Методы построения и анализа статистических моделей временных рядов

С.Н. Куприянова

методические указания

Содержание

Определение и структура временного ряда

Временной ряд – это совокупность значений какого-либо показателя за несколько последовательных моментов времени. Каждый уровень временного ряда формируется под воздействием факторов, которые можно подразделить на три большие группы:

- факторы, формирующие тенденцию ряда;
- факторы, формирующие циклические колебания ряда;
- случайные факторы.

Значения каждого последующего уровня ряда зависят от предыдущих значений. Корреляционную зависимость между последовательными уровнями временного ряда называют автокорреляцией уровней ряда.

Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Число периодов, по которым рассчитывается коэффициент автокорреляции, называется лагом.

Эту величину называют коэффициентом автокорреляции уровней k-го порядка, так как он измеряет зависимость между уровнями ряда t и t-k, т.е. при лаге 1.

$$r_k = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y}_{k+1})(y_t - \bar{y}_{k+2})}{\sqrt{\sum_{t=k+1}^{n}(y_t - \bar{y}_{k+1})^2} \cdot \sqrt{\sum_{t=k+1}^{n}(y_t - \bar{y}_{k+2})^2}},$$

где

$$\overline{y_{k+1}} = \frac{\sum_{t=k+1}^{n} y_t}{n-k-1}, \quad \overline{y_{k+2}} = \frac{\sum_{t=k+2}^{n} y_t}{n-k-2}$$

Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют автокорреляционной функцией временного ряда. График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется коррелограммой.

При помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда.

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка τ, ряд содержит циклические колебания с периодичностью в τ моментов времени.

Классификация и свойства стохастических процессов, генерирующих временной ряд

Случайным (стохастическим) процессом называется семейство случайных величин y_t (ω), зависит от пары t$\in$**T**, ω$\in$**Ω**, где t интерпретируется как время, T$\subset$**R**, **Ω** – вероятностное пространство, которому принадлежит элементарное событие ω. Иначе, при каждом t$\in$T y_t (ω) должна быть измерима по ω.

Математическое ожидание $E(X_t)$ может изменяться во времени и представляет собой функцию среднего в зависимости от времени

$$\mu(t) = \mu_t = E[X_t]$$

Дисперсия (X_t) является функцией, также зависящей от времени:

$$\sigma^2(t) = \sigma^2{}_t = E[(X_t - \mu_t)^2]$$

Автоковариация

$$\gamma_{t1t2} = \mathrm{cov}(X_{t1}, X_{t2}) = E[(X_{t1} - \mu_{t1})(X_{t2} - \mu_{t2})]$$

в общем виде зависит от каждого t_1 и t_2.

Конечная реализация $x_1, x_2, \ldots, x_T$ дискретного стохастического процесса $X_1, X_2, \ldots, X_T$ называется временным рядом.

Под ***стационарным процессом в слабом смысле*** понимается стохастический процесс, для которого среднее и дисперсия независимо от рассматриваемого периода времени имеют постоянное значение, а автоковариация зависит только от длины лага между рассматриваемыми переменными.

$$\mu_{t2} = \mu = const$$

$$\sigma_t^{\ 2} = \sigma^2 = const$$

$$\gamma_{t1,t2} = \gamma_{t1-t2} = \gamma_\tau$$

Автоковариация как функция длины лага τ

$$\gamma(\tau) = \gamma_\tau = E[(X_t - \mu)(X_{t-\tau} - \mu)]$$

называется автоковариационной функцией.

Автокорреляционная функция или коэффициент корреляции стационарного стохастического процесса определяется как

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}$$

Стационарность временного ряда означает отсутствие:

- тренда;

- систематических изменений дисперсии;

- строго периодичных флуктуаций;

- систематически изменяющихся взаимозависимостей между элементами временного ряда.

«Белым шумом» называется ряд независимых, одинаково распределенных случайных величин. Свойства «белого шума»:

$$\mu_t = E(a_t) = const = \mu$$

$$\sigma^2{}_t = const = \sigma_a^2$$

$$\gamma_{t1,t2} = 0 \text{ для } t_1 \neq t_2$$

Процесс скользящего среднего порядка q [MA(q)] – это процесс X_r

$$X_t = a_t - \theta_1 a_{t-1} - ... - \theta_q a_{t-q},$$

где a_t – «белый шум» с μ=0.

Если ввести оператор лага L:

$$L(X_t) = X_{t-1};$$

$$L^2(X_t) = X_{t-2};$$

$$L^k(X_t) = X_{t-k},$$

выполнить замену

$$a_{t-k} = L^k(a_t)$$

и использовать функцию оператора

$$\Theta_q(L) = 1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q, \text{ то}$$

процесс MA(q) определяется как

$$X_t = \Theta_q(L) a_t$$

Свойства процесса MA(q)

$$E[X_t] = 0;$$

$$Var[X_t] = \sigma^2 \sum_{i=0}^{q} \theta_i^2;$$

$$\gamma_{t,t+\tau} = \begin{cases} 0 & \tau > q \\ \sigma^2 \sum_{i=0}^{q-\tau} \theta_i \theta_{i+\tau} & \tau = 0,1,\ldots,q \end{cases}$$

Следовательно, процесс MA стационарен в слабом смысле.

Теорема 1: Порядок процесса MA(q)

$$X_t = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$$

определяется количеством q автокорреляционных коэффициентов γ_q, значимо отличных от нуля.

Авторегрессионный процесс порядка p[AR(p)] – стохастический процесс X_t:

$$X_t = \phi_0 + \phi_1 \tilde{O}_{t-1} + \phi_2 \tilde{O}_{t-2} + \ldots + \phi_p \tilde{O}_{t-p} + a_t$$

Используя функцию оператора лага

$$\varphi_p(L) = 1 - \varphi_0 - \varphi_1 L - \varphi_2 L^2 - \ldots - \varphi_p L^p,$$

можно записать

$$\varphi_p(L) X_t = a_t$$

Процесс AR не всегда стационарен.

Его характеристическое уравнение определяется как

$$1-\phi_1 z-\phi_2 z^2-\ldots-\phi_p z^p=0,$$

т.е.

$$\phi_p(z)=0,$$

где z – комплексное число.

Теорема 2 (критерии стационарности AR - процессов).

AR – процесс является стационарным тогда и только тогда, когда его комплексные решения (корни) лежат вне единичного круга, т.е. $|z|>1$.

В частности, если $|z|=1$, процесс называется процессом единичного корня и является нестационарным.

Авторегрессионным процессом скользящего среднего [ARMA(p,q)] называется процесса вида

$$X_t=\varphi_0+\varphi_1\tilde{O}_{t-1}+\varphi_2\tilde{O}_{t-2}+\ldots+\varphi_p\tilde{O}_{t-p}+a_t-\theta_1 a_{t-1}-\ldots-\theta_q a_{t-q}$$

или

$$\hat{O}_p(L)X_t=\phi_0+\Theta_q(L)a_t$$

где $\Phi_p(L)$ и $\Theta_q(L)$ – функции операторов лага соответствующих AR(p) и MA(q) процессов, а ϕ_0, как правило, предполагается равным нулю.

Теорема 3: Стационарный ARMA – процесс $\hat{O}_p(L)X_t=\phi_0+\Theta_q(L)a_t$ может быть представлен как бесконечный AR - процесс или как бесконечный MA – процесс:

$$X_t=\phi_0+a_t-\psi_1 a_{t-1}-\psi_2 a_{t-2}-\ldots$$

или

$$X_t=\phi_0+\psi(L)a_t$$

Бесконечный полигон лага $\Psi(L)$ определяется выражением

$$\psi(L)=\frac{\Theta_q(L)}{\hat{O}_p(L)}$$

Итегрируемость временного ряда. Алгоритмы проверки статистических гипотез о стационарности стохастических процессов

Если исходный временной ряд нестационарен, то взятие его разностей может позволить получить стационарный временной ряд.

Первые разности стохастического процесса имеют вид:

$$(1-L)X_t = \Delta X_t = X_t - X_{t-1}$$

Если первые разности ряда X_t стационарны, то ряд X_t называется интегрируемым первого порядка.

В противном случае дальнейшее взятие разностей приведет ко вторым разностям:

$$(1-L)^2 X_t = \Delta^2 X_t = \Delta X_t - \Delta X_{t-1}$$

Если первый стационарный ряд получается после k-кратного взятия разностей, процесс называется интегрируемым k-го порядка.

Стационарный процесс имеет нулевой порядок интегрируемости.

Рассмотрим авторегрессию без смещения вида

$$Y_t = \alpha_1 Y_{t-1} + \varepsilon_t$$

Для проверки порядка интегрируемости этого базового процесса используется метод Дики-Фулера (DF-тест на единичный корень), по сути представляющий собой алгоритм проверки статистической гипотезы о стационарности самого процесса и его разностей повышающегося порядка.

Выдвигаются две альтернативные гипотезы относительно параметра $\delta = \alpha_1 - 1$ уравнения $\Delta Y_t = \delta Y_{t-1} + \varepsilon_t$, эквивалентного исходному уравнению авторегрессии.

H_0: δ=0 – процесс нестационарен (наличие единичного корня)

Δ<0 – процесс стационарен.

Для проверки временного ряда y_t на порядок интегрируемости рассчитывают значение t-статистики Стьюдента для параметра δ и сравнивают его с верхним и нижним пороговыми значениями DF-статистики из таблицы. Если значение расчетной t-статистики меньше (более отрицательное), чем нижнее критическое значение для соответствующего числа наблюдений n, нулевую гипотезу $\delta=0$ (о наличии единичного корня) следует отклонить и принять альтернативную о стационарности процесса Y_t и т.д.

DF-тест применим также для оценки порядка интегрируемости случайного процесса со смещением, который задается следующим уравнением:

$$\Delta Y_t = a_0 + \alpha_1 Y_{t-1} + \varepsilon_t ,$$

где α_0 – смещение.

Используемый для оценивания порядка интегрируемости механизм аналогичен описанному выше, за исключением применяемой таблицы критических значений для t-критерия Стьюдента.

Решение типового примера в пакете Eviews

Для анализа реального массива данных временного ряда будем использовать программный статистический пакет Eviews, с целью применения его возможности визуализации данных исходного временного ряда, а также встроенных алгоритмов проверки статистических гипотез о стационарности и порядке интегрируемости соответствующего стохастического процесса.

На рисунке 1 представлен график временного ряда (2498 наблюдений), содержащего ежедневные данные о величине закрытия индекса цен акций DAX.

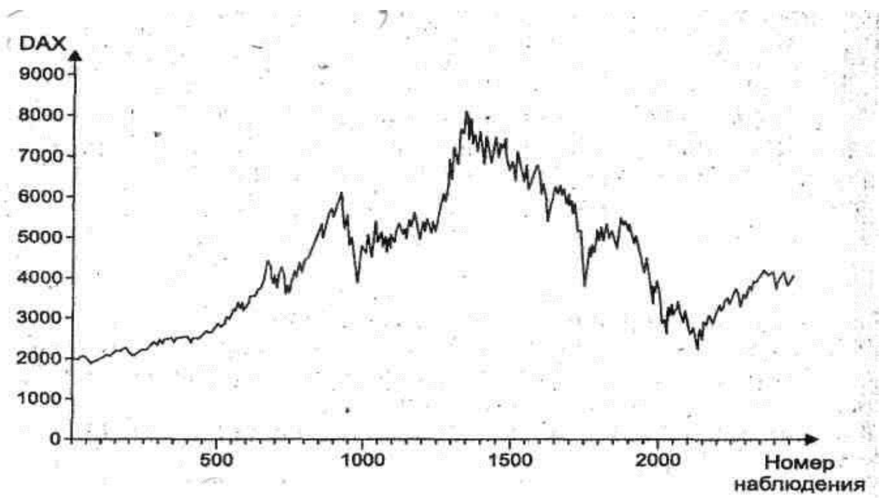

Рис.1. График индекса цен на акции dax

На графике представлен нестационарный процесс, возможно, случайное блуждание. Процесс характеризуется изменяющимся стохастическим трендом и увеличением дисперсии.

Определим порядок интегрируемости dax.

Чтобы протестировать стационарность ряда dax, к исходным данным применяется тест Дики-Фуллера. В таблице 1 содержатся результаты оценки регрессии теста Дики-Фуллера Δdax_t на dax_{t-1}.

Таблица 1

Dickey-Fuller Test Equation

Dependent Variable: Δdax_t

Method: Least Squares

Sample (adjusted): 2 2498; observations: 2497

Variable	Coefficient	Std.Error	t-Statistic	Prob.
dax_t	-0.001453	0.000919	-1.580936	0.1140
C	6.970114	4.213823	1.654107	0.0982

Значение t-статистики коэффициента равно: $t_{факт}$ = - 1,580936. как можно заключить из таблицы 1, это значение превосходит критическое значение t-статистики на 5%-м уровне значимости для модели с константой, которое составляет: $t_{крит}$ = - 2,862497.

Это означает, что на 5%-м уровне значимости ряд dax не является стационарным. Следующий шаг состоит в проверке на стационарность аналогичным образом первых разностей этого временного ряда.

Уравнение теста Дики-Фуллера для первых разностей представляет собой регрессию вторых разностей $\Delta^2 dax_t$ на первую разность, взятую с лагом, Δdax_{t-1}.

Таблица 2

Augmented Dickey-Fuller Test Equation

Dependent Variable: D (DAX,2)=$\Delta^2 dax_t$

Method: Least Squares

Sample (adjusted): 3 2498; observations: 2496

Variable	Coefficient	Std.Error	t-Statistic	Prob.
D (DAX(-1))=Δdax_{t-1}	-0.996234	0.020026	-49.74669	0.0000
C	0.702902	1.427408	0.492432	0.6225

Значение t-статистики, равное t = -49,74669, очевидно, меньше любого критического значения. Таким образом, первые разности стационарны, а сам ряд dax является интегрируемым первого порядка I(1).

Список литературы

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики: Учебник. – М.: ЮНИТИ, 1998.

2. Вентцель А.Д. «Курс теории случайных процессов», М., «Наука», 1975.

3. Справочник по прикладной статистике. В 2-х т.: Пер. с англ. / Под ред. Э. Ллойда, У. Ледермана, Ю.Н. Тюрина. – М.: Финансы и статистика, 1989.

4. Эконометрика. / Под ред. И.И. Елисеевой. – М.: «Финансы и статистика», 2007.