsymbol{\theta}^r)\right)^{-1}\mathbf{g}(\boldsymbol{\theta}^r),$$

где $\lambda > 0$. Такая модификация нужна для повышения устойчивости алгоритма. Например, при выборе $\lambda = 1$ функция правдоподобия может уменьшится или шаг может попасть в недопустимую область (в GARCH-модели это случай, когда одна из оценок условной дисперсии становится отрицательной). В таких случаях множитель λ уменьшают до тех пор, пока не достигнут требуемого — увеличения функции правдоподобия или попадания в допустимую область.

Информационную матрицу можно оценить разными методами. Это, например, может быть матрица Гессе логарифмической функции правдоподобия со знаком минус:

$$\hat{\Im}(\boldsymbol{\theta}) = -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}.$$

при этом итерации (3) соответствуют классическому методу Ньютона. К сожалению, в модели GARCH матрицу Гессе довольно трудно вычислять.

Другой метод (его применяли те, кто предложил модель GARCH — Энгл, Боллерслев) называется методом BHHH (или методом внешнего произведения градиентов, OPG). В этом методе информационная матрица оценивается следующим способом:

$$\hat{\Im}(\boldsymbol{\theta}) = \mathbf{G}(\boldsymbol{\theta})^{\mathrm{T}}\mathbf{G}(\boldsymbol{\theta}),$$

где $\mathbf{G}$ — матрица вкладов в градиент отдельных наблюдений:

$$\mathbf{G}(\boldsymbol{\theta}) = \{G_t(\boldsymbol{\theta})\}_{t=1,\dots,T},$$

$$G_t(\boldsymbol{\theta}) = \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}.$$

Однако, по-видимому, наилучшим методом для оценивания GARCH является следующий. В нем оценка информационной матрицы получается по формуле:

$$\hat{\Im}(\boldsymbol{\theta}) = \sum_{t=1}^{T}\hat{\Im}_t(\boldsymbol{\theta}),$$

где $\hat{\Im}_t(\boldsymbol{\theta})$ — информационная матрица для t-го наблюдения *условная* по прошлой информации Ω_t:

$$\hat{\Im}_t(\boldsymbol{\theta}) = \mathsf{E}\left(\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \mid \Omega_t\right) = \mathsf{E}(G_t(\boldsymbol{\theta})^{\mathrm{T}}G_t(\boldsymbol{\theta}) \mid \Omega_t).$$

Для применения метода нужны формулы для расчета условной информационной матрицы t-го наблюдения, $\hat{\Im}_t(\boldsymbol{\theta})$, и для расчета градиента $\mathbf{g}(\boldsymbol{\theta})$. Градиент удобно разложить на вклады отдельных наблюдений:

$$\mathbf{g}(\boldsymbol{\theta})^{\mathrm{T}} = \mathbf{1}^{\mathrm{T}}\mathbf{G} = \sum_{t=1}^{T}G_t(\boldsymbol{\theta}).$$

Таким образом, требуется вычислить вклады в градиент, $G_t(\boldsymbol{\theta})$.

Вычисление условной информационной матрицы *t*-го наблюдения и вклада в градиент *t*-го наблюдения

Продифференцируем $\ell_t(\boldsymbol{\theta})$, чтобы получить формулу для вклада в градиент t-го наблюдения:

$$G_t = \frac{\partial \ell_t}{\partial \boldsymbol{\theta}} = -\frac{1}{2}\frac{1}{\sigma_t^2}\frac{d\sigma_t^2}{d\boldsymbol{\theta}} + \frac{1}{2}\frac{\varepsilon_t^2}{\sigma_t^2}\frac{1}{\sigma_t^2}\frac{d\sigma_t^2}{d\boldsymbol{\theta}} - \frac{\varepsilon_t}{\sigma_t}\frac{1}{\sigma_t}\frac{d\varepsilon_t}{d\boldsymbol{\theta}}.$$

Здесь и в дальнейшем мы будем опускать аргумент $\boldsymbol{\theta}$.

Введем обозначения

$$\xi_t = \frac{\varepsilon_t}{\sigma_t},\ \ v_t = \xi_t^2 - 1,\ \ Q_t = \frac{1}{\sigma_t}\frac{d\varepsilon_t}{d\boldsymbol{\theta}},\ \ S_t = \frac{1}{\sigma_t^2}\frac{d\sigma_t^2}{d\boldsymbol{\theta}}.$$

В этих обозначениях вклад в градиент t-го наблюдения равен

$$G_t = \xi_t Q_t + \frac{1}{2}v_t S_t.$$

Поскольку условно по прошлой информации Ω_t вектора Q_t и S_t являются детерминированными, то

$$\hat{\Im}_t = \mathsf{E}(G_t^{\mathrm{T}} G_t \mid \Omega_t) = \mathsf{E}\left[\left(\xi_t Q_t + \frac{1}{2}v_t S_t\right)^{\mathrm{T}}\left(\xi_t Q_t + \frac{1}{2}v_t S_t\right) \mid \Omega_t\right].$$

Несложно показать, что

$$\mathsf{E}(\xi_t^2) = 1,\ \mathsf{E}(\xi_t v_t) = 0,\ \mathsf{E}(v_t^2) = 2.$$

Отсюда

$$\hat{\Im}_t = Q_t^{\mathrm{T}} Q_t + \frac{1}{2}S_t^{\mathrm{T}} S_t.$$

Ниже мы подробно разберем как вычислять градиенты

$$\frac{d\sigma_t^2}{d\boldsymbol{\theta}} \text{ и } \frac{d\varepsilon_t}{d\boldsymbol{\theta}},$$

которые нужны для вычисления Q_t и S_t.

Искусственная регрессия

Эти преобразования показывают, что r-й шаг итерационного алгоритма (4) имеет в данном случае вид

$$\boldsymbol{\Delta}^r = \boldsymbol{\theta}^{r+1} - \boldsymbol{\theta}^r = \left(\mathbf{Q}^{r\mathrm{T}}\mathbf{Q}^r + \frac{1}{2}\mathbf{S}^{r\mathrm{T}}\mathbf{S}^r\right)^{-1}\left(\mathbf{Q}^{r\mathrm{T}}\boldsymbol{\xi}^r + \frac{1}{2}\mathbf{S}^{r\mathrm{T}}\mathbf{v}^r\right),$$

где $\mathbf{Q}$, $\mathbf{S}$, $\boldsymbol{\xi}$, $\mathbf{v}$ — матрицы и вектора, составленные из Q_t, S_t, ξ_t и v_t соответственно. Шаг этого алгоритма, как несложно увидеть, можно вычислить с помощью следующей регрессии:

$$\begin{pmatrix} \boldsymbol{\xi} \\ \frac{1}{\sqrt{2}}\mathbf{v} \end{pmatrix} = \begin{bmatrix} \mathbf{Q} \\ \frac{1}{\sqrt{2}}\mathbf{S} \end{bmatrix}\boldsymbol{\Delta} + \textit{остатки}.$$

Заметим также, что оценка ковариационной матрицы в этой регрессии, полученная на последней итерации (когда метод сошелся), является состоятельной оценкой ковариационной матрицы оценок максимального правдоподобия в GARCH-модели:

$$\hat{\mathrm{Var}}(\hat{\boldsymbol\theta}) = \frac{\boldsymbol\xi^{\mathrm T}\boldsymbol\xi + \dfrac{\mathbf v^{\mathrm T}\mathbf v}{2}}{T}\left(\mathbf Q^{\mathrm T}\mathbf Q + \frac{\mathbf S^{\mathrm T}\mathbf S}{2}\right)^{-1}.$$

Таким образом, эта регрессия представляет собой то, что называется *искусственной регрессией*.

В качестве критерия остановки можно взять нецентральный коэффициент детерминации из искусственной регрессии:

$$R^2_{nc} = 1 - \frac{\textit{сумма квадратов остатков}}{\textit{сумма квадратов зависимой переменной}}.$$

Можно считать, что метод сошелся, если регрессия практически не объясняет зависимую переменную, т.е. если величина R^2_{nc} мала ($R^2_{nc} < \phi$). На практике неплохо работает правило остановки с $\phi = 10^{-11}$.

Вычисление градиентов остатков

Для полного описания алгоритма требуется указать способ вычисления градиентов $d\sigma_t^2/d\boldsymbol\theta$ и $d\varepsilon_t/d\boldsymbol\theta$.

Поскольку по определению

$$\varepsilon_t(\boldsymbol\theta) = y_t - X_t\boldsymbol\beta,$$

то

$$\frac{d\varepsilon_t}{d\boldsymbol\theta} = \left(X_t, \mathbf 0^{\mathrm T}\right).$$

Если регрессоры отсутствуют, то есть, если y_t — это непосредственно GARCH-процесс ε_t с нулевым математическим ожиданием, то градиент равен нулю и, соответственно $Q_t = \mathbf 0^{\mathrm T}$. При этом искусственная регрессия упрощается до регрессии $1/\sqrt{2}\,\mathbf v$ на $1/\sqrt{2}\,\mathbf S$ (множитель $1/\sqrt{2}$ не влияет на оценки регрессии, но должен входить в ковариационную матрицу, поэтому его лучше не отбрасывать).

Вычисление градиентов условных дисперсий

Обозначим

$$\frac{d\sigma_t^2}{d\boldsymbol\theta} = H_t.$$

Формулу для этих величин получим, дифференцируя по $\boldsymbol\theta$ соотношение (2).

При $t = \min\{1, k+l-1\}, \ldots, k$ вектор-строка H_t, будет состоять из нулей, только на месте параметра η_t будет стоять 1. При $t > k$ вектор H_t вычисляется рекуррентно:

$$H_t = \frac{\partial\sigma_t^2}{\partial\boldsymbol\theta} + 2\sum_{j=1}^{k}\gamma_t\varepsilon_{t-j}\frac{d\varepsilon_{t-j}}{d\boldsymbol\theta} + \sum_{j=1}^{l}\delta_t H_{t-j},\ \ t = k+1, \ldots, T.$$

Здесь $\dfrac{\partial\sigma_t^2}{\partial\boldsymbol\theta}$ — непосредственная производная условной дисперсии по параметрам:

$$\begin{array}{cccccc} & \boldsymbol\beta & \mu & \boldsymbol\gamma & \boldsymbol\delta & \boldsymbol\eta \\ \dfrac{\partial\sigma_t^2}{\partial\boldsymbol\theta} = & \Big(\mathbf 0^{\mathrm T}, & 1, & \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-p}^2, & \sigma_{t-1}^2, \ldots, \sigma_{t-q}^2, & \mathbf 0^{\mathrm T}\Big). \end{array}$$

Сверху в формуле подписаны параметры, по которым берется производная.

Начальные приближения

В алгоритме большую роль играет выбор начальных приближений параметров $\boldsymbol{\theta} = \boldsymbol{\theta}^0$. Неправильный выбор может вызвать переполнение при вычислениях. Можно предложить следующие начальные приближения:

Для вектора коэффициентов регрессии $\boldsymbol{\beta}$ можно использовать оценки обычного метода наименьших квадратов:

$$\boldsymbol{\beta}^0 = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}.$$

Затем на основе остатков из регрессии можно вычислить оценку безусловной дисперсии:

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2.$$

Параметры γ_j^0 и δ_j^0 выбираются положительными и такими, что

$$1 - \sum_{j=1}^{k}\gamma_j^0 - \sum_{j=1}^{l}\delta_j^0 > 0.$$

Можно взять, например,

$$\gamma_j^0 = \frac{1}{4k}, \quad \delta_j^0 = \frac{1}{4l} \quad (\text{при } l > 0).$$

Другой вариант — выбирать γ_j^0 и δ_j^0 случайно.

Затем на основе оценки безусловной дисперсии и коэффициентов γ_j и δ_j вычисляется начальное приближение для μ:

$$\mu^0 = \hat{\sigma}^2\left(1 - \sum_{j=1}^{k}\gamma_j^0 - \sum_{j=1}^{l}\delta_j^0\right).$$

В качестве начальных приближений параметров η_t можно взять оценку безусловной дисперсии:

$$\eta_t^0 = \hat{\sigma}^2.$$

Алгоритм вычислений

Сначала вычисляются начальные приближения параметров. Следует следить, чтобы они оказались в допустимой области, т.е. вычисленные на их основе условные дисперсии все были положительными.

После этого идут итерации:

- Вычислить остатки $\varepsilon_t(\boldsymbol{\theta}) = y_t - X_t\boldsymbol{\beta}$.
- Вычислить условные дисперсии σ_t^2 по рекуррентной формуле (2).
- Вычислить градиенты остатков $d\varepsilon_t/d\boldsymbol{\theta}$.
- Вычислить градиенты условных дисперсий H_t.
- Вычислить $\mathbf{Q}$, $\mathbf{S}$, $\boldsymbol{\xi}$ и $\mathbf{v}$.
- Оценить искусственную регрессию и получить направление изменения параметров $\boldsymbol{\Delta}$.
- Проверить выполнение правила остановки ($R_{nc}^2 < \phi$).
- Выбрать коэффициент λ так, чтобы новое значение параметров $\boldsymbol{\theta}^{r+1} = \boldsymbol{\theta}^r + \lambda\boldsymbol{\Delta}^r$ оказалось в допустимой области и давало больший уровень функции правдоподобия, чем $\boldsymbol{\theta}^r$, т.е. $\ell\left(\boldsymbol{\theta}^{r+1}\right) > \ell\left(\boldsymbol{\theta}^r\right)$.
- Начать новую итерацию.

После того, как итеративный алгоритм сойдется, на основе ковариационной матрицы из искусственной регрессии вычисляются стандартные ошибки параметров и t-статистики по обычным формулам.