Конспект лекций

Финансовая эконометрика

Станкевич Иван Павлович

Аннотация

Курс финансовой эконометрики знакомит с основными классами моделей, использующихся в финансах. Предполагается, что слушатели знакомы с эконометрикой, но ещё незнакомы с моделями временных рядов. Курс начинается с основных концепций анализа временных рядов – сюда входят базовые одномерные модели (ARIMA и ETS), коинтеграция, базовые многомерные модели. В дальнейшем рассматриваются специфичные для финансовой сферы вопросы, такие как CAPM-модель, моделирование волатильности, нелинейные модели временных рядов.

Содержание

1 Основные понятия теории временных рядов

2 Классический подход к теории временных рядов

2.1 Автокорреляционная функция и частная автокореляционная функция

Важный инструмент исследования временного ряда.

Определение 1. *Пусть Y_t – стационарный в широком смысле процесс. Тогда ее автокореляционная функция определяется как*

$$\rho_k = \frac{\operatorname{cov}(Y_t, Y_{t-k})}{\operatorname{var} Y_t}.$$

Уравнения Юла-Уокера Рассмотрим стационарное решение процесса $AR(1)$:

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad |\alpha| < 1.$$

$$\mathbb{E}\left[y_t y_{t-1}\right] = \mathbb{E}\left[(\alpha y_{t-1} + \varepsilon_t) y_{t-1}\right],$$

откуда

$$\begin{aligned}
\gamma_0 &= \alpha\gamma_1 + \sigma_\varepsilon^2, \\
\gamma_1 &= \alpha\gamma_0,
\end{aligned}$$

где

$$\gamma_k = \operatorname{cov}(y_t, y_{t-k}).$$

Итого,

$$\begin{aligned}
\gamma_0 &= \frac{\sigma_\varepsilon^2}{1-\alpha^2}, \\
\gamma_1 &= \alpha\frac{\sigma_\varepsilon^2}{1-\alpha^2}, \\
&\dots \\
\gamma_k &= \alpha^k\frac{\sigma_\varepsilon^2}{1-\alpha^2}, \\
&\dots
\end{aligned}$$

Получившиеся уравнения называются *уравнениями Юла-Уокера*. Аналогично для стационарного решения $MA(1)$, $y_t = \varepsilon_t + \beta\varepsilon_{t-1}$:

$$\begin{aligned}
\gamma_0 &=, \\
\gamma_1 &= \beta\sigma_\varepsilon^2, \\
\gamma_2 &= 0, \\
&\dots \\
\gamma_k &= 0, \\
&\dots
\end{aligned}$$

Предложение 1. *Для стационарного решения* $MA(q)$ $ACF(k) \equiv 0$ *при* $k > q$.

Определение 2. *Пусть* Y_t *– стационарный в широком смысле процесс. Тогда её частная автокореляционная функция определяется как коэффициент* $\varphi_{k,k}$ *из регрессии*

$$y_t = \varphi_{k,1} y_{t-1} + \dots + \varphi_{k,k} y_{t-k} + u_t.$$

Пример 1 ($PACF$ для процесса $AR(1)$).

$$\begin{aligned} &\varphi_{k,1} = \alpha, \\ &\varphi_{k,2} = 0, \\ &\dots \\ &\varphi_{k,k} = 0. \end{aligned}$$

Итого, $PACF(k) = \alpha\delta_1^k$, *где* δ_i^j *– символ Кронекера.*

2.2 Методология Бокса и Дженкинса (1970)

Авторами предлагаются следующие шаги[1]:

1. Ряд остационариваем разностями ($y_t \mapsto \Delta^d y_t$, оператор первой разности действует следующим образом: $\Delta y_t = y_t - y_{t-1}$);
2. Смотрим на ACF и $PACF$ [2]:
   - $AR(p)$
   - $MA(q)$
   - $ARMA(p, q)$.
3. Оцениваем $ARIMA(p, d, q)$.

Так уже никто не делает, но это важный этап развития эконометрической мысли. Выяснилось, что не всегда можно остационаривать ряд разностями (например, ряды с экспоненциальными трендами); стационарность определяется большим количеством тестов. Более того, для подбора параметров $ARIMA$ можно использовать информационные критерии AIC/BIC: чем больше значение статистики, тем хуже модель.

2.3 Сезонность временных рядов

Сезонная модификация модели $ARIMA$ – $SARIMA(p, d, q) \times (P, D, Q)_S$ имеет следующий вид:

$$\begin{aligned} \Delta^d \Delta_S^D y_t = \alpha_1 \Delta^d \Delta_S^D y_{t-1} + \dots + \alpha_p \Delta^d \Delta_S^D y_{t-p} + \\ + \alpha_1^S \Delta^d \Delta_S^D y_{t-1} + \dots + \alpha_P^S \Delta^d \Delta_S^D y_{t-PS} + \\ + \beta_1 \Delta^d \Delta_S^D \varepsilon_{t-1} + \dots + \beta_q \Delta^d \Delta_S^D \varepsilon_{t-q} + \\ + \beta_1^S \Delta^d \Delta_S^D \varepsilon_{t-1} + \dots + \beta_Q^S \Delta^d \Delta_S^D \underbrace{\varepsilon_{t-QS}}_{\text{«}Q\text{ лет назад»}}, \end{aligned} \tag{1}$$

где $\Delta_S = (1 - L^S)$ – сезонный оператор первой разности.

[1] Первые 2 шага называются идентификацией, третий – оцениванием.
[2] TODO: расписать случаи

2.4 Тесты на единичные корни

Это семейство тестов проверяет принадлежность корней характеристического уравнения единичной окружности.

Тест Дики-Фуллера (Dickey-Fuller test) Рассмотрим АР1

$$
\begin{aligned}
&y_t = \alpha y_{t-1} + \varepsilon_t \qquad | - y_{t-1} \\
&\Delta y_t = \hat{\beta} \Delta y_{t-1}, \\
&H_0 : \beta = 0 \text{ vs } H_1 : \beta < 0, \\
&\text{Статистика } t = \frac{\hat{\beta}}{\hat{SE}(\hat{\beta})}.
\end{aligned}
$$

Расширенный тест Дики-Фуллера (Augmented DF test) Все ровно то же самое, но добавили предыдущие разности:

$$
\begin{aligned}
&\Delta y_t = ... + (\alpha - 1) y_{t-1} + \beta_1 \Delta y_{t-1} \cdots + \beta_k \Delta y_{t-k} + \varepsilon_t, \\
&H_0 : \beta = 0 \text{ vs } H_1 : \beta < 0, \\
&\text{Статистика } t = \frac{\hat{\beta}}{\hat{SE}(\hat{\beta})}.
\end{aligned}
$$

Тест КПСС (KPSS test, 1993) Пример теста на стационарность вокруг тренда (trend stationarity).

$$
\begin{aligned}
&y_t = \underbrace{C_t}_{C_t = C_{t-1} + u_t} + \underbrace{\gamma t}_{\text{Линейный тренд}} + \varepsilon_t, \quad \varepsilon_t \perp\!\!\!\perp u_t \\
&H_0 : \operatorname{var} u_t = 0 \text{ vs } H_1 : \operatorname{var} u_t > 0,
\end{aligned}
$$

Тест Льюнга-Бокса (Ljung-Box test) Вычислим по выборке $\hat{\rho}_1, \ldots, \hat{\rho}_k$. $H_0 : \hat{\rho}_1 = \cdots = \hat{\rho}_k = 0$. Статистика:

$$
Q = T(T+2) \sum_{i=1}^{k} \frac{\hat{\rho}_i^2}{n-i} \sim \chi_k^2.
$$

3 Exponential Smoothing (ETS)

3.1 Simple ES

$$
\begin{aligned}
&\hat{y}_{T+1|T} = y_T \\
&|| \\
&\alpha y_T + \alpha(1-\alpha) y_{T-1} + \ldots.
\end{aligned}
$$

3.2 Component form

Forecast equation: $\hat{y}_{T+h|T} = l_T$.
Level equation: $l_t = \alpha y_t + (1-\alpha) l_{t-1}$.

3.3 Тренд Holt (1957)

Forecast equation: $\hat{y}_{T+h|T} = l_T + hb_T$.
Level equation: $l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})$.
Trend equation: $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$.

3.4 Тренд с сезонностью Holt-Winters

Forecast equation: $\hat{y}_{T+h|T} = l_T + hb_T + S_{t+h-s(k-\cdot)}$.
Level equation: $l_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(l_{t-1} + b_{t-1})$.
Trend equation: $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$.
Seasonality equation: $S_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)S_{t-s}$.

3.5 Damped trend

Forecast equation: $\hat{y}_{T+h|T} = l_T + (\varphi + \varphi^2 + \cdots + \varphi^h)b_T + S_{t+h-s(k-\cdot)}$, $\varphi \in (0,1)$.
Level equation: $l_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(l_{t-1} + b_{t-1})$.
Trend equation: $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$.
Seasonality equation: $S_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)S_{t-s}$.

3.6 Prophet

$$y_t = \underbrace{g_t}_{\text{Trend}} + \underbrace{\underbrace{S_t}_{\text{Seasonality}}}_{\text{сезон, недельная, внутридневная}} + \underbrace{h_t}_{\text{Holidays}} + e_t$$

3.7 Кросс-валидация моделей временных рядов

$y_t, t \in \{0, \ldots, N\}$
$n \leftarrow n_0$.
while $n + h \leq N$ **do**
    Строим $\hat{y}_{n+h|n} = \hat{y}_n(y_{n-n_0}, \ldots y_n)$
    $n \leftarrow n + 1$
end while

Другой алгоритм:

$y_t, t \in \{0, \ldots, N\}$
$n \leftarrow 0$.
while $n \leq H$ **do**
    Строим $\hat{y}_{t|t-n}$
    $n \leftarrow n + 1$
end while

Метрики: MAE[3], RMSE[4], MAPE[5], MASE[6].

[3] Не чувствителен к разовым выбросам
[4] Сильно чувствителен к единичным выбросам
[5] Очищен от масштаба
[6] Относительное сравнение модели со случайным блужданием

3.8 Model selection

3.8.1 Тест Diebold-Mariano (1995)

Сравнивает прогнозную силу моделей. Пусть $g(\cdot)$ – некоторая функция потерь.

1. $e_t^A = y_t - \hat{y}_t^A$, $e_t^B = y_t - \hat{y}_t^B$;
2. $d_t = g(e_t^A) - g(e_t^B)$

$H_0 : \mathbb{E}[d_t] = 0$ vs $H_A : \mathbb{E}[d_t] \neq 0$. Статистика выглядит следующим образом:

$$\bar{d} = \frac{1}{k}\sum_t d_t, \quad \bar{f} = \sum_j \gamma_d(j),$$
$$S = \frac{\bar{d}}{\sqrt{\bar{f}/k}} \overset{H_0}{\sim} N(0,1).$$

4 Многомерные модели

Если мы рассмотрим 2 независимых ряда, стационарных вокруг разных линейных трендов (но не стационарных), то выборочная корреляция будет значимо отлична от нуля. Поэтому в этой лекции рассматриваем стационарные модели рядов. В финансовой эконометрике доходности активов $\frac{y_t - y_{t-1}}{y_{t-1}}$ обычно почти что стационарны.

Начнем с обобщений модели ARIMA.

4.1 ARIMAX and Intervention analysis

Определение 3 ($\mathrm{ARIMAX}(p,d,q)$).

$$\Delta^d y_t = \alpha_1 \Delta^d y_{t-1} + \dots + \alpha_p \Delta^d y_{t-p} + \beta_1 \Delta^d \varepsilon_{t-1} + \dots$$
$$\dots + \beta_q \Delta^d \varepsilon_{t-q} + \gamma_1 X_{1t} + \dots + \gamma_k X_{kt}, \quad (2)$$

где X_{it} – значения стационарного ряда в какой-то момент времени (можем при помощи i сдвигать время).

Рассмотрим $\mathrm{AR}(1)$ модель с простым структурным срывом:

$$y_t = \alpha_0 + \alpha_1 t_{t-1} + c z_t + \varepsilon_t, \quad (3)$$
$$z_t = \begin{cases} 1, & t \geq t^*, \\ 0, & t < t^*. \end{cases} \quad (4)$$

Мгновенный эффект $+c$. Каков будет долгосрочный эффект?

$$\text{До:} \quad \mathbb{E}[y_t] = \alpha_0 + \alpha_1 \mathbb{E}[y_{t-1}], \quad (5)$$
$$\text{После:} \quad \mathbb{E}[y_t] = (\alpha_0 + c) + \alpha_1 \mathbb{E}[y_{t-1}]. \quad (6)$$

Долгосрочный эффект: $+\frac{c}{1-\alpha_1}$ (если брать различие по сравнению с моментом структурного срыва).

Определение 4. *Impulse response function – функция следующего вида:*

$$\mathrm{IRF} = \frac{\partial y_t}{\partial z_{t^*}}.$$

Показывает как меняется временной ряд в долгосрочной перспективе (явно выражаем $y_t = y_{t^*+(t-t^*)}$*).*

Рассмотрим

$$y_t = \alpha(L)y_t + \gamma(L)X_t + \varepsilon_t, \tag{7}$$

где α, γ – лаговые полиномы. Выряжая y, получаем

$$y_t = (I - \alpha(L))^{-1}\gamma(L)X_t + (I - \alpha(L))^{-1}\varepsilon_t. \tag{8}$$

Отсюда можно найти IRF по любому из X. Когда мы так делаем, мы предполагаем значимое влияние X на y, при этом X – экзогенный ряд. Но есть проблема: а что если причинность другая? Симс: "это все фигня, го ботать векторную авторегрессию".

4.2 Vector autoregression

Определение 5 ($\mathrm{VAR}(p)$)**.** *Пусть* $y_t = [y_t^1, \dots, y_t^k]^T$*. Модель векторной авторегрессии*[7] p*-ого порядка выглядит следующим образом:*

$$y_t = A_0 + A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t. \tag{9}$$

Оценка происходит построчно при помощи МНК.

Симсон предложил: мы в VAR слишком много фигни оцениваем, нам важно получить состоятельные оценки, а не эффективные.

Пример 2 (Двумерный $\mathrm{VAR}(1)$)**.**

$$x_t = a_0 + a_1 x_{t-1} + a_2 z_{t-1} + \varepsilon_t^1, \tag{10}$$
$$z_t = b_0 + b_1 x_{t-1} + b_2 z_{t-1} + \varepsilon_t^2. \tag{11}$$

Рассмотрим $\mathrm{VAR}(1)$:

$$\begin{aligned}
&y_t = A_0 + A_1 y_{t-1} + \varepsilon_t \\
&(I - A_1 L)y_t = A_0 + \varepsilon_t \\
&y_t = (I - A_1 L)^{-1}A_0 + (I - A_1 L)^{-1}\varepsilon_t \\
&y_t = \widetilde{A}_0 + \underbrace{A_1^n y_{t-n}}_{\to 0} + \sum_{i=1}^{n} A_1^i \varepsilon_{t-i} \to \mathrm{VMA}(\infty)
\end{aligned}$$

5 Байесовская векторная авторегрессия

Замечание: BVAR – некоторое обобщение регуляризации. Байесовский VAR – это про очень большие модели с недостатком данных для точной фреквентивистской оценки.

[7] Интерпретируем это как «всё в прошлом влияет на всё в настоящем»

5.1 Minnesota Prior

Определение 6 (Minnesota Prior). *Положим* $\mathrm{VAR}(p)$: $Y_t = X_t\Phi + \varepsilon_t$, $\varphi := \operatorname{vec}\Phi$, $\varepsilon \sim \mathcal{N}(0;\Sigma)$.

$$\mathrm{Prior}\cdot\varphi \sim \mathcal{N}(\underline{\varphi};\underline{\Xi}).$$

Тогда

$$\mathrm{Posterior}\cdot\varphi^{post} \sim \mathcal{N}(\overline{\varphi};\overline{\Xi}).$$

Рассмотрим $\underline{\varphi} = \mathbb{E}^{prior}[\varphi] = \operatorname{vec}\underline{\Phi}$[8]. *Говорим, что модель имеет априорное распределение Минессоты, если*

$$(A_l)_{ij} = \delta\mathbb{1}(l=1, i=j), \qquad \delta \in [-1,1].$$

Замечание 1. *Смысл этого априорного распределения: пусть все компоненты рядов* $\mathrm{AR}(1)$.

Пример 3. $X \sim U[0,\Theta]$. *Prior:* $\Theta \sim \mathrm{Pareto}(\alpha,\beta)$,

$$f(\theta) = \begin{cases} \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}}, & \theta > \beta > 0, \\ 0, \ \textit{else}. \end{cases}$$

Posterior:

$$f^{post}(\theta) = \mathcal{L}f(\theta) = f(\theta)\prod_{i=1}^{n}\frac{1}{\theta} = \frac{\alpha\beta^\alpha}{\theta^{\alpha+n+1}} \sim \mathrm{Pareto}(\alpha^*,\beta^*),$$

где $\alpha^* = \alpha + n$, $\beta^* = \max(\beta, \max X)$.

Далее все λ – гиперпараметры модели.

$$\underline{\Xi} = \begin{bmatrix} \Xi_{\text{const}} & 0 & 0 & & 0 \\ 0 & \Xi_1 & 0 & \dots & 0 \\ 0 & 0 & \Xi_2 & & 0 \\ & \dots & & \dots & \\ 0 & 0 & 0 & & \Xi_p \end{bmatrix},$$

$$(\Xi_{i,l})_{jj} = \begin{cases} \frac{\lambda_{\text{tight}}}{l^{\lambda_{\text{lag}}}}, & i = j, \\ \frac{(\lambda_{\text{tight}}\lambda_{\text{kron}}\sigma_i)^2}{(l^{\lambda_{\text{lag}}}\sigma_j)^2}, & i \neq j. \end{cases} \qquad l = 1,\dots,p,$$

$$\Xi_{i,\text{const}} = (\lambda_{\text{tight}}\lambda_{\text{const}}\sigma_i)^2.$$

5.2 Forecasting Error Variance Decomposition

$$\hat{Y}_{T+1} = \hat{A}_0 + \hat{A}_1 Y_T + \dots + \hat{A}_p Y_{T-p+1},$$
$$Y_{T+1} = A_0 + A_1 Y_T + \dots + A_p Y_{T-p+1} + \varepsilon_{T+1}.$$

Если оценка хорошая, то она несмещенная и

$$\hat{e}_{T+1} = Y_{T+1} - \mathbb{E}\left[\hat{Y}_{T+1}\right] = \varepsilon_{T+1}.$$

Аналогично,

$$\hat{e}_{T+2} = A_1\varepsilon_{T+1} + \varepsilon_{T+2}.$$

[8] восстановим $\underline{\Phi}$ как обращение оператора vec

Итого,

$$\hat{e}_{T+h;j} = a_{11}\varepsilon_{1;T+1} + a_{1n}\varepsilon_{1;T+n} + \cdots + a_{k1}\varepsilon_{k;T+1} + \ldots,$$

где j – номер y. То есть можем разложить итоговую ошибку как комбинацию ошибок прошлого. Поскольку можем рассматривать ортогонализированный шум, то получаем выражение для дисперсии ошибки.

A Bootstrap sampling

Пусть $X_1, \ldots, X_n$ – выборка. Генерим семплы размера n из этой выборки (равномерно по имеющимся значениям). Делаем $B \sim 10^3$ т.н. bootstrap samples. По каждой подвыборке оцениваем регрессию и получаем B разных оценок β_i в $y = \beta_0 + \beta_1 X$.

Для временных рядов: как попало семплить не можем, т.к. данные упорядоченны во времени. Делим ряд на набор блоков и применяем к ним bootstrap sampling. Если блоки маленькие, то зависимости во времени примерно сохраняются. Такой подход называется *block bootstrap sampling*.

B Информационные критерии

Информационные критерии: чем меньше число, тем лучше модель.

Определение 7 (Информационный критерий Акаике). *Пусть k – число оцениваемых параметров модели, $\mathcal{L}^* = \mathcal{L}(\hat{\theta})$ – правдоподобие полученной оценки. Тогда*

$$\mathrm{AIC} = 2k - 2\ln \mathcal{L}^*.$$

Определение 8 (Информационный критерий Шварца-Байеса). *Пусть k – число оцениваемых параметров модели, n – число наблюдений, $\mathcal{L}^* = \mathcal{L}(\hat{\theta})$ – правдоподобие полученной оценки. Тогда*

$$\mathrm{BIC} = k \ln n - 2\ln \mathcal{L}^*.$$