

# Эконометрический ликбез: асимптотика временных рядов

## Зависимость случайных процессов и теория стохастических пределов\*

Джонатан Хилл<sup>†</sup>

*Университет Северной Каролины, Чепел Хилл, США*

В настоящем эссе представлена базовая асимптотическая теория, лежащая в основе оценивания временных рядов. Рассматриваются типы зависимостей, используемые в стохастической теории пределов, в том числе перемешивание, миксингалы и свойство почти эпохальной зависимости. Затем излагаются наиболее общие предельные теоремы (как для сходимости по вероятности, так и сходимости по распределению) для процессов, обладающих вышеперечисленными свойствами и часто используемых в теории временных рядов и ее приложениях.

### 1 Введение

Настоящее эссе представляет собой краткую справку по предельной теории случайных процессов. Эта теория позволит вычислить пределы по вероятности и по распределению для оценок нелинейного МНК, квази-ММП, ОММ и информационно-теоретических оценок, и в том числе оценок обобщенного метода эмпирического правдоподобия. Таким образом, настоящее эссе будет полезно студентам старших курсов университетов и аспирантам младших курсов по экономике, финансам, эконометрике и статистике для вывода асимптотических свойств параметрических и непараметрических оценок в регрессионных моделях.

Множество концепций и аргументов подчерпнуты в замечательных монографиях Doob (1953), Davidson (1994) и Billingsley (1999). Основная часть теории зависимых процессов, изложенной в данном эссе, обязана прогрессу 1970–1990-х годов в теории мартингалов, перемешивания и измеримых функций от перемешивающих последовательностей. Я особо заостряю внимание на свойствах миксингалов и почти эпохальной зависимости как простых, но мощных инструментах описания серийной зависимости и нестационарности временных рядов. Эти идеи принадлежат, среди прочих, авторам Ibragimov (1962), McLeish (1974), Bierens (1981), а также Gallant & White (1988).

Из-за краткости изложения множество концепций зависимости и соответствующих предельных теорем не вошли в настоящее эссе. Важные ссылки, включающие работы по изучению зависимостей Поля Духана и Вея Бяо Ву, даны в работе Dedecker, Doukhan, Lang, Leon, Louhichi & Prieur (2007).

В разделе 2 я напомню читателю базовые определения вероятностных пространств и случайных процессов. В разделе 3 рассматриваются стационарность и эргодичность. Концепции зависимости вводятся в разделе 4, а в разделах 5 и 6 представлены законы больших чисел и центральные предельные теоремы. Наконец, в разделе 7 я определю броуновское движение и слабую сходимость. Некоторые доказательства вынесены в Приложение.

---

\*Перевод Г. Франгуриди. Цитировать как: Хилл, Джонатан (2012) «Зависимость случайных процессов и теория стохастических пределов», Квантиль, №10, стр. 1–31. Citation: Hill, Jonathan (2012) “Dependence and stochastic limit theory,” *Quantile*, No.10, pp. 1–31.

<sup>†</sup>Адрес: Department of Economics, Gardner Hall 208B, University of North Carolina–Chapel Hill, NC 27599-3305, USA. Электронная почта: [jbhill@email.unc.edu](mailto:jbhill@email.unc.edu)

Отныне будем считать, что  $K > 0$  — конечная константа, значение которой может меняться по ходу изложения.  $L_p$ -норма для  $m \times n$  матриц определяется как

$$\|x_t\|_p = \left( \sum_{i=1}^m \sum_{j=1}^n \mathbb{E}|x_{t,i,j}|^p \right)^{1/p}.$$

## 2 Вероятностные пространства, борелевские функции, случайные процессы

### 2.1 $\sigma$ -алгебры, борелевские функции, измеримость по Борелю

Пусть заданы пространство элементарных исходов  $\Omega$ ,  $\sigma$ -алгебра  $\mathfrak{F}$  исходов  $\omega \in \Omega$  и вероятностная мера  $\mathbb{P}$ . *Вероятностным пространством* называется тройка  $(\Omega, \mathfrak{F}, \mathbb{P})$ .

*Сигма-алгеброй*  $\mathfrak{F}$  ( $\sigma$ -алгеброй, или  $\sigma$ -полем) называется множество подмножеств  $\Omega$ , содержащее  $\emptyset$  и  $\Omega$  и замкнутое относительно операций взятия дополнений и счетных объединений. В частности,  $\mathfrak{F}$  является  $\sigma$ -алгеброй *тогда и только тогда, когда*

$$\begin{aligned} \mathfrak{F} &\neq \emptyset, \emptyset \in \mathfrak{F}, \Omega \in \mathfrak{F}, \\ \text{если } A \subseteq \mathfrak{F}, \text{ то } A^c &\subseteq \mathfrak{F}, \\ \text{если } \{A_i\}_{i=1}^\infty, \text{ где } A_i &\subseteq \mathfrak{F} \forall i, \text{ то } \bigcup_{i=1}^\infty A_i \subseteq \mathfrak{F}. \end{aligned}$$

Случайная величина  $x : \Omega \rightarrow \mathbb{R}$   *$\mathfrak{F}$ -измерима* (является  $\mathfrak{F}$ -измеримой функцией или  $\mathfrak{F}/\mathbb{R}$ -измерима) *тогда и только тогда, когда*  $x^{-1}(A) \in \mathfrak{F}$  для любого  $A \in \mathbb{R}$ . Таким образом, если реализации  $x$  на вещественной прямой связаны с подмножествами исходов  $\omega \in \mathfrak{F}$ , то вероятности событий, связанных с  $x$ , могут быть вычислены на основе шансов реализации исходов  $\omega \in \mathfrak{F}$ . Более кратко: реализации  $x$  ассоциированы с событиями в  $\mathfrak{F}$ .

*$\sigma$ -алгеброй, порожденной  $x$*  и обозначаемой как  $\sigma(x)$ , называется пересечение всех  $\sigma$ -алгебр  $\mathfrak{F}$ , относительно которых  $x$  измерима. То есть если  $x$  измерима только на элементах последовательности  $\{\mathfrak{F}_i\}_{i=1}^\infty$   $\sigma$ -алгебр, то

$$\sigma(x) = \bigcap_{i=1}^\infty \mathfrak{F}_i.$$

Это наименьшее множество событий, ассоциированных со всеми реализациями  $x$ .

**Пример 2.1** Пусть  $\Omega = \{\emptyset, 0, 1, 2\}$ . Пусть  $x(0) = x(1) = 5$  и  $x(2) = 10$ . Зададим

$$\mathfrak{F} = \{\{\emptyset\}, \{\Omega\}, \{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}.$$

Легко проверить, что  $\mathfrak{F}$  замкнуто относительно дополнений и счетных пересечений. Например,  $\{0, 1\}^c = \{2\} \in \mathfrak{F}$ . Тогда

$$x^{-1}(5) = \{0, 1\} \in \mathfrak{F} \text{ и } x^{-1}(10) = \{2\} \in \mathfrak{F}.$$

Значит  $x(\omega)$  является  $\mathfrak{F}/\{0, 1, 2\}$ -измеримой. Более того,  $\sigma(x) = \mathfrak{F}$ , так как любая другая  $\sigma$ -алгебра, относительно которой  $x$  измерима, будет «больше» чем  $\mathfrak{F}$ . Другими словами,  $\mathfrak{F}$  есть наименьшая  $\sigma$ -алгебра, относительно которой  $x$  измерима.

*Борелевской  $\sigma$ -алгеброй*  $\beta$  называется  $\sigma$ -алгебра, порожденная подмножествами вида  $(-\infty, x]$  числовой прямой  $(-\infty, \infty)$ :

$$\begin{aligned} \emptyset &\in \beta, \mathbb{R} \subseteq \beta, \\ (-\infty, x] &\subseteq \beta \text{ и } (-\infty, x]^c \subseteq \beta \forall x \in \mathbb{R}, \\ \bigcup_{q \in \mathbb{Q}} &(-\infty, q] \subseteq \beta. \end{aligned}$$

Объединение берется по рациональным концам отрезков, так как рациональные числа суть счетное множество (Doob 1953).

## 2.2 Борелевские функции

Борелевская функция  $g : \Omega \rightarrow \mathbb{R}$ , или функция, измеримая по Борелю (или  $\beta/\mathbb{R}$ -измеримая функция) удовлетворяет условию

$$x^{-1}(A) \in \beta \text{ для каждого } A \in \mathbb{R}.$$

Очень трудно построить пример функции, не измеримой по Борелю. Вообще говоря, мы просто соглашаемся с тем, что можем ассоциировать вещественные интервалы  $(-\infty, w]$  с абстрактными событиями  $\omega$ , и «измерить» шансы  $x(\omega)$ , «измеряя»  $(-\infty, w]$ .

Изложение на таком уровне абстракции бессмысленно. Множество эконометрических результатов получено в очень слабых предположениях относительно участвующих в них функций. Хотя есть и такие, которые предполагают непрерывность функций, для других достаточно измеримости по Борелю. На самом деле, свойство непрерывности строго сильнее свойства измеримости по Борелю.

**Лемма 2.1** *Любая функция  $g : \mathbb{R} \rightarrow \mathbb{R}$ , непрерывная почти всюду, измерима по Борелю.*

## 2.3 Случайные процессы

### 2.3.1 Определение

Случайный процесс есть обобщение измеримого отображения  $\{x\} : \Omega \rightarrow \mathbb{R}^{\mathbb{Z}}$ :

$$\{x(\omega)\} = \{x_t(\omega)\}_{t=-\infty}^{\infty} = \{x_t(\omega) : t \in \mathbb{Z}\}.$$

Вся последовательность  $\{\dots, x_t, x_{t+1}, \dots\}$  должна быть совместно измерима: должно быть возможно приписать функцию совместного распределения подпоследовательности  $\{x_{t_1}, \dots, x_{t_l}\}$  для любого непустого  $l$ -набора  $\{t_1, \dots, t_l\} \in \mathbb{Z}^l$ .

Заметим, что вся последовательность  $\{x_t(\omega)\}_{t=-\infty}^{\infty}$  есть функция одного случайного исхода  $\omega \in \Omega$ . Мы должны представлять последовательные во времени события как одну возможную реализацию случайного эксперимента  $\omega \in \Omega$ .

Мы говорим об «обобщении», потому что индекс не обязан быть ни целочисленным, ни сохранять естественный линейный порядок. На практике координата  $t$  обозначает время, а  $x_{t+1}(\omega)$  реализуется после  $x_t(\omega)$  и никогда в обратную сторону.

Траекторией называется наблюдаемая последовательность, зависящая от конкретного (не обязательно наблюдаемого) случайного исхода  $\omega \in \Omega$ :

$$\{x_t(\omega)\}_{t=1}^n.$$

### 2.3.2 Свойства

Процесс  $\{x(\omega)\}$  равномерно ограничен по вероятности, если для любого  $\varepsilon > 0$  существует  $K_\varepsilon < \infty$ , такое что

$$\sup_{t \in \mathbb{Z}} \mathbb{P}\{|x_t| > K_\varepsilon\} \leq \varepsilon.$$

Равномерная ограниченность исключительно важна, так как многие процессы зависят от времени и могут неограниченно расти или падать. Рассмотрим простой линейный тренд

$$x_t = \beta_0 + \beta_1 t + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} (0, \sigma^2),$$

где  $\epsilon_t$  обладает строго положительной и непрерывной плотностью (например, гауссовской). Ясно, что процесс  $x_t$  не является равномерно ограниченным по вероятности, так как для любого конечного  $K_\epsilon$

$$\sup_{t \in \mathbb{Z}} \mathbb{P} \{ |\beta_0 + \beta_1 t + \epsilon_t| > K_\epsilon \} = 1.$$

Случайный процесс *равномерно  $L_p$ -ограничен*,  $p > 0$ , если

$$\sup_{t \in \mathbb{Z}} \|x_t\|_p < \infty.$$

Заметим, что равномерная  $L_p$ -ограниченность влечет равномерную ограниченность по вероятности в силу неравенства Маркова:

$$\sup_{t \in \mathbb{Z}} \mathbb{P} \{ |x_t| > K_\epsilon \} \leq \frac{1}{K_\epsilon^p} \sup_{t \in \mathbb{Z}} \|x_t\|_p^p < \infty.$$

Поэтому для любого  $\epsilon > 0$  выберем

$$\frac{1}{\epsilon^{1/p}} \sup_{t \in \mathbb{Z}} \|x_t\|_p = K_\epsilon.$$

Если  $p = \infty$ , то  $L_p$ -норма превращается в *равномерную норму*:

$$\sup_{t \in \mathbb{Z}} \|x_t\|_p = \sup_{t \in \mathbb{Z}} |x_t| < \infty \text{ a.s.},$$

а значит  $x_t$  *почти наверное равномерно ограничен* (например,  $\mathbb{P}\{a \leq x_t \leq b\} = 1$  для любого конечного  $a \leq b$ : это верно, если данные усечены (скажем,  $x_t \times \mathbb{I}\{|x_t| \leq a\} \in [-a, a]$  п.н. равномерно по  $t$  для любого конечного  $a > 0$ ).

Случайная величина  $x_t$  *равномерно интегрируема*, если

$$\lim_{M \rightarrow \infty} \mathbb{E} [|x_t| \times \mathbb{I}\{|x_t| \geq M\}] = 0.$$

Это значит, что экстремальные значения (т.е. хвосты распределения) не доминируют, так как реализуются слишком редко. Процесс  $\{x(\omega)\}$  *равномерно интегрируем*, если

$$\lim_{M \rightarrow \infty} \sup_{t \in \mathbb{Z}} \mathbb{E} [|x_t| \times \mathbb{I}\{|x_t| \geq M\}] = 0.$$

**Теорема 2.1**  $L_{1+\delta}$ -ограниченность для любого  $\delta > 0$  влечет равномерную интегрируемость.

## 3 Стационарность и эргодичность

### 3.1 Стационарность

Процесс  $\{x_t(\omega)\}$  называется *строго стационарным*, если *преобразования сдвига сохраняют меру*: совместное распределение  $\{x_{t-m_1}, x_{t-m_2}, \dots, x_{t-m_k}\}$  зависит только от  $\{m_1, \dots, m_k\}$ , но не от  $t$ . Это означает, что *конечномерные распределения* (совместные распределения векторов  $\{x_{t-m_1}, x_{t-m_2}, \dots, x_{t-m_k}\}$ ) зависят только от относительного смещения, но не от конкретных моментов времени.

Процесс  $\{x_t(\omega)\}$  *слабо стационарен*, если

$$\begin{aligned} \mathbb{E} [x_t] &= \mu \quad \forall t, \\ \mathbb{E} [(x_t - \mathbb{E} [x_t])^2] &= \sigma^2 < \infty \quad \forall t, \\ \gamma(s, t) &:= \mathbb{E} [(x_s - \mathbb{E} [x_s]) (x_t - \mathbb{E} [x_t])] = \gamma(h), \text{ где } h = |s - t|. \end{aligned}$$

Иначе говоря, стационарными характеристиками процесса являются только среднее, дисперсия и автоковариации.

Если процесс  $\{x_t(\omega)\}$  строго стационарен и  $L_p$ -ограничен, то  $\forall r \leq p$

$$\mathbb{E}|x_t|^r = \mathbb{E}|x_1|^r.$$

## 3.2 Эргодичность: эргодическая теорема как ЗБЧ

### 3.2.1 Эргодичность: определение и пример

Пусть процесс  $\{x_t(\omega)\}$  задан на  $(\Omega, \mathfrak{F}, \mathbb{P})$ . Рассмотрим траекторию этого процесса для некоторого исхода  $\omega_0$ ,  $\{x_t(\omega_0)\}_{t=1}^n$ , и последовательность сечений этого процесса для некоторого момента времени  $t_0$ ,  $\{x_{t_0}(\omega_i)\}_{i=1}^N$ . Мы хотим узнать, при каких условиях среднее по реализациям для любого момента  $t_0$

$$\frac{1}{N} \sum_{i=1}^N x_{t_0}(\omega_i)$$

совпадает в пределе со средним по времени при фиксированном  $\omega_0$

$$\frac{1}{n} \sum_{t=1}^n x_t(\omega_0).$$

Если процесс  $\{x_t(\omega)\}$  стационарен, то  $\mathbb{E}[x_t] = \mathbb{E}[x_s] \forall s, t$ , в случае чего нам хотелось бы узнать, когда

$$\frac{1}{n} \sum_{t=1}^n x_t(\omega_0) \rightarrow \mathbb{E}[x_{t_0}] \text{ для произвольного момента } t_0.$$

**Пример 3.1** Синоптик  $A$  собирает данные об осадках  $x_{t_0}(\omega_i)$  из  $N$  точек  $\{\omega_i\}_{i=1}^N$  в Северной Каролине в момент времени  $t_1$ . Синоптик  $B$  собирает данные об осадках  $x_t(\omega_1)$  в моменты времени  $t \in \{1, \dots, n\}$  в Северной Каролине в точке  $\omega_1$ . Ясно, что данные, собранные двумя синоптиками, могут не иметь одинаковых популяционных характеристик. Точка  $\omega_1$  может быть очень засушливым местом, а в момент времени  $t_1$  погода может быть очень дождливая, так что

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_{t_1}(\omega_i) > \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n x_t(\omega_1).$$

Стационарный процесс  $\{x_t(\omega)\}$  называется *эргодичным для среднего*, если

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_{t_0}(\omega_i) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n x_t(\omega) = \mathbb{E}[x_t] \forall t.$$

В пространстве элементарных исходов нет «систематического сдвига» (то есть доминирующих засушливых регионов в Северной Каролине: флуктуации во времени в точке  $\omega$  в среднем не отличаются от флуктуаций в пространстве).

### 3.2.2 Эргодичность: память в $\sigma$ -алгебрах

Пусть процесс  $\{x_t(\omega)\}$  стационарен. На практике стационарность обычно не является реалистичным предположением в силу наличия временных трендов в среднем или дисперсии или сезонных эффектов. Тем не менее, эргодичность стационарного процесса допускает значительную персистентность и поэтому интересна, так как представляет собой уровень, с которым можно сравнивать персистентность других концепций зависимости.

Пусть  $\mathfrak{F} = \sigma(x)$  и  $A, B \in \mathfrak{F}$ . Пусть также  $T$  есть преобразование сдвига, сохраняющее меру:  $\mathbb{P}\{TB\} = \mathbb{P}\{B\}$ , и  $TB$  сдвигает исходы, составляющие событие  $A$ . Таким образом,  $T^k B$  производит  $k$  операций сдвига.

Говорят, что *сохраняющее меру преобразование сдвига  $T$  эргодично*, если средняя память исчезает.

**Теорема 3.1** *Сохраняющее меру преобразование сдвига  $T$  эргодично тогда и только тогда, когда  $\forall A, B \in \mathfrak{F}$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n [\mathbb{P}\{A \cap T^i B\} - \mathbb{P}\{A\} \mathbb{P}\{B\}] = 0. \quad (1)$$

*То есть в среднем события независимы.*

Процесс  $\{x_t(\omega)\}$  *эргодичен*, если для любого момента времени  $t$

$$x_t(T^k \omega) = x_{t+k}(\omega),$$

где преобразование  $T$  сохраняет меру и эргодично. Это ключевая идея (хотя и абстрактная). Процесс *эргодичен*, когда сдвиг в пространстве исходов эквивалентен сдвигу по времени. Таким образом, случайная выборка из сечения процесса в конкретный момент времени даст (асимптотически) ту же информацию, что и наблюдения во времени. Сечение момента  $t$  (которого мы обычно не имеем) будет иметь ту же вероятностную структуру, что и последовательность во времени (которую мы имеем).

**Пример 3.2** Пусть  $\Omega = \{0, 1\}$ , так что  $\omega = 0$  или  $1$ , и  $\mathfrak{F} = \{\{\emptyset\}, \{\Omega\}, \{0\}, \{1\}, \{0, 1\}\}$ ,  $\mathbb{P}\{0\} = \mathbb{P}\{1\} = 1/2$ . Пусть  $T0 = 1$  и  $T1 = 0$ . Ясно, что  $T$  сохраняет меру:  $\mathbb{P}\{T\omega\} = \mathbb{P}\{\omega\} = 1/2$ . Если  $A = \{0, 1\}$  и  $B = \{0\}$ , то для любого четного  $k$

$$\mathbb{P}\{A \cap T^k B\} - \mathbb{P}\{A\} \mathbb{P}\{B\} = \mathbb{P}\{0\} - \mathbb{P}\{\{0, 1\}\} \mathbb{P}\{0\} = 0,$$

а для любого нечетного  $k$

$$\mathbb{P}\{A \cap T^k B\} - \mathbb{P}\{A\} \mathbb{P}\{B\} = \mathbb{P}\{1\} - \mathbb{P}\{\{0, 1\}\} \mathbb{P}\{0\} = 0.$$

Аналогично, если  $A = \{0\}$  и  $B = \{1\}$ , то для любого четного  $k$

$$\mathbb{P}\{A \cap T^k B\} - \mathbb{P}\{A\} \mathbb{P}\{B\} = \mathbb{P}\{\emptyset\} - \mathbb{P}\{0\} \mathbb{P}\{1\} = -1/4,$$

а для любого нечетного  $k$

$$\mathbb{P}\{A \cap T^k B\} - \mathbb{P}\{A\} \mathbb{P}\{B\} = \mathbb{P}\{0\} - \mathbb{P}\{0\} \mathbb{P}\{1\} = 1/4,$$

что в среднем дает ноль. И так далее. Преобразование  $T$  эргодично. Определим теперь  $x_1(\omega) := \omega$  и  $x_t(\omega) := x_1(T^{t-1}\omega)$  для  $t = 2, 3, \dots$ . Тогда последовательность случайно выходит из 0 или 1, а после этого чередуется, скажем, как 0, 1, 0, 1, .... Среднее по времени равно, очевидно,  $1/2$ , и по построению среднее по реализациям равно  $1/2$ :  $\mathbb{E}[x_t(\omega)] = 0 \times .5 + 1 \times .5 = .5$ . Таким образом, процесс  $\{x_t(\omega)\}$  эргодичен.

**Пример 3.3** Пусть  $\mathfrak{S} = \{\{\emptyset\}, \{\Omega\}, \omega_1, \dots, \omega_6\}$ , где  $\mathbb{P}\{\omega_i\} = 1/6$ . Рассмотрим подмножества  $A = \{\omega_1, \omega_2, \omega_3\}$ ,  $B = \{\omega_1, \omega_2, \omega_4\}$ . Преобразование сдвига  $T^k B = \{\omega_{1+2k}, \omega_{2+2k}, \omega_{4+2k}\}$  сохраняет меру, так как  $\mathbb{P}\{T^k B\} = \mathbb{P}\{B\} = 1/2$ , где  $\omega_b = \omega_{b-6}$  для любого  $b > 6$ . Тогда

$$\begin{aligned}\mathbb{P}\{A \cap TB\} - \mathbb{P}\{A\}\mathbb{P}\{B\} &= \mathbb{P}\{\omega_3\} - 1/4 = 1/6 - 1/4 = -1/12 \\ \mathbb{P}\{A \cap T^2 B\} - \mathbb{P}\{A\}\mathbb{P}\{B\} &= \mathbb{P}\{\omega_2\} - 1/4 = 1/6 - 1/4 = -1/12 \\ \mathbb{P}\{A \cap T^3 B\} - 1/4 &= \mathbb{P}\{\omega_1, \omega_2\} - 1/4 = 1/3 - 1/4 = 1/12 \\ \mathbb{P}\{A \cap T^4 B\} - 1/4 &= \mathbb{P}\{\omega_3\} - 1/4 = 1/6 - 1/4 = -1/12,\end{aligned}$$

и так далее. Таким образом,

$$\frac{1}{n} \sum_{i=1}^n [\mathbb{P}\{A \cap T^i B\} - \mathbb{P}\{A\}\mathbb{P}\{B\}] = \frac{1}{n} [-1/12 - 1/12 + 1/12 + \dots] \rightarrow -1/3,$$

и значит, преобразование  $T$  неэргодично.

Есть два способа описать память эргодичного процесса. Первый способ просто переформулирует (1) в терминах расстояния по времени, в противовес абстрактному «сдвигу».

**Теорема 3.2** Пусть  $\mathfrak{S}_t$  обозначает  $\sigma$ -алгебру, порожденную  $\{x_t, x_{t-1}, \dots\}$ . Процесс  $\{x_t\}$  эргодичен тогда и только тогда, когда для любых  $A_t, B_t \subseteq \mathfrak{S}_t$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n [\mathbb{P}\{A_t \cap B_{t-i}\} - \mathbb{P}\{A_t\}\mathbb{P}\{B_{t-i}\}] = 0. \quad (2)$$

Теорема утверждает, что процесс  $x_t$  эргодичен, если его серийная память (то есть зависимости между  $x_t$  и  $x_{t-i}$ ) в среднем исчезает. Это не значит, что далекие по времени события независимы, то есть эргодичность допускает слишком сильную персистентность, чтобы найти хорошее применение в теории (предельная теория требует более сильных ограничений). Если наложить дополнительные структурные ограничения, из эргодичности будет следовать, что средние по времени имеют корректно определенный предел по вероятности.

**Теорема 3.3** Если  $\{x_t(\omega)\}$  стационарен, эргодичен и  $L_2$ -ограничен, то

$$\frac{1}{n} \sum_{i=1}^n \mathbb{C}(x_i, x_1) \rightarrow 0.$$

Следующий результат представляет собой простое достаточное условие эргодичности для среднего.

**Теорема 3.4** Если процесс  $\{x_t(\omega)\}$  слабо стационарен и  $\sum_{i=0}^{\infty} |\mathbb{C}(x_{i+1}, x_1)| < \infty$ , то он эргодичен для среднего:

$$\frac{1}{n} \sum_{t=1}^n x_t(\omega) \rightarrow \mathbb{E}[x_t(\omega)].$$

**Пример 3.4** Пусть  $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$ ,  $\sigma^2 < \infty$ . Тогда процесс  $MA(1)$ , определяемый как

$$x_t = \theta \epsilon_{t-1} + \epsilon_t, \quad |\theta| < \theta,$$

имеет очень короткую память. Ясно, что для всех  $A_t, B_t \subseteq \mathfrak{S}_t = \sigma(x_\tau : \tau \leq t)$  и всех  $i \geq 2$ ,  $\mathbb{P}\{A_t \cap B_{t-i}\} = \mathbb{P}\{A_t\}\mathbb{P}\{B_{t-i}\}$ . Следовательно,  $1/n \sum_{i=1}^n [\mathbb{P}\{A_t \cap B_{t-i}\} - \mathbb{P}\{A_t\}\mathbb{P}\{B_{t-i}\}] \rightarrow 0$ , и  $1/n \sum_{t=1}^n x_t(\omega) \rightarrow 0$ . Ясно, что  $\sum_{i=0}^{\infty} |\mathbb{C}(x_{i+1}, x_1)| = \sigma^2(1 + \theta^2) + \sigma^2|\theta| < \infty$ .

**Пример 3.5** Пусть  $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$  и

$$x_t = a + bt + \epsilon_t.$$

Так как  $\mathbb{E}[x_t] = a + bt$  не стационарно (и не слабо стационарно), теорема 3.4 неприменима.

**Лемма 3.1** Рассмотрим процесс  $AR(1)$

$$x_t = \phi x_{t-1} + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} (0, \sigma^2) \text{ и } |\phi| < 1.$$

Тогда  $\sum_{i=0}^{\infty} |\mathbb{C}(x_{i+1}, x_1)| = \sigma^2 / [(1 - \phi^2)(1 - |\phi|)] < \infty$ . Следовательно,  $x_t$  эргодичен для среднего.

## 4 Концепции зависимости и неоднородности

### 4.1 Перемешивание

#### 4.1.1 Преобразование перемешивания

Пусть  $\{x_t(\omega)\}$  — случайный процесс. Эргодичность — довольно слабое свойство памяти, применимое к последовательностям по времени *в среднем*, но не означающее, что далекие события асимптотически независимы. В экономике обычно предполагается, что бесконечно далекие события ( $x_t$  и  $x_{t-N}$  при  $N \rightarrow \infty$ ) независимы, или этого требует подходящая центральная предельная теорема. Это следует также из условия *перемешивания*.

Пусть  $\mathfrak{F} = \sigma(x)$  и  $A, B \in \mathfrak{F}$ . Эргодичное преобразование  $T$ , сохраняющее меру, называется *перемешивающим*, если для любых  $A, B \in \mathfrak{F}$

$$\lim_{k \rightarrow \infty} \mathbb{P} \{T^k A \cap B\} = \mathbb{P} \{A\} \mathbb{P} \{B\}.$$

Процесс  $\{x_t(\omega)\}$  является перемешивающим, если

$$x_t(T^k \omega) = x_{t+k}(\omega),$$

где  $T$  — эргодичное сохраняющее меру перемешивающее преобразование.

#### 4.1.2 Сильное и равномерное перемешивание

Рассмотрим более конкретные случаи перемешивания. Пусть  $\{x_t\}$  есть случайный процесс с  $\sigma$ -алгеброй

$$\mathfrak{F}_a^b := \sigma(x_t : a \leq t \leq b).$$

Определим последовательности  $\alpha_m$  и  $\phi_m$  следующим образом:

$$\alpha_m = \sup_{t \in \mathbb{Z}} \sup_{G \in \mathfrak{F}_{-\infty}^t, H \in \mathfrak{F}_{t+m}^{\infty}} |\mathbb{P} \{G \cap H\} - \mathbb{P} \{G\} \mathbb{P} \{H\}|$$

и

$$\phi_m = \sup_{t \in \mathbb{Z}} \sup_{G \in \mathfrak{F}_{-\infty}^t, H \in \mathfrak{F}_{t+m}^{\infty}} |\mathbb{P} \{H|G\} - \mathbb{P} \{H\}|.$$

Процесс  $\{x_t\}$  является *равномерно перемешивающим*, если

$$\phi_m \rightarrow 0 \text{ при } m \rightarrow \infty,$$



и *сильно перемешивающим*, если

$$\alpha_m \rightarrow 0 \text{ при } m \rightarrow \infty.$$

Перемешивание просто означает, что асимптотически далекие события  $G \in \mathfrak{S}_{-\infty}^t$  и  $H \in \mathfrak{S}_{t+m}^\infty$  независимы. Поэтому случайные величины  $x_{t-m}$  и  $x_t$  независимы при  $m \rightarrow \infty$  для любого  $t$ .

Нетрудно доказать, что

$$\alpha_m \leq \phi_m,$$

то есть «сильное перемешивание» на самом деле слабее «равномерного перемешивания», а значит представляет больший теоретический интерес.

### 4.1.3 Размер перемешивания

Скорость, с которой  $x_t$  и  $x_{t-k}$  становятся независимыми при  $k \rightarrow \infty$ , измеряется «размером». Есть несколько способов определить размер, однако наиболее популярным является гиперболическая ограниченность.

Говорят, что процесс  $x_t$  *сильно* или *равномерно перемешивающий* размера  $\lambda > 0$ , если

$$\alpha_m = o(m^{-\lambda}) \text{ или } \phi_m = o(m^{-\lambda}).$$

Гиперболическая ограниченность — наиболее слабое условие, применяемое в теории, так как оно представляет собой «*длинную*» память:  $x_t$  и  $x_{t-k}$  становятся независимыми очень медленно.

Обсудим, что на самом деле означает размер перемешивания. Если  $x_t$  — *сильно перемешивающий* процесс размера 2, то  $\forall \delta \leq 2$

$$m^\delta \times \sup_{t \in \mathbb{Z}} \sup_{G \in \mathfrak{S}_{-\infty}^t, H \in \mathfrak{S}_{t+m}^\infty} |\mathbb{P}\{G \cap H\} - \mathbb{P}\{G\}\mathbb{P}\{H\}| \rightarrow 0$$

при  $m \rightarrow \infty$ . Таким образом,  $x_t$  и  $x_{t-k}$  становятся независимыми со скоростью  $m^2$ . Чем больше размер, тем больше скорость сходимости, и тем слабее зависимость (меньше персистентность).

С другой стороны, говорят, что процесс  $x_t$  *геометрически* *сильно* или *равномерно перемешивающий*, если существует число  $\rho \in (0, 1)$ , такое что

$$\alpha_m = o(\rho^m) \text{ или } \phi_m = (\rho^m).$$

Легко доказать, что геометрическое перемешивание влечет перемешивание любого размера  $\lambda > 0$ , то есть геометрическое перемешивание — *сильное* условие, и представляет собой «*короткую*» память:  $x_t$  и  $x_{t-k}$  становятся независимыми относительно быстро.

Из перемешивания размера  $\lambda > 1$  (а значит и геометрического перемешивания) следует суммируемость коэффициентов перемешивания:

$$\sum_{m=0}^{\infty} \alpha_m < \infty \text{ или } \sum_{m=0}^{\infty} \phi_m < \infty.$$

### 4.1.4 Перемешивание во временных рядах

Для того чтобы понять, что собой представляет перемешивание, рассмотрим следующий пример.

**Пример 4.1** Рассмотрим бросание одного красного и одного синего плавающих мячей в бесконечно большой бассейн, и пусть  $x_t$  обозначает евклидово расстояние между ними. Рассмотрим также последовательность шок по времени: мы опускаем в воду большую палку и шевелим ее. Если мы пошевелим палкой однократно, расстояние между мячами станет  $x_{t+1}$ . Ясно, что  $x_{t+1}$  будет зависеть от  $x_t$  (если  $x_t = 10$ , то  $x_{t+1}$  будет равно скорее 12, чем, скажем, 20000). Однако после  $k$ -кратного шевеления или *перемешивания* бассейна, расстояние  $x_{t+k}$  при  $k \rightarrow \infty$  не будет зависеть от  $x_t$ , потому что в долгосрочной перспективе  $x_{t+k}$  может принять любое значение в  $\mathbb{R}_+$ , независимо от исходного расстояния  $x_t$ . Таким образом,  $x_t$  называется «перемешивающим».

Этот пример не абстрактен. Если случайные шоки  $\epsilon_t$  входят в процесс, порождающий, например, дневные данные по некоторому обменному курсу  $x_t$ , в соответствии с формулой

$$x_t = f(x_{t-1}, \dots, x_{t-p}; \phi) + \epsilon_t,$$

где  $f : \mathbb{R}^p \times \Phi \rightarrow \mathbb{R}$  неизвестна,  $\Phi \subseteq \mathbb{R}^p$ , то как текущий курс  $x_t$  зависит от прошлых курсов  $x_{t-k}$  при  $k \rightarrow \infty$ ? В примере с бассейном прошлые шоки  $\epsilon_{t-i}$  (шевеления) оказывали все уменьшающийся эффект на текущее расстояние  $x_t$ , а недавние шоки (шевеления) имели больший эффект. Вообще говоря, очень трудно ответить на поставленный вопрос с точки зрения перемешиваний, и только недавно полученные результаты прояснили, какие виды нелинейных процессов перемешиваются.

Рассмотрим частный случай гораздо более общего доказательства для нелинейных авторегрессий. См. An & Huang (1996) и Leibscher (2005).

**Лемма 4.1** *Рассмотрим процесс  $AR(1)$*

$$x_t = \phi x_{t-1} + \epsilon_t,$$

где  $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$  обладает строго положительной и непрерывной функцией плотности, и  $|\phi| < \infty$ . Тогда процесс  $x_t$  геометрически сильно перемешивающий.

Условие непрерывности и положительности плотности ошибок  $\epsilon_t$  на всем носителе очень ограничительно, хотя плотности нормального, экспоненциального и многих других распределений ему удовлетворяют. Несложно найти простой стационарный эргодичный процесс  $AR(1)$ , не являющийся сильно перемешивающим, потому что плотность ошибок  $\epsilon_t$  разрывна (например, распределение Бернулли). См. Andrews (1984).

#### 4.1.5 Свойства перемешивания

Приводимые ниже результаты полезны для расширения понятий эргодичности и законов больших чисел на случайные величины с перемешиванием.

**Лемма 4.2** *Для  $p > 1$  и  $r \geq p/(p-1)$*

$$|\mathbb{C}(x_t, x_{t-m})| \leq 2 \left( 2^{1/p} + 1 \right) \alpha_m^{1-1/p-1/r} \|x_t\|_p \|x_t\|_r,$$

где процесс  $x_t$  является  $L_{\max\{p,r\}}$ -ограниченным.

**Лемма 4.3** *Для  $r \geq 1$*

$$|\mathbb{C}(x_t, x_{t-m})| \leq 2 \phi_m^{1/r} \|x_t\|_r \|x_t\|_{r/(r-1)},$$

где процесс  $x_t$  является  $L_{\max\{r,r/(r-1)\}}$ -ограниченным.

Оба результата мгновенно влекут эргодичность для среднего в процессах с перемешиванием. Случай геометрического перемешивания тривиален.

**Теорема 4.1** Если процесс  $x_t$  слабо стационарен,  $L_{2+\delta}$ -ограничен и является геометрически сильно перемешивающим, то он эргодичен для среднего:  $1/n \sum_{t=1}^n x_t \xrightarrow{P} \mathbb{E}[x_t]$ .

Одна из наиболее полезных характеристик перемешивания случайных величин — это результат о том, что свойство перемешивания распространяется на функции конечного числа лагов от перемешивающихся случайных величин.

**Лемма 4.4** Пусть  $y_t := g(x_t, x_{t-1}, \dots, x_{t-k})$  есть измеримая функция для конечного  $k > 0$ . Если  $x_t$  сильно или равномерно перемешивающий размера  $\lambda > 0$ , то и  $y_t$  тоже.

**Пример 4.2** Если процесс  $x_t$  сильно перемешивающий размера  $\lambda > 0$ , то сильно перемешивающий и процесс

$$\sum_{i=0}^k \psi_i x_{t-i}$$

для любого конечного  $k > 0$  и любого  $\psi_i \in \mathbb{R}$ .

**Пример 4.3** Если  $\epsilon_t$  сильно перемешивающий размера  $\lambda > 0$ , а  $\{x_t\}$  — случайный процесс, то  $\mathbb{E}[x_t | \epsilon_t, \dots, \epsilon_{t-k}]$  сильно перемешивающий размера  $\lambda > 0$ .

## 4.2 Мартингалы и мартингал-разности

Рассмотрим процесс  $\{x_t\}$  с  $\sigma$ -алгеброй  $\mathfrak{F}_t := \sigma(x_\tau : \tau \leq t)$ . Мартингал  $\{x_t, \mathfrak{F}_t\}_{t=0}^\infty$  удовлетворяет условию

$$\mathbb{E}[x_t | \mathfrak{F}_{t-1}] = x_{t-1}.$$

Слово «мартингал» изначально относилось к игровой стратегии, а как концепция в теории вероятностей появилось по крайней мере в работах Lévy (1925, 1954) и Doob (1953).

**Пример 4.4** Рассмотрим процесс AR(1) с единичным наклоном:

$$x_t = x_{t-1} + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} (0, \sigma^2).$$

Автоковариации не суммируются, и  $x_t$  слишком персистентен, чтобы быть перемешивающим. Однако такой процесс является мартингалом, так как из независимости шоков следует, что  $\mathbb{E}[x_t | \mathfrak{F}_{t-1}] = x_{t-1}$ . Гипотеза эффективного рынка утверждает, что цены активов являются мартингалами: цена актива в момент времени  $t - 1$  на эффективном рынке учитывает всю информацию, так что лучший (в среднеквадратичном смысле) прогноз на будущее и есть цена в момент  $t$ .

Хотя в силу своих математических свойств мартингалы широко применяются в экономике и финансах, они допускают гораздо более сильную память, чем требуют законы больших чисел и центральные предельные теоремы.

Напротив, первая разность мартингала  $x_t$ ,

$$\epsilon_t := x_t - x_{t-1},$$

обладает такой простой структурой, что для нее верны ЦПТ и ЗБЧ. Процесс  $\{\epsilon_t, \mathfrak{F}_t\}_{-\infty}^{\infty}$  называется *мартингал-разностью* с характерными свойствами

$$\mathbb{E}[\epsilon_t | \mathfrak{F}_{t-1}] = 0 \quad (3)$$

и

$$\mathbb{E}[\epsilon_s \epsilon_t] = \mathbb{E}(\epsilon_s \mathbb{E}[\epsilon_t | \mathfrak{F}_{t-1}]) = \mathbb{E}(\epsilon_s \times 0) = 0. \quad (4)$$

В общем случае мартингал-разность определяется через свойство (3) без привязки к какому-либо мартингалу. Ясно, что последовательность независимых одинаково распределенных центрированных величин  $\{\epsilon_t, \mathfrak{F}_t\}_{-\infty}^{\infty}$ ,  $\mathfrak{F}_t := \sigma(\epsilon_\tau : \tau \leq t)$  является мартингал-разностью  $\{\epsilon_t, \mathfrak{F}_t\}_{-\infty}^{\infty}$ , так как по построению  $\mathbb{E}[\epsilon_t | \mathfrak{F}_{t-1}] = \mathbb{E}[\epsilon_t] = 0$ .

В силу того, что все ковариации равны нулю, *мартингал-разность*  $\{\epsilon_t\}$  эргодична для среднего (см. теорему 3.4).

#### 4.2.1 Линейная модель распределенных лагов

Линейная модель распределенных лагов *мартингал-разности* — полезная модель временных рядов, которая проявляется в различных постановках (например слабо стационарные модели ARIMA или случайные блуждания).

Пусть  $\{\epsilon_t, \mathfrak{F}_t\}_{-\infty}^{\infty}$  — мартингал-разность,  $\mathbb{E}[\epsilon_t] = 0$ ,  $\mathbb{E}[\epsilon_t^2] = \sigma^2$ , и положим

$$x_t := \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}.$$

Тогда автоковариации  $\gamma(h) := \mathbb{E}[x_t x_{t-h}]$  равны (используем (4))

$$\gamma(h) = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+h}.$$

Автоковариации  $\{\gamma(h)\}_{h \in \mathbb{N}}$  абсолютно суммируемы, если  $\{\psi_i\}_{i=0}^{\infty}$  абсолютно суммируемы:

$$\sum_{h=0}^{\infty} |\gamma(h)| = \sigma^2 \sum_{h=0}^{\infty} \left| \sum_{i=0}^{\infty} \psi_i \psi_{i+h} \right| \leq \sigma^2 \sum_{h=0}^{\infty} \sum_{i=0}^{\infty} (|\psi_i| \times |\psi_{i+h}|) \leq \sigma^2 \left( \sum_{i=0}^{\infty} |\psi_i| \right)^2 < \infty.$$

**Лемма 4.5** Пусть  $\{\epsilon_t, \mathfrak{F}_t\}_{-\infty}^{\infty}$  является мартингал-разностью,  $\mathbb{E}[\epsilon_t] = 0$ ,  $\mathbb{E}[\epsilon_t^2] = \sigma^2 < \infty$ , и положим  $x_t := \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ . Если числа  $\{\psi_i\}$  абсолютно суммируемы, то процесс  $x_t$  эргодичен для среднего:  $1/n \sum_{t=1}^n x_t \xrightarrow{p} \mathbb{E}[x_t]$ .

#### 4.3 Почти эпохальная зависимость

Мартингалы обладают слишком сильной персистентностью для применения ЗБЧ или ЦПТ в общем случае (без серьезных ограничений), а мартингал-разности структурно слишком бедны для моделирования памяти, встречающейся в экономических и финансовых временных рядах. Тем не менее, многие значительно персистентные и неоднородные процессы можно приблизить мартингал-разностями, что делает последние важнейшими компонентами ЗБЧ и ЦПТ.

Одним из решений проблемы является наложение ограничений типа перемешивания, но их очень сложно проверить, и это можно сделать только при достаточно абстрактных или ограничительных условиях (например, положительная и непрерывная на всем своем носителе плотность  $f$ ).

Свойство *почти эпохальной зависимости* (ПЭЗ), напротив, тривиально содержит перемешивание, и характеризует огромное количество моделей временных рядов. В частности, это свойство не требует явно положительности и непрерывности плотности, и формулируется в терминах математических ожиданий, что делает его не столь трудным для проверки на практике.

#### 4.3.1 Определение почти эпохальной зависимости

Пусть  $\{F_t\}$  есть последовательность  $\sigma$ -алгебр, порожденная последовательностью случайных величин  $\{v_t\}$ ,

$$F_t := \sigma(v_\tau : \tau \leq t).$$

Отметим, что  $v_t$  может быть вектором (например, шоки в модели векторной авторегрессии VAR).

Говорят, что  $\{x_t\}$  обладает  $L_p$ -ПЭЗ-свойством на  $\{F_t\}$  (или  $\{v_t\}$ ) размера  $\lambda > 0$ , если

$$\|x_t - \mathbb{E}[x_t | F_{t-m}^{t+m}]\|_p \leq d_t \times \varphi_m,$$

где  $d_t > 0$  и  $\varphi_m = o(m^{-\lambda})$ . Константы  $\{d_t\}$  позволяют вычлнить тренды моментов, и возможно стремление  $d_t \rightarrow \infty$  при  $t \rightarrow \infty$  (например, гетероскедастичность типа линейного тренда  $\mathbb{E}[x_t^2] = a + bt$ ,  $a, b > 0$ ). Коэффициенты позволяют моделировать персистентность. Случайная величина  $\{v_t\}$  называется *базой* ПЭЗ.

Двусторонность  $F_{t-m}^{t+m}$  позволяет включить двусторонние временные ряды, такие как линейные лаги, распределенные в обе стороны:

$$x_t = \sum_{i=-\infty}^{\infty} \psi_i \epsilon_{t-i}. \quad (5)$$

Вообще говоря, двусторонние ряды редко встречаются в экономических задачах; в этом случае мы заменяем  $F_{t-m}^t$  на  $F_{t-m}^{t+m}$ .

Словесно,  $\{x_t\}$  обладает  $L_p$ -ПЭЗ-свойством на  $\{F_t\}$ , когда последовательность случайных величин  $\{v_\tau\}_{\tau=t-m}^{t+m}$  можно использовать для предсказания  $x_t$  с нулевой ошибкой в  $L_p$ -норме при  $m \rightarrow \infty$ . Как и в случае перемешивания,  $\{x_t\}$  обладает  $L_p$ -ПЭЗ-свойством на  $\{F_t\}$  размера 2, если, например,  $\forall \delta \leq 2$

$$m^\delta \times \|x_t - \mathbb{E}[x_t | F_{t-m}^{t+m}]\|_p \rightarrow 0 \text{ при } m \rightarrow \infty.$$

#### 4.3.2 Примеры почти эпохальной зависимости

Очевидно, iid-последовательность является ПЭЗ на себе: если  $\mathfrak{F}_t := \sigma(x_\tau : \tau \leq t)$ , то  $\mathbb{E}[x_t | \mathfrak{F}_{t-m}^{t+m}] = \mathbb{E}[x_t | \mathfrak{F}_{t-m}^t] = x_t \forall m \geq 1$ , и значит  $\varphi_m = 0 \forall m \geq 1$ .

ПЭЗ-свойство обобщает перемешивание, так как процесс  $v_t$  может быть любым. Если  $x_t = v_t$  сильно перемешивающий, то  $x_t$ , очевидно, ПЭЗ на себе (то есть на  $\{F_t\} = \{\mathfrak{F}_t\}$ ) с  $\varphi_m = 0 \forall m \geq 1$ .

Проверка ПЭЗ-свойства обычно очень проста. Не все слабостационарные и эргодичные процессы с линейными распределенными лагами сильно перемешивающие (см., например, Andrews 1984, Guegan & Ladoucette 2001), но все обладают ПЭЗ-свойством.

**Лемма 4.6** Пусть процесс  $\epsilon_t$  имеет нулевое среднее и равномерно  $L_p$ -ограничен. Положим  $x_t := \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ , где  $\sum_{i=0}^{\infty} |\psi_i| < \infty$ . Тогда  $\{x_t\}$  обладает  $L_p$ -ПЭЗ-свойством на  $\{\epsilon_t\}$ . Размер равен  $\lambda > 0$ , если  $\psi_i = O(i^{-1-\lambda-\iota})$  для бесконечно малого  $\iota > 0$ .

Любой временной ряд, представимый в виде распределенных лагов бесконечного порядка, строящийся на  $L_p$ -ограниченных центрированных инновациях, и с абсолютно суммируемыми коэффициентами, стало быть, обладает  $L_p$ -ПЭЗ-свойством. Этот класс процессов включает все стационарные эргодичные процессы  $AR(p)$ ,  $MA(q)$  и  $ARMA(p, q)$  для конечных  $p, q$ .

### 4.3.3 Свойства почти эпохальной зависимости

Ясно, что не любая функция ПЭЗ-случайных величин обязательно обладает ПЭЗ-свойством. Сравним это наблюдение с тем, что мы знаем про перемешивание: измеримые функции конечного числа лагов перемешивающих случайных величин сами всегда перемешивающие (лемма 4.4). Тем не менее, некоторые простые свойства облегчают получение асимптотических распределений выборочных средних и выборочных ковариаций ПЭЗ-векторов.

**Лемма 4.7 (Линейные комбинации сохраняют ПЭЗ)** *Если  $\{x_{i,t} : i = 1, \dots, k\}$  являются  $L_p$ -ограниченными и  $L_p$ -ПЭЗ на некотором процессе  $\{F_t\}$  с константами  $\{d_{i,t}\}$  и коэффициентами  $\{\varphi_{i,m}\}$  размера  $\{\lambda_i\}$ , то  $\sum_{i=1}^k \psi_i x_{i,t}$  является  $L_p$ -ограниченным и  $L_p$ -ПЭЗ на  $\{F_t\}$  с константами  $d_t := \sum_{i=1}^k |\psi_i| d_{i,t}$  и коэффициентами  $\varphi_m := \sum_{i=1}^k \varphi_{i,m}$  размера  $\min_{1 \leq i \leq k} \{\lambda_i\}$ .*

**Лемма 4.8 (Произведения сохраняют ПЭЗ)** *Если  $\{x_t, y_t\}$  являются  $L_p$ -ограниченными и  $L_p$ -ПЭЗ на некотором  $\{F_t\}$  с константами  $\{d_{x,t}, d_{y,t}\}$  и коэффициентами  $\{\varphi_{x,m}, \varphi_{y,m}\}$  размера  $\{\lambda_x, \lambda_y\}$ , то  $x_t \times y_t$  является  $L_{p/2}$ -ПЭЗ на  $\{F_t\}$  с константами  $d_t := d_{x,t} d_{y,t}$  и коэффициентами  $\varphi_m := \varphi_{x,m} \varphi_{y,m}$  размера  $\min_{1 \leq i \leq k} \{\lambda_i\}$ .*

## 5 Законы больших чисел

Пусть случайный процесс  $\{x_t\}$  имеет среднее  $\mu_t := \mathbb{E}(x_t)$ , дисперсию  $\sigma_t^2 := \mathbb{E}(x_t - \mu_t)^2$  и автоковариации  $\gamma(|s - t|) = \mathbb{E}(x_s - \mu_s)(x_t - \mu_t)$ . Мы неявно позволяем процессу быть нестационарным (например иметь тренд в среднем или быть гетероскедастичным), но для ясности предполагаем, что автоковариации зависят только от смещений  $h = |s - t|$ . Например, если  $x_t$  есть просто зашумленный линейный тренд

$$x_t = \beta_0 + \beta_1 t + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} (0, \sigma^2),$$

то

$$\begin{aligned} \mu_t &= \beta_0 + \beta_1 t, \\ \sigma_t^2 &= \sigma^2, \\ \mathbb{E}(x_s - \mu_s)(x_t - \mu_t) &= 0 \quad \forall s \neq t. \end{aligned}$$

Разложим выборочную дисперсию следующим образом:

$$\mathbb{E} \left( \frac{1}{n} \sum_{t=1}^n (x_t - \mu_t) \right)^2 = \frac{1}{n^2} \sum_{t=1}^n \sigma_t^2 + \frac{2}{n} \sum_{i=1}^{n-1} \left( 1 - \frac{i}{n} \right) \gamma(i). \quad (6)$$

Если мы хотим использовать неравенство Чебышева

$$\mathbb{P} \left\{ \left| \frac{1}{n} \sum_{t=1}^n (x_t - \mu_t) \right| > \varepsilon \right\} \leq \varepsilon^{-2} \mathbb{E} \left( \frac{1}{n} \sum_{t=1}^n (x_t - \mu_t) \right)^2$$

для доказательства ЗБЧ, мы должны убедиться в том, что

$$\frac{1}{n^2} \sum_{t=1}^n \sigma_t^2 + \frac{2}{n} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \gamma(i) \rightarrow 0. \quad (7)$$

Безусловно, существуют и другие методы доказательства различных ЗБЧ. Однако для краткости, а также для того, чтобы интуитивно понять, что говорят различные типы зависимости о (7), мы сфокусируем внимание на предложенном нехитром приеме.

Случай мартингал-разностей наиболее прост и, конечно, включает в себя iid-случай.

## 5.1 Мартингал-разности

**Теорема 5.1** Если  $\{x_t, \mathfrak{F}_t\}$  есть мартингал-разность с нулевым средним и дисперсией  $\{\sigma_t^2\}$ ,  $1/n \sum_{t=1}^n \sigma_t^2 = o(n)$ , то  $1/n \sum_{t=1}^n x_t \xrightarrow{p} 0$ .

*Замечание 1.* Если дисперсии равномерно ограничены,  $\sup_{t \in \mathbb{N}} \sigma_t^2 \leq K < \infty$ , то ясно, что  $1/n \sum_{t=1}^n \sigma_t^2 \leq K$ , и значит  $1/n^2 \sum_{t=1}^n \sigma_t^2 = o(1)$  в силу  $K/n \rightarrow 0$ . Предположение о мартингал-разности влечет  $\sum_{i=1}^{n-1} (1 - i/n) \gamma(i) = 0$ .

*Замечание 2.* Заметим, что масштабирование на  $1/n$  всего лишь простой частный случай. Для большого количества интересных нестационарных процессов необходимое масштабирование отличается от  $1/n$ .

## 5.2 Неоднородные мартингал-разности

Справедливы следующие еще более интересные результаты.

**Лемма 5.1** Если  $\{S_n, \mathfrak{F}_n\}$  мартингал, то для любого  $p > 1$

$$\mathbb{P} \left\{ \max_{1 \leq k \leq n} |S_k| > \varepsilon \right\} \leq \frac{\mathbb{E} |S_n|^p}{\varepsilon^p}.$$

*Замечание.* Это достаточно тонкий результат, так как из неравенства Чебышева следует лишь, что

$$\mathbb{P} \left\{ \max_{1 \leq k \leq n} |S_k| > \varepsilon \right\} \leq \frac{\mathbb{E} [\max_{1 \leq k \leq n} |S_k|]^p}{\varepsilon^p}.$$

В следующей лемме в правой части неравенства убирается операция взятия максимума, что делает результат исключительно полезным.

**Лемма 5.2** Пусть  $\{x_t\}$  есть случайный процесс и  $S_n = \sum_{t=1}^n x_t$ . Предположим, что существует последовательность положительных чисел  $\{c_t\}$ , такая что для некоторого  $p > 0$ , каждого  $t \geq 0$  и  $n > t$ , и каждого  $\varepsilon > 0$

$$\mathbb{P} \left\{ \max_{m < j \leq n} |S_j - S_m| > \varepsilon \right\} \leq \frac{K}{\varepsilon^2} \sum_{t=m+1}^n c_t^p.$$

Если  $\sum_{t=1}^{\infty} c_t^p < \infty$ , то  $S_n \xrightarrow{a.s.} S$  (последовательность сходится).

**Теорема 5.2** Если  $\{x_t, \mathfrak{F}_t\}$  есть мартингал-разность с нулевым средним и последовательностью дисперсий  $\{\sigma_t^2\}$ , удовлетворяющей  $\sum_{t=1}^{\infty} \sigma_t^2 / a_t^2 < \infty$  для последовательности положительных чисел  $\{a_t\}$ ,  $a_t \uparrow \infty$ , то  $1/a_n \sum_{t=1}^n x_t \xrightarrow{p} 0$ .

Легко привести простые примеры. Если  $\{x_t\}$  — слабостационарная мартингал-разность, то  $\sigma^2 = \sigma_t^2 < \infty$ , и  $\sum_{t=1}^{\infty} 1/t^2 < \infty$  выполнено тривиально, и значит  $1/n \sum_{t=1}^n x_t \xrightarrow{P} 0$ . Вообще, если  $\sup_{t \in \mathbb{N}} \sigma_t^2 \leq K < \infty$ , то  $\sum_{t=1}^{\infty} \sigma_t^2/t^2 \leq K \sum_{t=1}^{\infty} 1/t^2 < \infty$ .

Предположение о конечных дисперсиях можно отбросить.

**Теорема 5.3** Если  $\{x_t, \mathfrak{F}_t\}$  имеет нулевое среднее и удовлетворяет  $\sum_{t=1}^{\infty} \mathbb{E}|x_t|^p/a_t < \infty$  для некоторой последовательности положительных чисел  $\{a_t\}$ ,  $a_t \uparrow \infty$ , и  $1 \leq p \leq 2$ , то  $1/a_n \sum_{t=1}^n x_t \xrightarrow{a.s.} 0$ .

### 5.3 Процессы с сильным перемешиванием

Предположим, что дисперсии  $\sigma_t^2$  и ковариации  $\gamma(t, t-h)$  равномерно ограничены,

$$B := \sup_{t \in \mathbb{N}} \sigma_t^2 < K \text{ и } B_h := \sup_{t \in \mathbb{N}} |\gamma(t, t-h)| < K,$$

и запишем

$$\mathbb{E} \left( \frac{1}{n} \sum_{t=1}^n (x_t - \mu_t) \right)^2 = \frac{1}{n^2} \sum_{t=1}^n \sigma_t^2 + \frac{1}{n^2} \sum_{i=1}^{n-1} (n-i) \gamma(i) \leq \frac{B}{n} + \frac{1}{n} \sum_{i=1}^n B_i.$$

**Лемма 5.3** Если  $\{x_t\}$  равномерно  $L_2$ -ограничен и  $\sum_{i=1}^{\infty} i^{-1} B_i < \infty$ , то  $1/n \sum_{t=1}^n x_t \xrightarrow{P} 0$ .

Приложения леммы 5.3 очевидны. Необходимо лишь установить ограниченную память в том смысле, что  $\sum_{i=1}^{\infty} i^{-1} B_i < \infty$ . Хотя это условие не вполне интуитивно, легко проверить, что оно выполняется для обширного класса случайных процессов. Рассмотрим следующий простой результат.

**Теорема 5.4** Пусть  $\{x_t\}$  имеет нулевое среднее, равномерно  $L_{2+\delta}$ -ограничен, и сильно перемешивающий с коэффициентами  $\sum_{m=1}^{\infty} m^{-1} \alpha_m^{\delta/(2+\delta)} < \infty$ . Тогда  $1/n \sum_{t=1}^n x_t \xrightarrow{P} 0$ .

#### 5.3.1 Почти эпохально зависимые процессы

Красота ПЭЗ-случайных величин  $x_t$  состоит в том, что они являются миксингалами, а миксингалы при минимальных дополнительных предположениях имеют  $O(1/n)$ -ограниченные выборочные дисперсии  $\mathbb{E}(1/n \sum_{t=1}^n (x_t - \mu_t))^2$ . Иными словами, при подходящих условиях

$$\mathbb{E} \left( \frac{1}{n} \sum_{t=1}^n (x_t - \mu_t) \right)^2 \rightarrow 0,$$

и установления неравенства Чебышева достаточно для справедливости ЗБЧ. Формальное определение миксингала см. в работах McLeish (1974, 1975) и Davidson (1994). Следующие рассуждения частично адаптированы из работы de Jong (1997).

Необходимо ввести понятие двойного (или стохастического) массива  $\{x_{n,t}\}$ . Далее по тексту  $\{x_{n,t}\}$  нужно представлять себе как треугольный массив:  $\{\{x_{n,t}\}_{t=1}^n\}_{n=1}^{\infty}$ . Классический пример такого массива:

$$x_{n,t} := \frac{1}{\sqrt{n}} \left( \frac{x_t - \mathbb{E}[x_t]}{\sqrt{\mathbb{E}(x_t - \mathbb{E}[x_t])^2}} \right), \text{ где } x_t \sim iid.$$

Определим  $\sigma$ -алгебру, порожденную базой ПЭЗ:  $F_t := \sigma(\epsilon_\tau : \tau \leq t)$ .



**Теорема 5.5** *Предположим следующее:*

- a.  $\{x_{n,t}\}$  является центрированным,  $L_r$ -ограниченным,  $r > 2$ , стохастическим массивом.
- b.  $\{x_{n,t}\}$  является  $L_2$ -ПЭЗ размера  $1/2$  с константами  $\{d_{n,t}\}$  и коэффициентами  $\{\varphi_m\}$ . База  $\epsilon_t$  равномерно перемешивающая размера  $r/[2(r-1)]$ , или сильно перемешивающая размера  $r/(r-2)$ .
- c. Константы  $\{d_{n,t}\}$  удовлетворяют

$$\sum_{t=1}^n (\max\{\|x_{n,t}\|_r, d_{n,t}\})^2 = O(1).$$

Тогда существует последовательность положительных постоянных  $\{a_n\}$ ,  $a_n \nearrow \infty$ , удовлетворяющих  $1/a_n \sum_{t=1}^n x_{n,t} \xrightarrow{P} 0$ .

Случай строгой стационарности формулируется в терминах  $\{x_t\}$  вместо массива  $\{x_{n,t}\}$ , что позволяет избежать громоздких обозначений.

**Следствие 5.1** *Пусть  $\{x_t\}$  строго стационарен, имеет среднее  $\mathbb{E}[x_t]$ ,  $L_r$ -ограничен для некоторого  $r > 2$ . Также предположим, что  $\{x_t\}$  обладает  $L_2$ -ПЭЗ-свойством размера  $1/2$  с константами  $d_t = d$  и коэффициентами  $\varphi_m$  и ПЭЗ-базой  $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$ . Тогда  $\{x_t\}$  эргодичен для среднего:  $1/n \sum_{t=1}^n x_t \xrightarrow{P} \mathbb{E}[x_t]$ .*

## 6 Центральные предельные теоремы

В настоящем разделе мы преследуем цель получить достаточно ограничительные условия, которые гарантировали бы, что некоторый стандартизованный треугольный массив  $\{x_{n,t}\}$ , удовлетворяющий  $\|\sum_{t=1}^n x_{n,t}\|_2 = 1$ , подчинялся бы

$$\sum_{t=1}^n x_{n,t} \xrightarrow{d} N(0, 1).$$

### 6.1 Мартингал-разности

Несмотря на то, что мартингал-разности обладают тривиальной структурой, в отличие от ЗБЧ для мартингал-разностей необходимо будет наложить некоторые условия, касающиеся неоднородности. Следующий результат, представленный в работе Davidson (1994, теорема 24.3), использует более общее утверждение, доказанное в работе McLeish (1974).

**Теорема 6.1** *Пусть  $\{x_{n,t}, \mathfrak{F}_{n,t}\}$  есть последовательность мартингал-разностей с конечной дисперсией  $\{\sigma_{n,t}^2\}$ , и  $\sum_{t=1}^n \sigma_{n,t}^2 = 1$ . Если  $\sum_{t=1}^n x_{n,t}^2 \xrightarrow{P} 1$  и  $\max_{1 \leq t \leq n} |x_{n,t}| \xrightarrow{P} 0$ , то  $\sum_{t=1}^n x_{n,t} \xrightarrow{d} N(0, 1)$ .*

Рассмотрим стандартный случай

$$x_{n,t} = \frac{1}{\sqrt{n}} \left( \frac{x_t - \mathbb{E}[x_t]}{\sqrt{\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n (x_t - \mathbb{E}[x_t])^2)}} \right)$$

где  $\{x_t, \mathfrak{F}_t\}$  есть мартингал-разность с дисперсиями  $\{\sigma_t^2\}$ ,  $\sigma_t^2 > 0$  равномерно по  $t$ . Заметим, что  $\mathfrak{S}_{n,t} = \mathfrak{F}_t$ , так как неслучайное  $n$  не рафинирует  $\sigma$ -алгебру. Предположение о мартингал-разности влечет

$$\mathbb{E} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n (x_t - \mathbb{E}[x_t]) \right)^2 = \frac{1}{n} \sum_{t=1}^n \sigma_t^2,$$

и значит

$$\sigma_{n,t}^2 = \frac{1}{n} \frac{\sigma_t^2}{\frac{1}{n} \sum_{t=1}^n \sigma_t^2} = \frac{\sigma_t^2}{\sum_{t=1}^n \sigma_t^2},$$

поэтому, очевидно,  $\sum_{t=1}^n \sigma_{n,t}^2 = 1$ .

Условие  $\max_{1 \leq t \leq n} |x_{n,t}| \xrightarrow{p} 0$  влечет

$$\max_{1 \leq t \leq n} \left| \frac{x_t - \mathbb{E}[x_t]}{\sqrt{\sum_{t=1}^n \sigma_t^2}} \right| \xrightarrow{p} 0,$$

что почти тривиально, так как  $\sum_{t=1}^n \sigma_t^2 \rightarrow \infty$  из-за того, что  $\inf_t \sigma_t^2 > 0$ . Необходимо лишь предположить, что ни одно  $|x_t - \mathbb{E}[x_t]|$  не доминирует над всеми остальными.

Наконец, используем свойство мартингал-разности, чтобы получить

$$\sum_{t=1}^n x_{n,t}^2 = \frac{\sum_{t=1}^n (x_t - \mathbb{E}[x_t])^2}{\mathbb{E}(\sum_{t=1}^n (x_t - \mathbb{E}[x_t])^2)} = \frac{\sum_{t=1}^n (x_t - \mathbb{E}[x_t])^2}{\sum_{t=1}^n \mathbb{E}(x_t - \mathbb{E}[x_t])^2}.$$

Условие  $\sum_{t=1}^n x_{n,t}^2 \xrightarrow{p} 1$  сводится к тому, чтобы процесс  $\sum_{t=1}^n (x_t - \mathbb{E}[x_t])^2$  удовлетворял ЗБЧ. Достаточными условиями этого являются:  $\{(x_t - \mathbb{E}[x_t])^2 - \sigma_t^2, \mathfrak{F}_t\}$  есть мартингал-разность, где  $\mathbb{E}((x_t - \mathbb{E}[x_t])^2 - \sigma_t^2)^2$  постоянно равномерно по  $t$ .

Условие  $\max_{1 \leq t \leq n} |x_{n,t}| \xrightarrow{p} 0$  является немного избыточным, хотя само по себе достаточно слабо. Следующий результат представлен в работе White (1984, следствие 5.25). Идея в том, что даже если  $\{x_t, \mathfrak{F}_t\}$  неоднороден (например, не обладает постоянными дисперсиями), пока неоднородность не носит взрывного характера (то есть ни одно наблюдение не доминирует над всеми остальными), центральная предельная теорема остается верной.

**Теорема 6.2** Пусть  $\{x_t, \mathfrak{F}_t\}$  есть мартингал-разность с конечными дисперсиями  $\{\sigma_t^2\}$ , и  $1/n \sum_{t=1}^n \sigma_t^2 \rightarrow \sigma^2$ ,  $\sup_{t \in \mathbb{Z}} \mathbb{E}|x_t|^r < \infty$  для некоторого  $r > 2$  и  $1/n \sum_{t=1}^n x_t^2 \xrightarrow{p} \sigma^2$ , тогда  $1/\sqrt{n} \sum_{t=1}^n x_t \xrightarrow{d} N(0, \sigma^2)$ .

## 6.2 Почти эпохально зависимые массивы

Следующий результат представляет собой несколько усеченную версию теоремы 2 из работы de Jong (1997).

**Теорема 6.3** Пусть  $\{x_{n,t}\}$  есть треугольный массив с нулевым средним, и предположим следующее:

- Существует последовательность положительных чисел  $\{c_{n,t}\}$ , такая что  $\{x_{n,t}/c_{n,t}\}$  является  $L_r$ -ограниченным для некоторого  $r > 2$ , равномерно по  $t$  и  $n$ .
- $\{x_{n,t}\}$  обладает  $L_2$ -ПЭЗ-свойством размера  $1/2$  с константами  $\{d_{n,t}\}$  и коэффициентами  $\{\varphi_m\}$ . База  $\epsilon_t$  равномерно перемешивающая размера  $r/[2(r-1)]$  или сильно

перемешивающая размера  $r/(r - 2)$ .

с. Числа  $\{c_{n,t}\}$  удовлетворяют  $\sum_{t=1}^n c_{n,t}^2 = O(1)$ .

Тогда  $\sum_{t=1}^n x_{n,t} \xrightarrow{d} N(0, 1)$ .

*Замечание.* Так как в теореме 6.3 требуется лишь, чтобы  $\{x_{n,t}\}$  обладал ПЭЗ-свойством на равномерно или сильно перемешивающем процессе, перемешивающие процессы соответствующего размера, очевидно, удовлетворяют сформулированным условиям.

Рассмотрим упрощенное следствие.

**Следствие 6.1** Пусть  $\{x_t\}$  слабо стационарен и  $L_r$ -ограничен для некоторого  $r > 2$ . Предположим также, что  $\{x_t\}$  обладает  $L_2$ -ПЭЗ-свойством размера  $1/2$  с константами  $d_t = d$  и коэффициентами  $\varphi_m$  и ПЭЗ-базой  $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$ . Наконец, пусть

$$\liminf_{n \geq 1} \mathbb{E} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\} \right)^2 > 0.$$

Запишем

$$x_{n,t} = \frac{1}{\sqrt{n}} \left( \frac{x_t - \mathbb{E}[x_t]}{\sqrt{\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\}^2)}} \right).$$

Тогда  $\sum_{t=1}^n x_{n,t} \xrightarrow{d} N(0, 1)$  и  $\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\}^2) = O(1)$ .

*Замечание.* Заметим, что мы предположили невырожденность стандартного отклонения  $\liminf_{n \geq 1} \mathbb{E}(1/\sqrt{n} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\}^2) > 0$ , а сформулированный результат гарантирует, что оно также ограничено:  $\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\}^2) \leq K$ . Следовательно,  $\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\}^2) \rightarrow v > 0$ , где  $v < \infty$ .

### 6.3 Метод Крамера–Волда

Обобщение вышеприведенных ЦПТ на многомерный случай  $x_t \in \mathbb{R}^k$  как правило не сложнее, чем доказательство ЦПТ для линейной комбинации  $\sum_{i=1}^k r_i x_{i,t}$ .

**Теорема 6.4 (Метод Крамера–Волда)** Если  $\{x_{n,t}\}$  есть  $\mathbb{R}^k$ -значный стохастический массив, и  $\forall r \in \mathbb{R}^k$

$$\sum_{t=1}^n r' x_{n,t} \xrightarrow{d} N(0, r' r),$$

то  $\sum_{t=1}^n x_{n,t} \xrightarrow{d} N(0, I_k)$ .

**Следствие 6.2** Если  $\{x_t\}$  суть  $\mathbb{R}^k$ -значный случайный процесс  $\forall r \in \mathbb{R}^k$ , и

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n r' x_t \xrightarrow{d} N(0, r' \Sigma r),$$

то  $1/\sqrt{n} \sum_{t=1}^n x_t \xrightarrow{d} N(0, \Sigma)$ , где  $\Sigma = \lim_{n \rightarrow \infty} \mathbb{E}(1/\sqrt{n} \sum_{t=1}^n x_t)^2$ .

Взятие линейных комбинаций сохраняет и свойство перемешиваемости, и ПЭЗ-свойство. В свете этого теорема 6.4 и следствие 6.2 крайне полезны для получения предельных распределений многомерных оценок (например квази-ММП, МНК, ОММ).

## 7 Теория слабых пределов и броуновское движение

В настоящем разделе мы представим концепции функциональных пределов, полезные для характеристики предельных распределений оценок, связанных с процессами с единичными корнями. Мы демонстрируем главные результаты на примере вывода предельного распределения МНК-оценки коэффициента наклона в модели AR(1) с единичным корнем.

### 7.1 Случайное блуждание без сноса: пример

Рассмотрим случайное блуждание без сноса:

$$x_t = x_{t-1} + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} (0, \sigma^2).$$

Рассмотрим также МНК-оценку  $\hat{\rho}$  коэффициента наклона  $\rho$  в модели AR(1)

$$x_t = \rho x_{t-1} + \epsilon_t, \quad \text{где } \rho = 1.$$

Тогда

$$\hat{\rho} = \frac{\sum_{t=2}^n x_t x_{t-1}}{\sum_{t=2}^n x_{t-1}^2} = 1 + \frac{\sum_{t=2}^n \epsilon_t x_{t-1}}{\sum_{t=2}^n x_{t-1}^2}.$$

Таким образом,  $\hat{\rho} \xrightarrow{p} 1$ , если правильным образом масштабированные суммы  $\sum_{t=2}^n \epsilon_t x_{t-1} \xrightarrow{p} 0$ . Но учитывая что

$$x_t = \sum_{i=0}^{\infty} \epsilon_{t-i}$$

обладает бесконечной дисперсией

$$\mathbb{E}[x_t^2] = \sum_{i=0}^{\infty} \sigma^2 = \infty, \quad \text{поскольку } \epsilon_t \stackrel{iid}{\sim} (0, \sigma^2),$$

каковы пределы по вероятности и по распределению процесса  $\sum_{t=2}^n \epsilon_t x_{t-1}$ ? Принимая во внимание то, что  $\mathbb{E}[x_t^2] = \infty$ , можем ли мы использовать множитель  $1/n$  в обеих компонентах отношения (например  $1/n \sum_{t=2}^n x_{t-1}^2$ )? Быстрые ответы таковы: *вероятно негауссовские; нет.*

Мы получим предельное распределение оценки  $\hat{\rho}$  последовательно, сначала рассмотрев функциональную центральную предельную теорию, а потом решая эту задачу для  $\hat{\rho}$  непосредственно.

### 7.2 Функциональная предельная теория

Как и все термины в настоящем эссе, этот термин имеет свою аббревиатуру. Изложение будет касаться только тех результатов, которые нужны для определения предела по распределению оценки  $\hat{\rho}$  раздела 7.1.

Пусть  $\{x_t\}$  — произвольный случайный процесс с нулевым средним. Если  $x_t \stackrel{iid}{\sim} (0, \sigma^2)$ , то (например, по ЦПТ Линдберга–Леви)

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n x_t \xrightarrow{d} N(0, \sigma^2).$$

Здесь  $n$  совершенно произвольно, пока предел берется при  $n \rightarrow \infty$ . Результат остается верным для любой доли  $[\lambda n]$  от  $n$ , где  $\lambda \in [0, 1]$ , а  $[z]$  обозначает целую часть числа  $z$ . Простое масштабирование и тот факт, что  $[\lambda n]/\lambda n \rightarrow 1$ , позволяют написать

$$Z_n(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[\lambda n]} x_t = \frac{\sqrt{[\lambda n]}}{\sqrt{n}} \times \frac{1}{\sqrt{[\lambda n]}} \sum_{t=1}^{[\lambda n]} x_t \xrightarrow{d} N(0, \lambda) = Z(\lambda),$$

скажем, где предельное распределение  $Z(\lambda)$  является нормальным с нулевым средним и дисперсией  $\lambda$ . Таким образом, «функция»  $Z_n(\lambda)$ , вычисленная в точке  $\lambda \in [0, 1]$  сходится к нормальному закону с нулевым средним  $Z(\lambda)$ , который сам является функцией  $\lambda$ . Следовательно,

$$Z_n(\lambda) \xrightarrow{d} Z(\lambda) \text{ поточечно на } [0, 1].$$

Так как гауссовский закон полностью характеризуется средним и дисперсией (Doob 1953),  $\lambda$  полностью характеризует это гауссовское распределение с нулевым средним.

Теперь для каждого  $n$  мы считаем

$$\{Z_n(\lambda)\} = \{Z_n(\lambda) : 0 \leq \lambda \leq 1\}$$

случайным процессом, индексированным параметром  $\lambda$ . Каковы свойства этого процесса? Каковы его траектории (то есть как процесс  $Z_n(\lambda)$  выглядит для каждого  $\lambda$  для конкретной реализации  $\{x_t\}_{t=1}^n$ )? К чему сходится процесс  $\{Z_n(\lambda)\}$ , и сходится ли вообще?

Если процесс  $\{Z_n(\lambda)\}$  имеет предел, скажем,  $\{Z(\lambda)\}$ , то это *слабый предел*<sup>1</sup>, или *функциональный предел*. Мы обозначаем тот факт, что  $\{Z(\lambda)\}$  является слабым пределом  $\{Z_n(\lambda)\}$ , символами

$$Z_n(\lambda) \implies Z(\lambda).$$

Заметим, что здесь есть существенные отличия от обычной поточечной сходимости:  $Z_n(\lambda) \xrightarrow{d} Z(\lambda)$  для каждого  $\lambda \in [0, 1]$ . К сожалению, можно привести примеры, когда поточечная сходимость выполняется, но процесс  $\{Z_n(\lambda)\}$ , являющийся случайным выбором из распределения траекторий, не сходится.

Если  $Z(\lambda)$  имеет определенные свойства, то говорят, что  $Z_n(\lambda)$  сходится в пространстве с этими свойствами. Например, если  $Z(\lambda)$  — непрерывная функция  $\lambda$ , то  $Z_n(\lambda)$  сходится на  $C[0, 1]$ , пространстве непрерывных вещественнозначных функций  $[0, 1]$ .

Оказывается, для iid, не-iid и вырожденных функций не-iid случайных процессов  $\{x_t\}$  процесс частичных сумм  $\{Z_n(\lambda)\}$  сходится к предельному процессу  $\{Z(\lambda)\}$ , который идентичен *броуновскому движению*. Это верно для мартингал-разностей, перемешивающих и ПЭЗ-процессов  $\{x_t\}$  (Davidson 1994, de Jong & Davidson 2000, Hill 2009).

Мы рассмотрим только iid-случай, но читатель может посмотреть формулировки аналогичных результатов для мартингал-разностей в соответствующей литературе. Дадим сначала определения броуновского движения и меры Винера (Wiener 1923).

**Броуновское движение (винеровский процесс)** *Броуновским движением или винеровским процессом  $\{Z(\lambda)\} = \{Z(\lambda) : 0 \leq \lambda \leq 1\}$  называется непрерывный процесс на  $[0, 1]$ , такой что*

$$(i) Z(0) = 0 \text{ почти наверное;}$$

<sup>1</sup>Важно не путать это слово «слабый» с состоятельностью, которую часто называют «слабым пределом по вероятности», или слабым пределом.

(ii)  $[Z(\lambda_2) - Z(\lambda_1), \dots, Z(\lambda_k) - Z(\lambda_{k-1})]$  для любого  $k \in \mathbb{N}$  имеет многомерное нормальное распределение с диагональной ковариационной матрицей для любых  $0 \leq \lambda_1 < \lambda_2 < \dots < \lambda_k \leq 1$  и гауссовскими приращениями  $Z(\lambda) - Z(\lambda') \sim N(0, \lambda - \lambda')$  для любых  $\lambda > \lambda'$ ;

(iii)  $Z(\lambda)$  непрерывен по  $\lambda$  почти наверное.

*Замечание 1.* Так как  $Z(\lambda)$  почти наверное непрерывная функция на отрезке  $[0, 1]$ , ее траектории принадлежат пространству  $C[0, 1]$  с вероятностью единица. Траекторией называется реализация  $\{Z(\lambda, \omega) : 0 \leq \lambda \leq 1\}$ , где  $\omega$  — случайный исход из  $\Omega$ . Реализация броуновского движения изображена на Рис. 1.

*Замечание 2.* Многомерное нормальное распределение имеет диагональную ковариационную матрицу тогда и только тогда, когда его скалярные компоненты независимы: приращение  $Z(\lambda_i) - Z(\lambda_{i-1})$  независимо от приращения  $Z(\lambda_j) - Z(\lambda_{j-1}) \forall \lambda_{i-1} < \lambda_i < \lambda_{j-1} < \lambda_j$ .

*Замечание 3.* Ограничение  $\lambda \in [0, 1]$  — условность, которой мы пользуемся ниже. Вообще говоря,  $\lambda$  может принадлежать любому компакту  $A \subset \mathbb{R}$ , или  $[0, \infty)$ , и соответствующий процесс существует на  $C[A]$  или  $C[0, \infty)$  соответственно. См. Billingsley (1999).

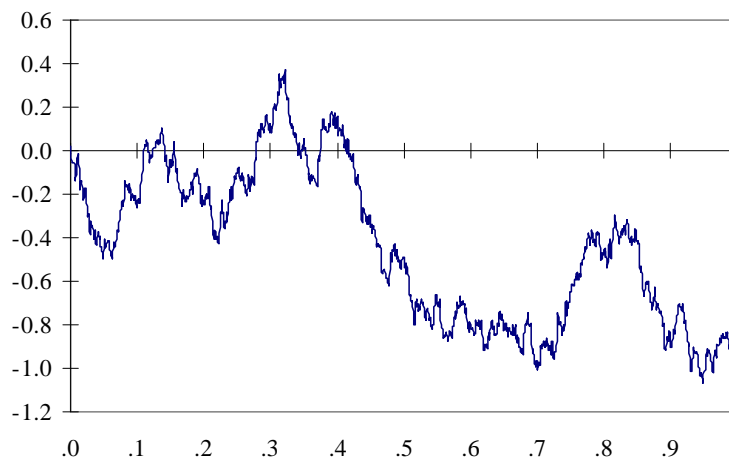


Рис. 1: Реализация броуновского движения

*Замечание 4.* На самом деле, на Рис. 1 изображено лишь кусочное приближение к траектории броуновского движения, так как мы изобразили не совсем непрерывную функцию (мы соединили лишь конечное множество точек!). А именно, на рисунке изображен

$$x_t = \frac{1}{\sqrt{n}} \sum_{i=0}^{t-1} \epsilon_{t-i} \text{ для } t = 1, \dots, 1000, \text{ где } \epsilon_t \stackrel{iid}{\sim} N(0, 1),$$

что можно переписать как

$$Z_n(\lambda) := \frac{1}{\sqrt{n}} \sum_{i=0}^{[n\lambda]} \epsilon_{t-i} \text{ для } \lambda \in \left\{ \frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n} \right\} = \{0,001; 0,002; \dots, 1\}.$$

Заметим, что  $Z_n(\lambda)$  обладает всеми свойствами броуновского движения, *кроме* непрерывности: мы можем утверждать лишь, что этот процесс *непрерывен всюду за исключением счетного числа точек*  $[0, 1]$ . При гладком увеличении  $\lambda$  процесс  $Z_n(\lambda)$  претерпевает небольшие скачки, когда  $\lambda$  пересекает точки  $1/n, 2/n$  и т. д. Скачки асимптотически исчезают с вероятностью 1. Более того, несмотря на то, что (нулевые с вероятностью 1) скачки всегда асимптотически присутствуют, их всего *счетное* число. Это может казаться большим

количеством, однако счетное множество имеет нулевую меру. Асимптотически, вероятность скачка нулевая. Иными словами, траектории почти наверное непрерывны, и значит принадлежат  $C[0, 1]$  с вероятностью 1. Этого достаточно для доказательства широкого спектра асимптотических результатов, которые мы приводим и используем ниже.

Винеровской мерой называется вероятностная мера (распределение вероятностей) броуновского движения  $\{Z(\lambda)\}$ .

**Мера Винера** Мерой Винера  $W$  называется вероятностная мера на  $C[0, 1]$  со свойствами

$$(i) W(Z(0) = 0) = 1;$$

$$(ii) W(Z(\lambda) \leq a) = \int_{-\infty}^a (\sqrt{2\pi\lambda})^{-1} \exp\{0,5 Z(\lambda)^2/\lambda\};$$

(iii) если  $\{Z(\lambda)\} \sim W$ , то  $Z(\lambda_2) - Z(\lambda_1)$  независим от  $Z(\lambda_4) - Z(\lambda_3)$  для всех  $0 \leq \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4 \leq 1$ .

*Замечание 1.* Таким образом, мера Винера  $W$  есть вероятностная мера, соответствующая броуновскому движению, или винеровскому процессу.

*Замечание 2.* Легко показать, что определение броуновского движения, данное нами, избыточно: любой процесс  $\{Z(\lambda)\} \sim W$  имеет точно те же свойства, которым характеризуется броуновское движение. Однако очень трудно доказать, что винеровская мера существует (то есть процесс  $\{Z(\lambda)\}$  действительно обладает соответствующей вероятностной мерой).

Функциональная центральная предельная теорема (ФЦПТ), или теорема о слабом пределе, представленная ниже, позволяет асимптотически характеризовать траектории  $\{Z_n(\lambda)\}$ , такие как на Рис. 1, а также получить предельное распределение  $\hat{\rho}$  из раздела 7.1.

**Теорема 7.1** Пусть  $\{x_t\}$  есть iid-процесс с нулевым средним и  $\mathbb{E}[x_t^2] = \sigma^2 < \infty$ , и положим  $Z_n(\lambda) = 1/\sqrt{n} \sum_{t=1}^{\lfloor \lambda n \rfloor} x_t$ ,  $\lambda \in [0, 1]$ . Тогда

$$Z_n(\lambda) \Longrightarrow Z(\lambda),$$

где  $Z(\lambda)$  обладает почти наверное непрерывными траекториями. В частности,  $Z(\lambda)$  суть броуновское движение.

Вспомним теорему о непрерывном отображении: если некоторая последовательность  $Z_n \xrightarrow{d} Z$ , то для непрерывных функций  $g(Z_n) \xrightarrow{d} g(Z)$ . Фактически, данная теорема позволяет менять местами символы непрерывной функции и слабого предела (Billingsley 1999):

$$g(Z_n(\lambda)) \Longrightarrow g(Z(\lambda)).$$

Следующий полезный результат доказывается с помощью теоремы 7.1 и является обобщением теоремы о непрерывном отображении.

**Теорема 7.2** Рассмотрим любую функцию  $g: \mathbb{R} \rightarrow \mathbb{R}$ , которая непрерывна всюду, за исключением, быть может, счетного числа точек своего носителя. В условиях теоремы 7.1

$$\int_0^1 g(Z_n(\lambda)) d\lambda \xrightarrow{d} \int_0^1 g(Z(\lambda)) d\lambda.$$

### 7.3 Случайные блуждания со сносом: асимптотическая теория МНК

Вернемся к проблеме МНК-инференции для процесса с единичным корнем. Предположим, что время отсчитывается с момента  $t = 1$ :  $\epsilon_t = 0$  почти наверное  $\forall t \leq 0$ . Простой подстановкой убеждаемся в том, что

$$x_t = \sum_{i=0}^{t-1} \epsilon_{t-i} = \sum_{\tau=1}^t \epsilon_{\tau}.$$

Мы получили известный результат (Phillips 1987).

**Теорема 7.3** *Рассмотрим процесс случайного блуждания из раздела 7.1. МНК-оценка коэффициента наклона удовлетворяет условию*

$$n(\hat{\rho} - 1) \xrightarrow{d} \frac{1}{2} \frac{Z(\lambda)^2 - 1}{\int_0^1 Z(\lambda)^2 d\lambda}.$$

*Замечание.* Справедливы следующие два важных утверждения: во-первых, предел очевидно негауссов; во-вторых, сходимость теперь имеет скорость  $n$ . Таким образом,  $\hat{\rho} \xrightarrow{p} 1$  с необычно высокой скоростью<sup>2</sup>.

## Список литературы

- An, H.Z. & F.C. Huang (1996). The geometrical ergodicity of nonlinear autoregressive models. *Statistica Sinica* 6, 943–956.
- Andrews, D.W.K. (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21, 930–934.
- Bierens, H.J. (1981). *Robust Methods and Asymptotic Theory in Nonlinear Econometrics*. Springer-Verlag.
- Billingsley, P.J. (1999). *Convergence of Probability Measure*. Wiley: New York.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press: Oxford.
- Davidson, J. & R.M. de Jong (2000). The functional central limit theorem and weak convergence to stochastic integrals. *Econometric Theory* 16, 621–642.
- de Jong, R.M. (1997). Central limit theorems for dependent heterogeneous random variables. *Econometric Theory* 13, 353–367.
- Dedecker, J., P. Doukhan, G. Lang, J.R. Leon, S. Louhichi & C. Prieur (2007). *Weak Dependence: with Examples and Applications*. Springer.
- Doob, J. L. (1953). *Stochastic Processes*. New York: Wiley.
- Gallant, A.R. & H. White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell: Oxford.
- Guegan D. & S. Ladoucette (2001). Non-mixing properties of long memory processes. *Comptes Rendus de l'Academie des Sciences Series I Mathematics* 333, 373–376.
- Hill, J.B. (2009). On functional central limit theorems for dependent, heterogenous arrays with applications to tail index and tail dependence estimation. *Journal of Statistical Planning and Inference* 139, 2091–2110.
- Ibragimov, I.A. (1962). Some limit theorems for stationary processes *Theory of Probability and its Applications* 7, 349–382.
- Leibschner, E. (2005). Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *Journal of Time Series Analysis* 26, 669–689.

<sup>2</sup>На самом деле это неудивительно: МНК-оценки суперсостоятельны (то есть сходятся со скоростью большей, чем  $\sqrt{n}$ ) для большого количества процессов с бесконечной дисперсией. Процесс с единичным корнем представляет собой простой частный случай процесса с бесконечной дисперсией.



Lévy, P. (1925). *Calcul de Probabilités*. Gauthier-Villars: Paris.

Lévy, P. (1954). *Théorie de l'Addition des Variables Eléatoires*. Gauthier-Villars: Paris.

McLeish, D.L. (1974). Dependent central limit theorems. *Annals of Probability* 2, 620–628.

McLeish, D.L. (1975). A maximal inequality and dependent strong law. *Annals of Probability* 3, 329–339.

Phillips, P.C.B. (1987). Time series regression with a unit root. *Econometrica* 55, 277–301.

Serfling, R.J. (1968). Contributions to central limit theory for dependent variables. *Annals of Mathematical Statistics* 39, 1158–1175.

Wiener, N. (1923). Differential space. *Journal of Mathematics and Physics* 2, 131–174.

White, H. (1984). *Asymptotic Theory for Econometricians*. Academic Press: Orlando.

## Приложение: избранные доказательства

**Доказательство леммы 2.1.** См. Davidson (1994). ■

**Доказательство теоремы 2.1.** Ясно, что

$$\begin{aligned} \mathbb{E}|x_t|^{1+\delta} &= \mathbb{E}|x_t|^{1+\delta} I(|x_t| \geq M) + \mathbb{E}|x_t|^{1+\delta} I(|x_t| < M) \\ &\geq \mathbb{E}|x_t|^{1+\delta} I(|x_t| \geq M) \geq M^\delta \times \mathbb{E}|x_t| I(|x_t| \geq M). \end{aligned}$$

Следовательно,

$$\mathbb{E}|x_t| I(|x_t| \geq M) \leq \mathbb{E}|x_t|^{1+\delta} / M^\delta \rightarrow 0$$

при  $M \rightarrow \infty$  в силу  $\mathbb{E}|x_t|^{1+\delta} < \infty$  из-за  $L_{1+\delta}$ -ограниченности. ■

**Доказательство теоремы 3.1.** См. теорему 13.13 в работе Davidson (1994). ■

**Доказательство теоремы 3.3.** См. следствие 13.14 в работе Davidson (1994). ■

**Доказательство теоремы 3.4.** Используя слабую стационарность и неравенство Чебышева, можно записать, что для любого  $\varepsilon > 0$

$$\begin{aligned} \mathbb{P}\left\{\left|\frac{1}{n} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\}\right| > \varepsilon\right\} &\leq \mathbb{E}\left(\frac{1}{n} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\}\right)^2 \varepsilon^{-2} \\ &= \varepsilon^{-2} \times \frac{1}{n^2} \sum_{s,t=1}^n \mathbb{C}(x_s, x_t) \\ &= \varepsilon^{-2} \times \left(\frac{\sigma^2}{n} + \frac{1}{n} \sum_{i=1}^{n-1} (1 - i/n) \times \mathbb{C}(x_{i+1}, x_1)\right) \\ &\leq \varepsilon^{-2} \times \left(\frac{\sigma^2}{n} + \frac{1}{n} \sum_{i=1}^n |\mathbb{C}(x_{i+1}, x_1)|\right), \end{aligned}$$

так как  $1 - i/n \in [0, 1]$ . Из слабой стационарности следует, что  $\sigma^2 < \infty$ , и значит

$$1/n \sum_{i=1}^n |\mathbb{C}(x_{i+1}, x_1)| \rightarrow 0,$$

чего достаточно для

$$\mathbb{P}\left\{\left|\frac{1}{n} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\}\right| > \varepsilon\right\} \rightarrow 0.$$

Абсолютная суммируемость  $\sum_{i=0}^{\infty} |\mathbb{C}(x_{i+1}, x_1)| < \infty$  влечет  $\mathbb{C}(x_{N+1}, x_1) \rightarrow 0$  при  $N \rightarrow \infty$ . Но если  $\{\mathbb{C}(x_{i+1}, x_1)\}$  сходится к нулю, то для суммирования Чезаро верно

$$1/n \sum_{i=1}^n |\mathbb{C}(x_{i+1}, x_1)| \rightarrow 0$$

(Davidson 1994, теорема 2.26). ■

**Доказательство леммы 3.1.** Так как процесс  $x_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}$  имеет нулевое среднее, и  $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$ , автоковариации  $\gamma(t-s) := \mathbb{E}[x_s x_t]$ ,  $t \geq s$ , имеют вид

$$\sum_{i,j=0}^{\infty} \phi^i \phi^j \mathbb{E}[\epsilon_{s-i} \epsilon_{t-j}] = \sigma^2 \sum_{j=i+t-s}^{\infty} \phi^i \phi^j = \sigma^2 \sum_{i=0}^{\infty} \phi^i \phi^{i+t-s} = \sigma^2 \phi^{t-s} / (1 - \phi^2),$$

то есть зависят только от смещения  $h := t - s$ , но не конкретных значений  $s$  и  $t$ . Значит,  $\gamma(h) = \sigma^2 \phi^h / (1 - \phi^2)$ , и

$$\sum_{i=1}^{\infty} |\gamma(h)| = \frac{\sigma^2}{1 - \phi^2} \sum_{i=1}^{\infty} |\phi|^h = \frac{\sigma^2}{(1 - \phi^2)(1 - |\phi|)} < \infty.$$

Абсолютная суммируемость ковариаций влечет эргодичность для среднего, ср. с теоремой 3.4. ■

**Доказательство леммы 4.2.** См. Ibragimov (1962), а также теорему 14.2 и следствие 14.3 в работе Davidson (1994). ■

**Доказательство леммы 4.3.** См. Serfling (1968), а также теорему 14.4 и следствие 14.5 в работе Davidson (1994). ■

**Доказательство теоремы 4.1.** Геометрическое сильное перемешивание и  $L_{2+\delta}$ -ограниченность означают, что для некоторого  $K > 0$  и малого  $\iota > 0$

$$|\mathbb{C}(x_t, x_{t-m})| \leq K \alpha_m^\iota \leq K \rho^{\iota m}.$$

Следовательно,

$$\sum_{m=0}^{\infty} |\mathbb{C}(x_t, x_{t-m})| \leq K (1 - \rho^\iota)^{-1} < \infty.$$

Теперь можно применить теорему 3.4. ■

**Доказательство леммы 4.4.** См. теорему 14.1 в работе Davidson (1994). ■

**Доказательство леммы 4.5.** Необходимо применить теорему 3.4. ■

**Доказательство леммы 4.6.** Положим  $F_t := \sigma(\epsilon_\tau : \tau \leq t)$  и заметим, что

$$\mathbb{E}[x_t | F_{t-m}^{t+m}] = \mathbb{E}\left[\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} | F_{t-m}^t\right] = \sum_{i=0}^m \psi_i \epsilon_{t-i} + \sum_{i=m+1}^{\infty} \psi_i \mathbb{E}[\epsilon_{t-i} | F_{t-m}^{t+m}].$$

Применяя неравенство Минковского и условное неравенство Йенсена, получаем

$$\begin{aligned} \|x_t - \mathbb{E}[x_t | F_{t-m}^{t+m}]\|_2 &= \left\| \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} - \sum_{i=0}^m \psi_i \epsilon_{t-i} - \sum_{i=m+1}^{\infty} \psi_i \mathbb{E}[\epsilon_{t-i} | F_{t-m}^{t+m}] \right\|_p \\ &= \left\| \sum_{i=m+1}^{\infty} \psi_i (\epsilon_{t-i} - \mathbb{E}[\epsilon_{t-i} | F_{t-m}^{t+m}]) \right\|_p \\ &\leq \sum_{i=m+1}^{\infty} |\psi_i| \times \|\epsilon_{t-i}\|_p \leq \sup_{s \in \mathbb{Z}} \|\epsilon_s\|_p \times \sum_{i=m+1}^{\infty} |\psi_i| = d_t \times \varphi_m, \end{aligned}$$

скажем, где  $d_t = \sup_{s \in \mathbb{Z}} \|\epsilon_s\|_p < \infty \forall t$ . Ясно, что из  $\sum_{i=0}^{\infty} |\psi_i| < \infty$  следует  $\sum_{i=m+1}^{\infty} |\psi_i| \rightarrow 0$  при  $m \rightarrow \infty$ . Если  $\psi_i = O(i^{-1-\lambda-\iota})$ , то

$$m^\lambda \varphi_m \leq K \sum_{i=m+1}^{\infty} i^{-1-\iota} < \infty,$$

и значит  $m^\lambda \varphi_m \rightarrow 0$ , то есть ПЭЗ-размер  $\lambda > 0$ . ■

**Доказательство леммы 4.8.** См. теорему 17.9 в работе Davidson (1994). ■

**Доказательство леммы 4.7.** Используя неравенство Минковского и ПЭЗ-свойство, получаем

$$\begin{aligned} \left\| \sum_{i=1}^k \psi_i x_{i,t} - \mathbb{E} \left[ \sum_{i=1}^k \psi_i x_{i,t} | F_{t-m}^{t+m} \right] \right\|_p &= \left\| \sum_{i=1}^k \psi_i (x_{i,t} - \mathbb{E}[x_{i,t} | F_{t-m}^{t+m}]) \right\|_p \\ &\leq \sum_{i=1}^k |\psi_i| \|(x_{i,t} - \mathbb{E}[x_{i,t} | F_{t-m}^{t+m}])\|_p \\ &\leq \sum_{i=1}^k |\psi_i| d_{i,t} \varphi_{i,m} \leq \left( \sum_{i=1}^k |\psi_i| d_{i,t} \right) \left( \sum_{i=1}^k \varphi_{i,m} \right) = d_t \times \varphi_m. \end{aligned}$$

■

**Доказательство теоремы 5.1.** Используя телескопическое свойство математического ожидания и свойство мартигал-разности, получаем, что для всех  $h \geq 1$

$$\begin{aligned} \gamma(h) &= \mathbb{E}[x_t x_{t-h}] = \mathbb{E}(x_{t-h} \mathbb{E}[x_t | \mathfrak{S}_{t-h}]) \\ &= \mathbb{E}(x_{t-h} \mathbb{E}(\mathbb{E}[x_t | \mathfrak{S}_{t-1}] | \mathfrak{S}_{t-h})) = \mathbb{E}(x_{t-h} \mathbb{E}(0 | \mathfrak{S}_{t-h})) = 0. \end{aligned}$$

Тогда по предположению

$$\frac{1}{n^2} \sum_{t=1}^n \sigma_t^2 + \frac{2}{n} \sum_{i=1}^{n-1} (1 - i/n) \times \gamma(i) = \frac{1}{n^2} \sum_{t=1}^{n-1} \sigma_t^2 = o(1).$$

■

**Доказательство леммы 5.1.** См. теорему 15.14 в работе Davidson (1994). ■

**Доказательство леммы 5.2.** См. следствие 20.2 в работе Davidson (1994). ■

**Доказательство теоремы 5.2.** Положим  $T_m := \sum_{t=1}^m x_t/a_t$  и заметим, что  $\{T_{n,m} := T_n - T_m, \mathfrak{S}_n\}$  является мартигалом, так как

$$\begin{aligned} \mathbb{E}[T_{n,m} | \mathfrak{S}_{n-1}] &= \mathbb{E} \left( \sum_{t=m+1}^n x_t/a_t | \mathfrak{S}_{n-1} \right) \\ &= \sum_{t=m+1}^n \mathbb{E}[x_t | \mathfrak{S}_{n-1}] / a_t = \sum_{t=m+1}^{n-1} x_t/a_t = T_{n-1,m}. \end{aligned}$$

Более того, используя телескопическое свойство математического ожидания и свойства мартигал-разности, получаем

$$\mathbb{E}(T_n - T_m)^2 = \sum_{t=m+1}^n \sigma_t^2 / a_t^2.$$

Из леммы 5.1 следует, что

$$\mathbb{P} \left\{ \max_{1 \leq k \leq n} |T_k - T_m| > \varepsilon \right\} \leq \frac{\mathbb{E}(T_n - T_m)^2}{\varepsilon^2} = \varepsilon^{-2} \sum_{t=m+1}^n \sigma_t^2 / a_t^2.$$

Теперь необходимо применить лемму 5.2 с  $c_t^2 = \sigma_t^2/a_t^2$ , чтобы получить

$$T_n = \sum_{t=1}^m x_t/a_t \xrightarrow{a.s.} T,$$

для некоторого  $T$ . Использование леммы Кронекера<sup>3</sup> завершает доказательство. Действительно, если  $\sum_{t=1}^m x_t/a_t \xrightarrow{a.s.} T$ , то  $\sum_{t=1}^m x_t/a_n \xrightarrow{a.s.} 0$ . ■

**Доказательство теоремы 5.3.** См. теорему 20.11 в работе Davidson (1994). ■

**Доказательство леммы 5.3.** Достаточно показать, что  $1/n \sum_{i=1}^n B_i \rightarrow 0$ . Но это следует из леммы Кронекера:  $\sum_{i=1}^{\infty} i^{-1} B_i < \infty \implies 1/n \sum_{i=1}^n B_i \rightarrow 0$ . ■

**Доказательство теоремы 5.4.** Следует из лемм 4.2 и 5.3. ■

**Доказательство теоремы 5.5.** При условиях (a) и (b) теорема 17.5 в работе Davidson (1994) утверждает, что  $\{x_{n,t}, F_t\}$  образует набор  $L_2$ -миксингалов размера  $1/2$  с миксингал-константами  $c_{n,t} \leq \max\{\|x_{n,t}\|_r, d_{n,t}\}$ . Следовательно (см. McLeish 1975),

$$\mathbb{E} \left( \sum_{t=1}^n x_{n,t} \right)^2 = O \left( \sum_{t=1}^n c_{n,t}^2 \right) = O \left( \sum_{t=1}^n (\max\{\|x_{n,t}\|_r, d_{n,t}\})^2 \right).$$

При условии (c)

$$\mathbb{E} \left( \sum_{t=1}^n x_{n,t} \right)^2 = O(1).$$

Но это означает, что для любой последовательности чисел  $\{a_n\}$ ,  $a_n \nearrow \infty$ ,

$$\mathbb{P} \left\{ \left| \frac{1}{a_n} \sum_{t=1}^n x_{n,t} \right| > \varepsilon \right\} \leq \varepsilon^{-2} \mathbb{E} \left( \frac{1}{a_n} \sum_{t=1}^n x_{n,t} \right)^2 = O(1/a_n^2) = o(1).$$

■

**Доказательство следствия 5.1.** Достаточно проверить выполнение условий теоремы 5.5 для

$$x_{n,t} := \frac{1}{\sqrt{n}} (x_t - \mathbb{E}[x_t]).$$

Ясно, что  $\mathbb{E}[x_{n,t}] = 0$ , и  $x_{n,t}$  является  $L_r$ -ограниченным в силу неравенств Минковского и Ляпунова:

$$\|x_{n,t}\|_r \leq \frac{1}{\sqrt{n}} (\|x_t\|_r + \|x_t\|_1) \leq \frac{2}{\sqrt{n}} \|x_t\|_r < \infty.$$

Значит, условие (a) теоремы 5.5 выполнено.

Заметим, что  $\{x_{n,t}\}$  является  $L_2$ -ПЭЗ размера  $1/2$  на  $\{\epsilon_t\}$  с константами  $d_{n,t} = d/\sqrt{n}$  и коэффициентами  $\varphi_m$ , поскольку

$$\|x_{n,t} - \mathbb{E}[x_{n,t}|F_{t-m}^{t+m}]\|_2 = (1/\sqrt{n}) \|x_t - \mathbb{E}[x_t|F_{t-m}^{t+m}]\|_2 \leq (d/\sqrt{n}) \varphi_m.$$

Значит, условие (b) теоремы 5.5 выполнено.

<sup>3</sup>См. Davidson (1994, лемма 2.35).

Из сильной стационарности и  $L_r$ -ограниченности заключаем, что

$$\begin{aligned} \sum_{t=1}^n (\max \{ \|x_{n,t}\|_r, d_{n,t} \})^2 &\leq \sum_{t=1}^n (\max \{ \| (1/\sqrt{n}) (x_t - \mathbb{E}[x_t]) \|_r, (d/\sqrt{n}) \})^2 \\ &\leq K \max \{ \|x_t\|_r, d \} \sum_{t=1}^n 1/n = K \max \{ \|x_t\|_r, d \} < \infty, \end{aligned}$$

где равенство выполнено равномерно по  $t$ . Значит, условие (c) теоремы 5.5 выполнено. ■

**Доказательство следствия 6.2.** Используя аргументацию из доказательств теоремы 5.5 и следствия 5.1, заключаем, что  $\{1/\sqrt{n}(x_t - \mathbb{E}[x_t])\}$  является  $L_2$ -ПЭЗ размера  $1/2$  на  $\{F_t\}$  с константами  $\tilde{d}_{n,t} = d/\sqrt{n}$  и коэффициентами  $\varphi_m$ . Так как база состоит из независимых и одинаково распределенных случайных величин, она является перемешивающей, и значит вследствие теоремы 17.5 работы Davidson (1994)  $\{1/\sqrt{n}(x_t - \mathbb{E}[x_t]), F_t\}$  является  $L_2$ -миксингал-массивом с константами, для которых

$$\tilde{c}_{n,t} \leq \max \{ \|x_{n,t}\|_r, \tilde{d}_{n,t} \} \leq K/\sqrt{n}$$

выполняется равномерно по  $t$ . Следовательно, (см. McLeish 1975)

$$\mathbb{E} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n (x_t - \mathbb{E}[x_t]) \right)^2 = O \left( \sum_{t=1}^n c_{n,t}^2 \right) = O \left( \sum_{t=1}^n 1/n \right) = O(1).$$

Учитывая  $\liminf_{n \geq 1} \mathbb{E}(1/\sqrt{n} \sum_{t=1}^n (x_t - \mathbb{E}[x_t]))^2 > 0$ , заключаем, что

$$x_{n,t} = \frac{1}{\sqrt{n}} \left( \frac{x_t - \mathbb{E}[x_t]}{\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n x_t)^2} \right)$$

является корректно определенным треугольным массивом.

Все условия теоремы выполняются, если положить  $\tilde{c}_{n,t} = K/\sqrt{n}$  для любого конечного  $K > 0$ . Например, очевидно, что  $\{x_{n,t}/c_{n,t}\}$   $L_r$ -ограничен, так как

$$\left\| \frac{x_{n,t}}{c_{n,t}} \right\|_r = K \left\| \left( \frac{x_t - \mathbb{E}[x_t]}{\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n x_t)^2} \right) \right\|_r \leq K \|x_t - \mathbb{E}[x_t]\|_r \leq K \|x_t\|_r,$$

где первое неравенство следует из  $\liminf_{n \geq 1} \mathbb{E}(1/\sqrt{n} \sum_{t=1}^n (x_t - \mathbb{E}[x_t]))^2 > 0$ , а второе — из неравенств Минковского и Ляпунова. ■

**Доказательство теоремы 6.4.** См. теоремы 25.5 и 25.6 в работе Davidson (1994). ■

**Доказательство теоремы 7.3.** Запишем

$$n(\hat{\rho} - 1) = \frac{1/n \sum_{t=2}^n \epsilon_t x_{t-1}}{1/n^2 \sum_{t=2}^n x_{t-1}^2}.$$

Достаточно доказать, что

$$\frac{1}{n} \sum_{t=2}^n \epsilon_t x_{t-1} \xrightarrow{d} \sigma^2 \frac{1}{2} (Z(\lambda)^2 - 1) \quad \text{и} \quad \frac{1}{n^2} \sum_{t=2}^n x_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 Z(\lambda)^2 d\lambda.$$

**Шаг 1.** Так как  $x_t^2 = (x_{t-1} + \epsilon_t)^2 = x_{t-1}^2 + 2x_{t-1}\epsilon_t + \epsilon_t^2$ , можно записать

$$x_{t-1}\epsilon_t = \frac{1}{2} (x_t^2 - x_{t-1}^2) - \frac{1}{2} \epsilon_t^2,$$

и значит

$$\sum_{t=2}^n \epsilon_t x_{t-1} = \frac{1}{2} \sum_{t=2}^n (x_t^2 - x_{t-1}^2) - \frac{1}{2} \sum_{t=1}^n \epsilon_t^2 = \frac{1}{2} \left( x_n^2 - \sum_{t=1}^n \epsilon_t^2 \right).$$

Так как  $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$ , легко показать, что  $1/n \sum_{t=1}^n \epsilon_t^2 \xrightarrow{P} \sigma^2$ . Рассмотрим член

$$bx_n^2 = \left( \sum_{\tau=1}^n \epsilon_\tau \right)^2.$$

В силу ЦПТ для независимых одинаково распределенных случайных величин и теоремы о непрерывном отображении

$$\frac{1}{n} x_n^2 = \sigma^2 \left( \frac{1}{\sqrt{n}} \sum_{\tau=1}^n \epsilon_\tau / \sigma \right)^2 \xrightarrow{d} \sigma^2 \times Z(1)^2,$$

где  $Z(1) \sim N(0, 1)$ . Следовательно,

$$\frac{1}{n} \sum_{t=2}^n \epsilon_t x_{t-1} \xrightarrow{d} \frac{1}{2} \left( \sigma^2 Z(1)^2 - \sigma^2 \right) = \sigma^2 \frac{1}{2} \left( Z(1)^2 - 1 \right).$$

**Шаг 2.** Чтобы доказать, что  $1/n^2 \sum_{t=2}^n x_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 Z(\lambda)^2 d\lambda$ , где  $\{Z(\lambda) : 0 \leq \lambda \leq 1\}$  есть броуновское движение, заметим, что  $1/n \sum_{t=2}^n (x_{t-1}^2/n)$  приближает интеграл от  $x_{t-1}^2/n$ , и для любого  $t$  подстановкой обратно по времени получаем

$$x_t / \sqrt{n} = \frac{1}{\sqrt{n}} \sum_{\tau=1}^{[n \times t/n]} \epsilon_\tau.$$

Предел  $1/n \sum_{t=2}^n (x_{t-1}^2/n)$  легко посчитать, если определить кусочный процесс

$$\begin{aligned} S_n(\lambda) &= 0 \text{ если } 0 \leq \lambda < 1/n, \\ &= x_1^2/n \text{ если } 1/n \leq \lambda < 2/n, \\ &\dots \\ &= x_{n-1}^2/n \text{ если } (n-1)/n \leq \lambda < n/n = 1, \\ &= x_n^2/n \text{ если } \lambda = 1. \end{aligned}$$

Следовательно, по определению,

$$\int_0^1 S_n(\lambda) d\lambda = \frac{1}{n} \sum_{t=2}^n x_{t-1}^2/n = \frac{1}{n^2} \sum_{t=2}^n x_{t-1}^2.$$

Далее, согласно теоремам 7.1 и 7.2 имеем слабую сходимость на  $C[0, 1]$ :

$$S_n(\lambda) = \sigma^2 \times \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\lambda]} \epsilon_t / \sigma \right)^2 \implies \sigma^2 \times Z(\lambda)^2.$$

Используя теорему о непрерывном отображении, заключаем, что

$$\int_0^1 S_n(\lambda) d\lambda \xrightarrow{d} \sigma^2 \int_0^1 Z(\lambda)^2 d\lambda.$$

■

# Dependence and stochastic limit theory

**Jonathan Hill**

*University of North Carolina, Chapel Hill, USA*

In this essay we provide the basic asymptotic theory that serves as background theory for estimators in time series. We outline concepts of dependence used for stochastic limit theory, covering mixing, mixingale and near epoch dependence properties. We then detail some of the most general probability and distribution limit theorems for these processes popularly employed for time series theory and applications.





# Линейные процессы: свойства и асимптотические результаты\*

Вадим Мармер<sup>†</sup>

*Университет Британской Колумбии, Ванкувер, Канада*

Настоящее эссе содержит обзор результатов, касающихся линейных временных рядов и включающих разложение Вольда, свойства спектральных плотностей и операторов сдвига, моделей ARMA, разложение Бевеиджа–Нельсона и метод получения асимптотических результатов Филлипса–Соло.

## 1 Введение

В эконометрической литературе можно выделить два основных подхода к получению асимптотических результатов. Первый подход базируется на предельных теоремах для временных рядов, удовлетворяющих условиям сильного перемешивания. Подробное описание данного подхода дано, например, в Davidson (1994), а также White (2001), причем в последней книге акцент делается на приложениях. Причина популярности концепций перемешивания кроется в том, что это понятие позволяет включить как зависимость, так и неоднородность, так как временные ряды, обладающие свойствами перемешивания, не обязаны быть стационарными. В то же время, этот подход не универсален, так как даже авторегрессионные (AR) процессы первого порядка могут не быть сильно перемешивающими, см. Andrews (1984).

Второй подход, являющийся предметом настоящей обзорной статьи, базируется на линейных процессах  $\{X_t : t = 1, 2, \dots\}$  вида

$$X_t = \sum_{j=1}^{\infty} c_j \varepsilon_{t-j}, \quad (1)$$

где  $c_j$  — константы, а  $\{\varepsilon_t : t = 1, 2, \dots\}$  — серийно некоррелированный процесс. Данная модель предоставляет очень интуитивный способ генерации серийно зависимых данных: начиная с серийно некоррелированного процесса  $\{\varepsilon_t\}$ , можно получить временную зависимость в  $\{X_t\}$  путем усреднения случайных величин  $\varepsilon_t$ .

Асимптотический метод для линейных процессов был разработан в Phillips & Solo (1992) и основан на алгебраическом разложении временного ряда на долгосрочную и переходную компоненты. Разложение было предложено в Beveridge & Nelson (1981) и известно в литературе как разложение Бевеиджа–Нельсона (БН). Подход Филлипса и Соло привлекателен тем, что опирается на законы больших чисел (ЗБЧ) и центральные предельные теоремы (ЦПТ) для независимых процессов или мартингал-разностей. Используя линейную структуру процесса и соответствующее БН-разложение, Филлипс и Соло предлагают применить предельные теоремы для независимых процессов или мартингал-разностей к долгосрочной компоненте и показать, что переходной компонентой можно асимптотически пренебречь. Этот метод можно использовать для классических линейных процессов, таких как ARMA-процессы, а также процессов скользящего среднего (MA) бесконечного порядка.<sup>1</sup>

\*Цитировать как: Мармер, Вадим (2012) «Линейные процессы: свойства и асимптотические результаты», Квантиль, №10, стр. 33–56. Citation: Marmer, Vadim (2012) “Linear processes: properties and asymptotic results,” Quantile, No.10, pp. 33–56.

<sup>†</sup>Адрес: Department of Economics, University of British Columbia, 997 – 1873 East Mall, Vancouver, BC, V6T 1Z1, Canada. Электронная почта: [vadim.marmer@ubc.ca](mailto:vadim.marmer@ubc.ca)

<sup>1</sup>Как в уравнении (1).

Несмотря на свою простоту, метод Филлипса и Соло обладает замечательной общностью в силу знаменитого разложения Вольда, в котором слабостационарные процессы представляются линейными процессами, см., например, Hansen & Sargent (1991). Разложение Вольда часто служит оправданием для моделирования временных рядов как линейных процессов. Если необходимо немного усилить результат разложения Вольда, можно применять метод Филлипса и Соло.

Настоящее эссе имеет далее следующую структуру. В разделе 2 обсуждается разложение Вольда. В разделе 3 описываются некоторые важные инструменты, необходимые для работы с линейными процессами. Их использование иллюстрируется в разделе 4 на примере ARMA-моделей. Наконец, в разделе 5 описывается БН-разложение и асимптотический метод Phillips & Solo (1992).

Помимо цитированных в списке литературы источников, эссе основано также на неопубликованных рукописных лекциях Питера Филлипса.<sup>2</sup>

## 2 Разложение Вольда

Сфокусируем внимание на слабостационарных процессах, удовлетворяющих следующему определению.

**Определение 1 (Слабая стационарность и автоковариационная функция)** Процесс  $\{X_t\}$ , такой что  $\mathbb{E}X_t^2 < \infty$  для всех  $t$ , называется слабостационарным, если  $\mathbb{E}X_t = \mu$  и  $\mathbb{C}(X_t, X_s) = \gamma(|s - t|)$  для всех  $s$  и  $t$ , где  $\gamma(j)$ ,  $j = 0, 1, \dots$ , называется автоковариационной функцией  $\{X_t\}$ .

Слабостационарный серийно некоррелированный процесс с нулевым средним называется белым шумом (БШ). Процессы БШ являются строительными блоками в конструкции MA-процессов:

**Определение 2 (Белый шум)** Процесс  $\{\varepsilon_t\}$  называется БШ, если  $\mathbb{E}\varepsilon_t = 0$ ,  $\mathbb{E}\varepsilon_t^2 = \sigma^2 < \infty$  и  $\mathbb{E}\varepsilon_t\varepsilon_{t-j} = 0$  для всех  $t$  и  $j \neq 0$ .

**Определение 3 (MA)** Процесс  $\{X_t\}$  называется процессом скользящего среднего порядка  $q$  (MA( $q$ )), если

$$X_t = c_0\varepsilon_t + c_1\varepsilon_{t-1} + \dots + c_q\varepsilon_{t-q}, \quad (2)$$

и  $\{\varepsilon_t\}$  есть БШ.

Расширение до бесконечного порядка (2), как в уравнении (1), обозначается MA( $\infty$ ). Разложение Вольда показывает, что любой слабостационарный процесс с нулевым средним и абсолютно суммируемыми автоковариациями может быть представлен в форме MA( $\infty$ ).

Обозначим за  $L_2$  пространство случайных величин с конечными вторыми моментами. Для  $X, Y \in L_2$  определим скалярное произведение

$$\langle X, Y \rangle = \mathbb{E}XY.$$

С таким определением скалярного произведения пространство  $L_2$  становится гильбертовым. Рассмотрим слабостационарный процесс  $\{X_t\}$  с нулевым средним, такой что

$$\sum_{j=0}^{\infty} |\gamma(j)| < \infty, \quad (3)$$

<sup>2</sup>Лекции доступны по адресу <http://cowles.econ.yale.edu/korora/phillips/teach/lec-notes.htm>.

где  $\gamma(j)$  обозначает автоковариационную функцию

$$\gamma(j) = \mathbb{E}X_t X_{t-j}.$$

Определим  $\mathcal{M}_t$  как наименьшее замкнутое подпространство пространства  $L_2$ , содержащее все элементы вида

$$\sum_{j=0}^{\infty} c_j X_{t-j}, \text{ такие что } \sum_{j=0}^{\infty} c_j^2 < \infty.$$

Требование  $\sum_{j=0}^{\infty} c_j^2 < \infty$  необходимо для того, чтобы элементы  $\mathcal{M}_t$  принадлежали  $L_2$ . Действительно, в этом случае дисперсию любого элемента из  $\mathcal{M}_t$  можно ограничить суммами  $\sum_{j=0}^{\infty} c_j^2$  и  $\sum_{j=0}^{\infty} |\gamma(j)|$ , как показано ниже:

$$\begin{aligned} \mathbb{V} \left( \sum_{j=0}^{\infty} c_j X_{t-j} \right) &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_i c_j \gamma(i-j) \leq 2 \sum_{h=0}^{\infty} \sum_{j=0}^{\infty} c_j c_{j+h} \gamma(h) = 2 \sum_{h=0}^{\infty} \gamma(h) \sum_{j=0}^{\infty} c_j c_{j+h} \\ &\leq 2 \sum_{h=0}^{\infty} |\gamma(h)| \left| \sum_{j=0}^{\infty} c_j c_{j+h} \right| \leq 2 \sum_{h=0}^{\infty} |\gamma(h)| \left( \sum_{j=0}^{\infty} c_j^2 \right)^{1/2} \left( \sum_{j=0}^{\infty} c_{j+h}^2 \right)^{1/2} \\ &\leq 2 \left( \sum_{j=0}^{\infty} c_j^2 \right) \sum_{h=0}^{\infty} |\gamma(h)| \\ &< \infty, \end{aligned} \tag{4}$$

где последнее неравенство выполнено, если верно неравенство (3). Заметим далее, что последовательность  $\mathcal{M}_t$  возрастает:

$$\dots \subset \mathcal{M}_t \subset \mathcal{M}_{t+1} \subset \dots$$

Пусть  $P_{\mathcal{M}_t}$  – ортогональная проекция на  $\mathcal{M}_t$ . Можно записать

$$X_t = \widehat{X}_t + \varepsilon_t, \tag{5}$$

где

$$\begin{aligned} \widehat{X}_t &= P_{\mathcal{M}_{t-1}} X_t, \\ \varepsilon_t &= (1 - P_{\mathcal{M}_{t-1}}) X_t. \end{aligned}$$

Из известных результатов для гильбертовых пространств следует, что  $\widehat{X}_t$  решает задачу наименьших квадратов. Значит,  $\widehat{X}_t$  можно интерпретировать как лучший (в среднеквадратическом смысле) линейный предиктор величины  $X_t$  на один шаг вперед, а величину  $\varepsilon_t$  – как ошибку предсказания. Заметим, что  $\widehat{X}_t \in \mathcal{M}_{t-1}$ , и  $\varepsilon_t \in \mathcal{M}_t$ , так как

$$\varepsilon_t = X_t - \widehat{X}_t.$$

Более того, в силу свойств ортогональной проекции

$$\varepsilon_t \in \mathcal{M}_{t-1}^{\perp},$$

где

$$\mathcal{M}_t^{\perp} = \{Y \in L_2 : \langle Y, X \rangle = 0 \text{ для всех } X \in \mathcal{M}_t\}.$$

Так как последовательность  $\mathcal{M}_t$  возрастает, она содержит все элементы из  $\mathcal{M}_{t-h}$ , и мы получаем, что  $\varepsilon_t \in \mathcal{M}_{t-h}^\perp$  для всех  $h \geq 1$ . Следовательно,

$$\mathbb{E}\varepsilon_t\varepsilon_{t-h} = 0 \text{ для всех } h \geq 1.$$

Более того, дисперсия равна

$$\mathbb{E}\varepsilon_t^2 = \sigma^2 < \infty$$

и не зависит от  $t$ , так как процесс  $\{X_t\}$  слабостационарен.

Определим

$$\mathcal{E}_t = \left\{ \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} : \sum_{j=0}^{\infty} c_j^2 < \infty \right\},$$

где  $\mathcal{E}_t$  является на самом деле замкнутым линейным подпространством пространства  $L_2$ . Обозначим за  $P_{\mathcal{E}_t}$  ортогональную проекцию на  $\mathcal{E}_t$ . Имеем:

$$X_t = P_{\mathcal{E}_t} X_t + V_t, \tag{6}$$

где в силу замкнутости и линейности  $\mathcal{E}_t$

$$P_{\mathcal{E}_t} X_t = \sum_{j=0}^{\infty} \hat{c}_j \varepsilon_{t-j} \tag{7}$$

для некоторой последовательности  $\{\hat{c}_j\}$ , и

$$V_t = (1 - P_{\mathcal{E}_t}) X_t.$$

Заметим, что  $V_t \in \mathcal{M}_t$ , и в силу (5)

$$\mathcal{M}_t = \mathcal{M}_{t-1} \oplus \mathcal{S}(\varepsilon_t),$$

где  $\mathcal{S}(\varepsilon_t)$  — линейное подпространство, порожденное  $\varepsilon_t$ , а  $\oplus$  обозначает прямую сумму:

$$\mathcal{S}_1 \oplus \mathcal{S}_2 = \{x_1 + x_2 : x_1 \in \mathcal{S}_1, x_2 \in \mathcal{S}_2\}.$$

Так как  $P_{\mathcal{E}_t}$  — ортогональная проекция,  $\mathbb{E}V_t\varepsilon_t = 0$  в силу (6) и (7), и значит  $V_t \notin \mathcal{S}(\varepsilon_t)$ . Следовательно, должно быть выполнено включение  $V_t \in \mathcal{M}_{t-1}$ . Аналогично, так как  $\mathcal{M}_{t-1} = \mathcal{M}_{t-2} \oplus \mathcal{S}(\varepsilon_{t-1})$  и  $\mathbb{E}V_t\varepsilon_{t-1} = 0$ , должны быть выполнены включения  $V_t \in \mathcal{M}_{t-1}, \mathcal{M}_{t-2}, \dots$ . Пусть

$$\mathcal{M}_{-\infty} = \bigcap_{t=-\infty}^{\infty} \mathcal{M}_t.$$

Тогда

$$V_t \in \mathcal{M}_{-\infty} \text{ для всех } t.$$

Таким образом,  $V_t$  является элементом любого линейного подпространства  $\mathcal{M}_s$ ,  $s \in \mathbb{Z}$ , где  $\mathbb{Z}$  обозначает множество целых чисел. Весь процесс  $\{V_t\}$  можно точно предсказать на основе произвольно далекой истории процесса  $\{X_t\}$ . Такие процессы называются *детерминистическими*.

Мы вывели разложение Вольда для слабостационарного процесса  $\{X_t\}$ :

$$X_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} + V_t,$$

где  $\{\varepsilon_t\}$  — БШ, а  $V_t$  — детерминистический процесс. Такое представление единственно почти наверное.

Можно показать, что

$$c_j = \mathbb{E}X_t\varepsilon_{t-j}/\sigma^2 \tag{8}$$

и

$$c_0 = 1, \tag{9}$$

где  $\sigma^2 = \mathbb{V}(\varepsilon_t)$ .<sup>3</sup>

Если  $V_t = 0$  для всех  $t$ , процесс  $\{X_t\}$  называется *чисто недетерминистическим*. Разумно предположить, что экономические временные ряды не содержат детерминистических компонент, которые можно бы было точно предсказать на основе очень далеких историй, а значит, являются чисто недетерминистическими.

Возникает желание дать величинам  $\varepsilon$  структурную экономическую интерпретацию и считать, что они отвечают за фундаментальные шоки переменной  $X$ . В этом случае, так как величины  $\varepsilon$  серийно некоррелированы, эффектом шока после  $j$  периодов можно считать частную производную  $\partial X_t/\partial \varepsilon_{t-j} = c_j$ , которая называется *импульсным откликом*. К сожалению, такая интерпретация инкрементов в разложении Вольда редко бывает оправданной. Заметим, что по построению величины  $\varepsilon$  являются ошибками *линейного* прогноза на один шаг вперед. Предположим, что фундаментальные шоки являются неожиданными шоками для информационного множества агента. Пусть  $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots)$  —  $\sigma$ -алгебра, порожденная историей  $X$  до времени  $t$ . Эта  $\sigma$ -алгебра представляет информацию агентов в экономике до момента времени  $t$ . Однако,  $\hat{X}_t$  в уравнении (5) должен быть линейным предиктором, и значит, необязательно равен  $\mathbb{E}(X_t|\mathcal{F}_{t-1})$ . Таким образом, истинные фундаментальные шоки могут отличаться от шоков в разложении Вольда. См. обсуждение в главе 4 книги Hansen & Sargent (1991).

### 3 Инструментарий линейных процессов

Обозначим за  $X_1, X_2, \dots, X_n$  наблюдения за случайной величиной  $X_t$ . Тогда ЦПТ формулируется следующим образом:

$$\frac{n^{-1/2} \sum_{t=1}^n (X_t - \mathbb{E}X_t)}{\sqrt{\mathbb{V}(n^{-1/2} \sum_{t=1}^n X_t)}} \rightarrow_d N(0, 1).$$

Дисперсия  $\mathbb{V}(n^{-1} \sum_{t=1}^n X_t)$  представляет особый интерес. С серийно коррелированными данными,

$$\begin{aligned} \mathbb{V}\left(n^{-1/2} \sum_{t=1}^n X_t\right) &= n^{-1} \sum_{t=1}^n \sum_{s=1}^n \mathbb{C}(X_t, X_s) \\ &= n^{-1} \sum_{t=1}^n \mathbb{V}(X_t) + 2 \sum_{j=1}^{n-1} n^{-1} \sum_{t=j+1}^n \mathbb{C}(X_t, X_{t-j}). \end{aligned}$$

Пусть  $\gamma(j) = \mathbb{C}(X_t, X_{t-j})$ . Предполагая, что  $\{X_t\}$  слабостационарен, для больших  $n$  дисперсию можно аппроксимировать выражением

$$\omega \equiv \lim_{n \rightarrow \infty} \mathbb{V}\left(n^{-1/2} \sum_{t=1}^n X_t\right) = \gamma(0) + 2 \sum_{j=1}^{\infty} \gamma(j),$$

---

<sup>3</sup>Запишем  $X_t\varepsilon_{t-j} = \left(\sum_{j=0}^{\infty} c_j\varepsilon_{t-j} + V_t\right)\varepsilon_{t-j}$ . Из этого следуют уравнения (8) и (9), так как  $\{\varepsilon_t\}$  — БШ и  $\mathbb{E}V_t\varepsilon_{t-j} = 0$  по построению.

при условии, что предел существует. В этом определении  $\omega$  обозначает так называемую *долгосрочную дисперсию* процесса  $\{X_t\}$ .

**Определение 4 (Долгосрочная дисперсия)** Для слаботационарного процесса с автоковариационной функцией  $\gamma(j)$ ,  $j = 0, 1, \dots$ , долгосрочная дисперсия (если существует) задается выражением  $\gamma(0) + 2 \sum_{j=1}^{\infty} \gamma(j)$ .

Заметим, что необходимым и достаточным условием существования долгосрочной дисперсии является суммируемость

$$\sum_{j=0}^{\infty} |\gamma(j)| < \infty.$$

Ясно, что автоковариационная функция  $\gamma(j)$  играет важную роль в эконометрике временных рядов, так как в пределе зависимость по времени отражается долгосрочной дисперсией:

$$n^{-1/2} \sum_{t=1}^n (X_t - \mathbb{E}X_t) \rightarrow_d N \left( 0, \gamma(0) + 2 \sum_{j=1}^{\infty} \gamma(j) \right).$$

### 3.1 Спектральная плотность<sup>4</sup>

Зависимость по времени асимптотически отражается автоковариационной функцией. Удобно агрегировать эту зависимость так называемой функцией спектральной плотности. В то время как автоковариационная функция определена на бесконечной последовательности действительных чисел, спектральную плотность достаточно рассматривать на компактном интервале.

**Определение 5 (Спектральная плотность)** Для автоковариационной функции  $\gamma(j)$ ,  $j = 0, 1, \dots$ , соответствующая спектральная плотность определяется как

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-i\lambda j},$$

где  $i = \sqrt{-1}$ .

Так как  $\gamma(j) = \gamma(-j)$ , спектральная плотность действительнoзначна.

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi} \left( \gamma(0) + \sum_{j=-\infty}^{-1} \gamma(j) e^{-i\lambda j} + \sum_{j=1}^{\infty} \gamma(j) e^{-i\lambda j} \right) \\ &= \frac{1}{2\pi} \left( \gamma(0) + \sum_{j=1}^{\infty} \gamma(-j) e^{i\lambda j} + \sum_{j=1}^{\infty} \gamma(j) e^{-i\lambda j} \right) \\ &= \frac{1}{2\pi} \left( \gamma(0) + \sum_{j=1}^{\infty} \gamma(j) (e^{i\lambda j} + e^{-i\lambda j}) \right). \end{aligned}$$

Далее,  $e^{i\lambda j} = \cos(\lambda j) + i \sin(\lambda j)$ ,  $e^{-i\lambda j} = \cos(\lambda j) - i \sin(\lambda j)$ , и, стало быть,

$$e^{i\lambda j} + e^{-i\lambda j} = 2 \cos(\lambda j).$$

<sup>4</sup>Материал этого раздела основан на главе 5 книги Gourieroux & Monfort (1995) и главе 6 книги Hamilton (1994).

Значит,

$$f(\lambda) = \frac{1}{2\pi} \left( \gamma(0) + 2 \sum_{j=1}^{\infty} \gamma(j) \cos(\lambda j) \right) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) \cos(\lambda j).$$

Так как  $\cos(\lambda j) = \cos(-\lambda j)$ , спектральная плотность симметрична относительно нуля. Более того, так как функция  $\cos$  периодична с периодом  $2\pi$ , область значений спектральной плотности определяется значениями  $f(\lambda)$  в точках  $0 \leq \lambda \leq \pi$ .

Автоковариационная функция и спектральная плотность эквивалентны в том смысле, что всегда можно однозначно восстановить автоковариационную функцию по спектральной плотности. Таким образом, они содержат одинаковое количество информации, что демонстрируется следующим результатом.

**Теорема 1** Пусть  $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ . Тогда  $\gamma(j) = \int_{-\pi}^{\pi} f(\lambda) e^{i\lambda j} d\lambda$ .

**Доказательство.** Имеем

$$\int_{-\pi}^{\pi} f(\lambda) e^{i\lambda j} d\lambda = \int_{-\pi}^{\pi} \left( \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\lambda h} \right) e^{i\lambda j} d\lambda = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \int_{-\pi}^{\pi} e^{i\lambda(j-h)} d\lambda,$$

где суммирование и интегрирование можно менять местами, так как  $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ . Далее,

$$\int_{-\pi}^{\pi} e^{i\lambda(j-h)} d\lambda = 2\pi \text{ если } j = h.$$

Для  $j \neq h$

$$\begin{aligned} \int_{-\pi}^{\pi} e^{i\lambda(j-h)} d\lambda &= \int_{-\pi}^{\pi} (\cos(\lambda(j-h)) + i \sin(\lambda(j-h))) d\lambda \\ &= \frac{\sin(\lambda(j-h))}{j-h} \Big|_{-\pi}^{\pi} - i \frac{\cos(\lambda(j-h))}{j-h} \Big|_{-\pi}^{\pi} \\ &= \frac{\sin(\pi(j-h)) - \sin(-\pi(j-h))}{j-h} - i \frac{\cos(\pi(j-h)) - \cos(-\pi(j-h))}{j-h}. \end{aligned}$$

Однако, так как функции  $\cos$  и  $\sin$  периодичны с периодом  $2\pi$ ,

$$\cos(\pi(j-h)) = \cos(-\pi(j-h) + 2\pi(j-h)) = \cos(-\pi(j-h)).$$

Следовательно,

$$\int_{-\pi}^{\pi} e^{i\lambda(j-h)} d\lambda = 0 \text{ если } j \neq h.$$

■

В частности, из теоремы 1 следует, что

$$\gamma(0) = \int_{-\pi}^{\pi} f(\lambda) d\lambda.$$

Таким образом, площадь под спектральной плотностью процесса  $X_t$  между точками  $-\pi$  и  $\pi$  равна дисперсии величины  $X_t$ .

Аргумент  $\lambda$  функции  $f(\lambda)$  называется частотой. Заметим, что если  $\{X_t\}$  слабостационарен с абсолютно суммируемыми автоковариациями, долгосрочная дисперсия определяется значением спектральной плотности в нулевой частоте:

$$\omega_X = \lim_{n \rightarrow \infty} \mathbb{V} \left( n^{-1/2} \sum_{t=1}^n X_t \right) = \sum_{h=-\infty}^{\infty} \gamma(h) = 2\pi f(0).$$

Далее мы обсудим, как линейные (МА) преобразования слабостационарных процессов влияют на спектральную плотность и долгосрочную дисперсию.

**Теорема 2** Пусть  $\{X_t\}$  – слабостационарный процесс с автоковариационной функцией  $\gamma_X$ , такой что  $\sum_{j=-\infty}^{\infty} |\gamma_X(j)| < \infty$ . Определим  $Y_t = \sum_{j=0}^{\infty} c_j X_{t-j}$ , где  $\sum_{j=0}^{\infty} c_j^2 < \infty$ . Тогда процесс  $\{Y_t\}$  слабостационарен, а его спектральная плотность  $f_Y(\lambda) = \left| \sum_{j=0}^{\infty} c_j e^{-i\lambda j} \right|^2 f_X(\lambda)$ , где  $f_X$  – спектральная плотность процесса  $\{X_t\}$ .

**Доказательство.** Имеем

$$\begin{aligned} \mathbb{C}(Y_t, Y_{t-h}) &= \mathbb{C} \left( \sum_{j=0}^{\infty} c_j X_{t-j}, \sum_{j=0}^{\infty} c_j X_{t-h-j} \right) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} c_j c_k \mathbb{C}(X_{t-j}, X_{t-h-k}) \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} c_j c_k \gamma_X(h+k-j). \end{aligned}$$

Таким образом,  $\mathbb{C}(Y_t, Y_{t-h})$  не зависит от  $t$ . Более того, аналогично выводу неравенства (4),

$$\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} c_j c_k \gamma_X(h+k-j) \leq 2 \left( \sum_{j=0}^{\infty} c_j^2 \right) \sum_{h=0}^{\infty} |\gamma_X(h)| < \infty.$$

Следовательно,  $\{Y_t\}$  слабостационарен.

Далее,

$$\begin{aligned} f_Y(\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathbb{C}(Y_t, Y_{t-h}) e^{-i\lambda h} = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} c_j c_k \gamma_X(h+k-j) e^{-i\lambda h} \\ &= \sum_{j=0}^{\infty} c_j e^{-i\lambda j} \sum_{k=0}^{\infty} c_k e^{i\lambda k} \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_X(h+k-j) e^{-i\lambda(h+k-j)} \\ &= \left( \sum_{j=0}^{\infty} c_j e^{-i\lambda j} \right) \left( \sum_{j=0}^{\infty} c_j e^{i\lambda j} \right) f_X(\lambda) = \left| \sum_{j=0}^{\infty} c_j e^{-i\lambda j} \right|^2 f_X(\lambda). \end{aligned}$$

Последнее неравенство верно в силу

$$\sum_j c_j e^{-i\lambda j} = \sum_j c_j (\cos(\lambda j) - i \sin(\lambda j)) = \sum_j c_j \cos(\lambda j) - i \sum_j c_j \sin(\lambda j).$$

Комплексным сопряжением получаем

$$\sum_j c_j \cos(\lambda j) + i \sum_j c_j \sin(\lambda j) = \sum_j c_j (\cos(\lambda j) + i \sin(\lambda j)) = \sum_j c_j e^{i\lambda j},$$



и значит,

$$\left| \sum_{j=0}^{\infty} c_j e^{-i\lambda j} \right|^2 = \left( \sum_{j=0}^{\infty} c_j \cos(\lambda j) \right)^2 + \left( \sum_{j=0}^{\infty} c_j \sin(\lambda j) \right)^2 = \left( \sum_{j=0}^{\infty} c_j e^{-i\lambda j} \right) \left( \sum_{j=0}^{\infty} c_j e^{i\lambda j} \right).$$

■

В данной теореме спектральная плотность в нуле и, стало быть, долгосрочная дисперсия конечны, если  $\sum_{j=0}^{\infty} |c_j| < \infty$ . Однако абсолютная суммируемость,  $\sum_{j=0}^{\infty} |c_j| < \infty$ , является более сильным предположением, чем суммируемость квадратов,  $\sum_{j=0}^{\infty} c_j^2 < \infty$ , как мы сейчас покажем. Пусть  $\sum_{j=0}^{\infty} |c_j| < \infty$ . Из  $\sum_{j=0}^{\infty} |c_j| < \infty$  следует, что  $c_j \rightarrow 0$  при  $j \rightarrow \infty$ . Значит, последовательность  $\{c_j\}$  равномерно ограничена. Наконец,  $\sum_{j=0}^{\infty} c_j^2 \leq \sup_j |c_j| \sum_{j=0}^{\infty} |c_j| < \infty$ .

Предположим, что процесс  $\{X_t\}$  слабостационарен и чисто недетерминистичен. Тогда в силу разложения Вольда он имеет МА( $\infty$ ) представление

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \tag{10}$$

где  $\{\varepsilon_t\}$  — БШ, и  $\sum_{j=0}^{\infty} a_j^2 < \infty$ . Пусть  $\mathbb{V}(\varepsilon_t) = \sigma^2$ . Так как спектр БШ плоский:

$$f(\lambda) = \frac{\sigma^2}{2\pi} \text{ для всех } \lambda,$$

из теоремы 2 следует, что

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} a_j e^{-i\lambda j} \right|^2,$$

а долгосрочная дисперсия процесса  $\{X_t\}$  равна

$$\omega_X = 2\pi f_X(0) = \sigma^2 \left( \sum_{j=0}^{\infty} a_j \right)^2.$$

Если рассматривать уравнение (10) как механизм генерации, то условие  $\sum_{j=0}^{\infty} a_j^2 < \infty$  гарантирует, что  $\{X_t\}$  слабостационарен. Однако достаточным условием конечности долгосрочной дисперсии является суммируемость  $\sum_{j=0}^{\infty} |a_j| < \infty$ . Если последнее условие не выполняется, может случиться, что  $\sum_{j=-\infty}^{\infty} |\gamma_X(j)| = \infty$ . Говорят, что такие процессы обладают *длинной памятью*. Если  $\sum_{j=0}^{\infty} a_j^2 < \infty$  выполнено для процесса с длинной памятью, то автоковариационная функция сходится к нулю слишком медленно для того, чтобы долгосрочная дисперсия была конечна.

**Определение 6 (Короткая и длинная память)** Слабостационарный процесс с автоковариационной функцией  $\gamma(j)$ ,  $j = 0, 1, \dots$ , обладает короткой памятью, если  $\sum_{j=0}^{\infty} |\gamma(j)| < \infty$ . Длинная память имеет место в случае  $\sum_{j=0}^{\infty} |\gamma(j)| = \infty$ .

Пусть  $\{Y_t\}$  определен как в теореме 2. Тогда его спектральная плотность и долгосрочная дисперсия равны

$$f_Y(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} a_j e^{-i\lambda j} \right|^2 \left| \sum_{j=0}^{\infty} c_j e^{-i\lambda j} \right|^2,$$

$$\omega_Y = \sigma^2 \left( \sum_{j=0}^{\infty} a_j \right)^2 \left( \sum_{j=0}^{\infty} c_j \right)^2.$$

### 3.2 Оператор сдвига<sup>5</sup>

Оператор сдвига  $L$  отображает процесс  $\{X_t\}$  в себя, так что

$$\begin{aligned} LX_t &= X_{t-1}, \\ L^2X_t &= LLX_t = LX_{t-1} = X_{t-2}, \\ &\dots \\ L^hX_t &= X_{t-h}. \end{aligned}$$

Многочлен от оператора сдвига

$$C(L) = \sum_{j=0}^{\infty} c_j L^j$$

преобразует  $\{X_t\}$  в другой процесс  $\{Y_t\}$ , такой что

$$Y_t = C(L)X_t = \sum_{j=0}^{\infty} c_j L^j X_t = \sum_{j=0}^{\infty} c_j X_{t-j}.$$

Пусть  $A(L) = \sum_{j=0}^{\infty} a_j L^j$  и  $B(L) = \sum_{j=0}^{\infty} b_j L^j$ . Тогда

$$A(L) + B(L) = \sum_{j=0}^{\infty} (a_j + b_j) L^j,$$

и

$$\begin{aligned} A(L)B(L) &= \left( \sum_{j=0}^{\infty} a_j L^j \right) \left( \sum_{j=0}^{\infty} b_j L^j \right) = \sum_{j=0}^{\infty} \sum_{h=0}^{\infty} a_j b_h L^{j+h} \\ &= a_0 b_0 + (a_0 b_1 + b_0 a_1) L + (a_1 b_1 + a_0 b_2 + a_2 b_0) L^2 + \dots \end{aligned}$$

Имеем:

$$\begin{aligned} A(L) + B(L) &= B(L) + A(L), \\ A(L)B(L) &= B(L)A(L). \end{aligned}$$

При определенных условиях многочлен от оператора сдвига может быть обращен. Обратный к многочлену  $C(L)$  есть еще один многочлен от оператора сдвига, скажем,  $B(L)$ , такой что

$$C(L)B(L) = 1. \tag{11}$$

Можно записать

$$C(L)^{-1} = B(L).$$

Обращение многочленов от оператора сдвига важно по следующей причине. Рассмотрим, например, авторегрессионный процесс первого порядка AR(1):

$$X_t = cX_{t-1} + \varepsilon_t. \tag{12}$$

<sup>5</sup>Материал этого раздела основан на главе 5 книги Gourieroux & Monfort (1995) и главе 2 книги Hamilton (1994).

Этот процесс рекурсивно генерируется на основе экзогенного белого шума  $\{\varepsilon_t\}$ , начального значения  $X_0$  (случайной величины с дисперсией  $V(X_t)$ , которую мы рассчитаем позже), и коэффициента  $c$ :

$$\begin{aligned} X_1 &= cX_0 + \varepsilon_1, \\ X_2 &= cX_1 + \varepsilon_2, \end{aligned}$$

и.т.д. Таким образом, процесс  $\{X_t\}$  является эндогенным решением *разностного уравнения* (12). Это разностное уравнение может быть переписано в виде  $X_t - cX_{t-1} = \varepsilon_t$  или

$$\begin{aligned} C(L)X_t &= \varepsilon_t, \\ C(L) &= 1 - cL. \end{aligned}$$

Если  $C(L)$  можно обратить, то решение может быть записано как  $X_t = C(L)^{-1}\varepsilon_t$ . Таким образом, важно вывести условия обратимости многочлена от оператора сдвига и найти способ вычисления коэффициентов обратного многочлена.

Рассмотрим для начала многочлен порядка 1. Без ограничения общности можно положить коэффициент при  $L^0$  единицей  $c_0 = 1$ :<sup>6</sup>

$$C(L) = 1 - cL.$$

Предположим, что  $|c| < 1$ . Тогда можно определить многочлен, обратный к  $1 - cL$ , следующим образом:

$$(1 - cL)^{-1} = 1 + cL + c^2L^2 + \dots \quad (13)$$

Это определение удовлетворяет (11), так как

$$(1 - cL) \sum_{j=0}^{\infty} c^j L^j = 1.$$

Решение (13) не единственное, удовлетворяющее условию (11). Добавление к нему члена  $Vc^t$ , где  $V$  — некоторая случайная величина, тоже удовлетворяет ему, так как

$$(1 - cL)Vc^t = Vc^t - VcLc^t = Vc^t - Vcc^{t-1} = 0.$$

Однако если взять  $(1 - cL)^{-1} = \sum_{j=0}^{\infty} c^j L^j + Vc^t$ , то процесс  $(1 - cL)^{-1}\varepsilon_t = \sum_{j=0}^{\infty} c^j \varepsilon_{t-j} + Vc^t \varepsilon_t$  не будет слабостационарным. Поэтому мы накладываем ограничение  $V = 0$ .

Далее, рассмотрим многочлен от оператора сдвига порядка 2:

$$C(L) = 1 - c_1L - c_2L^2.$$

Мы можем разложить многочлен на множители:

$$1 - c_1L - c_2L^2 = (1 - \lambda_1L)(1 - \lambda_2L), \quad (14)$$

где

$$\begin{aligned} c_1 &= \lambda_1 + \lambda_2, \\ c_2 &= -\lambda_1\lambda_2. \end{aligned}$$

---

<sup>6</sup>Это следует из разложения Вольда. С другой стороны, это можно рассматривать как нормализацию: умножение всех коэффициентов на некоторую константу  $c_0 \neq 0$  влияет только на дисперсию процесса.

Другой способ нахождения  $\lambda_1$  и  $\lambda_2$  состоит в следующем. Пусть  $z_1$  и  $z_2$  — решения (возможно, комплексные) уравнения

$$1 - c_1 z - c_2 z^2 = 0. \quad (15)$$

Мы можем переписать уравнение (15) в виде

$$0 = 1 - c_1 z - c_2 z^2 = (z_1 - z)(z_2 - z),$$

и, сравнивая последнее уравнение с (14), получаем

$$\lambda_1 = z_1^{-1} \text{ и } \lambda_2 = z_2^{-1}.$$

Многочлен в уравнении (14) можно обратить, если  $|\lambda_1| < 1$  и  $|\lambda_2| < 1$ . Эти условия эквивалентны требованию, чтобы все корни многочлена в уравнении (15) лежали вне единичного круга:

$$|z| > 1.$$

Если это условие выполняется, то

$$(1 - c_1 L - c_2 L^2)^{-1} = (1 - \lambda_1 L)^{-1} (1 - \lambda_2 L)^{-1} = \left( \sum_{j=0}^{\infty} \lambda_1^j L^j \right) \left( \sum_{j=0}^{\infty} \lambda_2^j L^j \right).$$

С другой стороны, можно записать

$$\frac{1}{1 - \lambda_1 L} \frac{1}{1 - \lambda_2 L} = \frac{1}{\lambda_1 - \lambda_2} \left( \frac{\lambda_1}{1 - \lambda_1 L} - \frac{\lambda_2}{1 - \lambda_2 L} \right).$$

Тогда

$$(1 - c_1 L - c_2 L^2)^{-1} = \frac{1}{\lambda_1 - \lambda_2} \sum_{j=0}^{\infty} (\lambda_1 \lambda_1^j L^j - \lambda_2 \lambda_2^j L^j) = \frac{1}{\lambda_1 - \lambda_2} \sum_{j=0}^{\infty} (\lambda_1^{j+1} - \lambda_2^{j+1}) L^j.$$

Этот результат может быть обобщен на случай многочлена порядка  $p$ ,

$$C(L) = 1 - c_1 L - \dots - c_p L^p,$$

так как мы можем разложить его как

$$1 - c_1 L - \dots - c_p L^p = \prod_{j=1}^p (1 - \lambda_j L),$$

где числа  $\lambda_j$  удовлетворяют

$$1 - c_1 z - \dots - c_p z^p = \prod_{j=1}^p (1 - \lambda_j z).$$

Многочлен можно обратить, если корни  $1 - c_1 z - \dots - c_p z^p$  лежат вне единичного круга. Тогда

$$(1 - c_1 L - \dots - c_p L^p)^{-1} = \prod_{j=1}^p (1 - \lambda_j L)^{-1}. \quad (16)$$

### 3.3 Некоторые обобщения на случай многомерных процессов

До настоящего момента все результаты были сформулированы для одномерных процессов, однако они могут быть обобщены на векторный случай, как показано далее. Предположим, что векторный процесс  $\{X_t\}$  слабостационарен, и  $\mathbb{E}X_t = 0$  для всех  $t$ . Пусть

$$\Gamma(j) = \mathbb{E}X_t X_{t-j}'.$$

Заметим, что для того, чтобы  $\{X_t\}$  был слабостационарным, матрица  $\Gamma(j)$  не обязана быть симметричной для  $j \neq 0$ , однако

$$\Gamma(j) = \Gamma(-j)'.$$

Рассмотрим последовательность  $k$ -мерных матриц  $\{C_j\}$  и определим

$$Y_t = \sum_{j=0}^{\infty} C_j X_{t-j}.$$

Дисперсия величины  $Y_t$  равна

$$\begin{aligned} \mathbb{E}Y_t Y_t' &= \mathbb{E} \left( \sum_{i=0}^{\infty} C_i X_{t-i} \right) \left( \sum_{j=0}^{\infty} C_j X_{t-j} \right)' = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} C_i \Gamma(j-i) C_j' \\ &= \sum_{j=0}^{\infty} C_j \Gamma(0) C_j' + \sum_{h=1}^{\infty} \sum_{j=0}^{\infty} (C_j \Gamma(h) C_{j+h}' + C_{j+h} \Gamma(h)' C_j'). \end{aligned}$$

Пусть  $\|A\| = \sqrt{\text{tr}(A'A)}$ . Имеем:

$$\begin{aligned} \|\mathbb{E}Y_t Y_t'\| &\leq \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \|C_i \Gamma(j-i) C_j'\| \leq \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \|C_i\| \|C_j\| \|\Gamma(j-i)\| \\ &\leq 2 \sum_{h=0}^{\infty} \|\Gamma(h)\| \left( \sum_{j=0}^{\infty} \|C_j\|^2 \right)^2, \end{aligned}$$

где последнее неравенство доказывается так же, как (4). Значит, матрица  $\mathbb{E}Y_t Y_t'$  конечна при условиях

$$\sum_{j=0}^{\infty} \|C_j\|^2 < \infty \tag{17}$$

и

$$\sum_{h=-\infty}^{\infty} \|\Gamma(h)\| < \infty. \tag{18}$$

Процесс  $\{\varepsilon_t\}$  называется  $k$ -мерным БШ, если  $\mathbb{E}\varepsilon_t = 0$  и  $\mathbb{E}\varepsilon_t \varepsilon_t' = \Sigma$  — положительно определенная матрица для всех  $t$ , и  $\mathbb{E}\varepsilon_t \varepsilon_s' = 0$  для  $t \neq s$ . Если  $\{X_t\}$  — слабостационарный чисто недетерминистический процесс с нулевым средним, такой что выполнено условие (18), у него имеется, аналогично скалярному случаю,  $MA(\infty)$ -представление:

$$X_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j},$$

где  $\{\varepsilon_t\}$  — векторный БШ, матрицы  $C_j$  удовлетворяют условию (17), и

$$C_0 = I_k.$$

Снова, как и в скалярном случае, величины  $\varepsilon_t$  являются ошибками линейного предсказания на один шаг вперед.

Спектральная плотность векторного процесса определяется аналогично скалярному случаю:

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma(j) e^{-i\lambda j} = \frac{1}{2\pi} \left( \Gamma(0) + \sum_{j=1}^{\infty} (\Gamma(j) e^{-i\lambda j} + \Gamma(j)' e^{i\lambda j}) \right),$$

Долгосрочная ковариационная матрица снова равна

$$\Omega = 2\pi f(0).$$

Чтобы долгосрочная дисперсия была конечна, достаточно

$$\sum_{j=0}^{\infty} \|C_j\| < \infty.$$

Для векторного БШ  $\{\varepsilon_t\}$  спектральная плотность постоянна:

$$f_\varepsilon(\lambda) = \frac{1}{2\pi} \Sigma.$$

Пусть  $\{X_t\}$  — слабостационарный процесс со спектральной плотностью  $f_X$ . Определим

$$Y_t = \sum_{j=0}^{\infty} C_j X_{t-j}.$$

Тогда спектральная плотность процесса  $\{Y_t\}$  равна

$$f_Y(\lambda) = \left( \sum_{j=0}^{\infty} C_j e^{-i\lambda j} \right) f_X(\lambda) \left( \sum_{j=0}^{\infty} C_j' e^{i\lambda j} \right).$$

Пусть  $\{C_j\}$  — последовательность матриц порядка  $k$ . Многочлен  $C(L)$  от оператора сдвига в векторном случае определяется как

$$C(L) = I_k + C_1 L + C_2 L^2 + \dots,$$

где мы положили  $C_0 = I_k$  в соответствии с разложением Вольда. Говорят, что

$$B(L) = C(L)^{-1}$$

если

$$B(L) C(L) = I_k.$$

Многочлен  $C(L)$  обратим, если корни уравнения

$$\det(C(z)) = 0$$

лежат вне единичного круга. Например, если многочлен  $C(L)$  имеет порядок  $p$ , то все корни уравнения

$$\det(I + C_1 z + \dots + C_p z^p) = 0$$

должны быть по модулю больше единицы.

## 4 Модель ARMA

В этом разделе для скалярного случая иллюстрируются некоторые понятия предыдущего раздела. Пусть  $\{\varepsilon_t\}$  — БШ с  $\mathbb{V}(\varepsilon_t) = \sigma^2$ . Процесс  $MA(q)$ , скажем  $\{X_t\}$ , имеет вид:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} = \Theta(L) \varepsilon_t,$$

где

$$\Theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q.$$

По теореме 2, процесс  $MA(q)$  имеет спектральную плотность

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \Theta(e^{-i\lambda}) \right|^2 = \frac{\sigma^2}{2\pi} \left| 1 + \theta_1 e^{-i\lambda} + \dots + \theta_q e^{-i\lambda q} \right|^2,$$

и долгосрочную дисперсию

$$\omega_X = 2\pi f_X(0) = \sigma^2 |\Theta(1)|^2 = \sigma^2 (1 + \theta_1 + \dots + \theta_q)^2.$$

Заметим, что для некоторых значений  $\theta_j$ , долгосрочная дисперсия может быть равна нулю. Для  $0 \leq j \leq q$  эффект шока в период  $t$  на  $X$  через  $j$  периодов равен

$$\theta_j = \frac{\partial X_{t+j}}{\partial \varepsilon_t},$$

и влияние шока нулевой после  $q$  периодов.

**Определение 7 (AR)** Процесс  $\{X_t\}$  называется авторегрессией порядка  $p$ , обозначаемой как  $AR(p)$ , если он имеет вид

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

где  $\{\varepsilon_t\}$  есть БШ.

Пусть

$$\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p.$$

Тогда процесс  $AR(p)$  может быть переписан в виде

$$\Phi(L) X_t = \varepsilon_t.$$

Если корни многочлена  $\Phi(z)$  лежат вне единичного круга,  $\Phi(L)$  можно обратить, и процесс получает следующее  $MA(\infty)$ -представление (разложение Вольда):

$$X_t = \Phi(L)^{-1} \varepsilon_t = \Psi(L) \varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \tag{19}$$

для некоторого многочлена  $\Psi(L)$ .

Пусть  $p = 1$ . Тогда

$$(1 - \phi_1 L) X_t = \varepsilon_t,$$

и, в соответствии с уравнением (13),

$$X_t = (1 - \phi_1 L)^{-1} \varepsilon_t = \left( \sum_{j=0}^{\infty} \phi_1^j L^j \right) \varepsilon_t = \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}.$$

Таким образом, в случае  $AR(1)$ -процесса коэффициенты многочлена  $\Psi(L)$  в соотношении (19) равны  $\psi_j = \phi_1^j$ . Проверим условие суммируемости квадратов из теоремы 2:

$$\sum_{j=0}^{\infty} \psi_j^2 = \sum_{j=0}^{\infty} \phi_1^{2j} = \frac{1}{1 - \phi_1^2} < \infty$$

при условии  $|\phi_1| < 1$ . Из теоремы 2 следует, что процесс  $AR(1)$  слабостационарен, и его спектральная плотность и долгосрочная дисперсия равны

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} \phi_1^j e^{-i\lambda j} \right|^2 = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \phi_1 e^{-i\lambda}|^2},$$

$$\omega_X = 2\pi f_X(0) = \frac{\sigma^2}{(1 - \phi_1)^2}.$$

Долгосрочная дисперсия стационарного  $AR(1)$ -процесса также конечна. Коэффициенты в  $MA(\infty)$ -представлении стационарного  $AR(1)$ -процесса абсолютно суммируемы:

$$\sum_{j=0}^{\infty} |\psi_j| = \sum_{j=0}^{\infty} |\phi_1|^j = \frac{1}{1 - |\phi_1|} < \infty.$$

Это следует из того факта, что  $\psi_j = \phi_1^j \rightarrow 0$  при  $j \rightarrow \infty$  с экспоненциальной скоростью.

В случае  $AR(p)$ ,  $p > 1$ , в силу (16),  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ , если все корни многочлена  $\Phi(z)$  лежат вне единичного круга. Тогда  $AR(p)$ -процесс слабостационарен со спектральной плотностью и долгосрочной дисперсией

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|\Phi(e^{-i\lambda})|^2},$$

$$\omega_X = \frac{\sigma^2}{\Phi(1)^2}.$$

**Определение 8 (ARMA)** Процесс  $\{X_t\}$  называется  $ARMA(p, q)$ -процессом, если он имеет вид

$$\Phi(L) X_t = \Theta(L) \varepsilon_t,$$

где  $\{\varepsilon_t\}$  — БШ.

Если корни многочлена  $\Phi(z)$  лежат вне единичного круга, то  $ARMA(p, q)$ -процесс имеет  $MA(\infty)$ -представление

$$X_t = \Phi(L)^{-1} \Theta(L) \varepsilon_t.$$

Его спектральная плотность и долгосрочная дисперсия равны

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\Theta(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2},$$

$$\omega_X = \sigma^2 \frac{|\Theta(1)|^2}{|\Phi(1)|^2}.$$

Если корни многочлена  $\Theta(z)$  лежат вне единичного круга, то процесс  $ARMA(p, q)$  имеет  $AR(\infty)$ -представление

$$\Theta(L)^{-1} \Phi(L) X_t = \varepsilon_t.$$



Таким образом, на практике любой слабостационарный процесс  $ARMA(p, q)$  или  $MA(\infty)$  может быть аппроксимирован моделью  $AR(m_n)$ , с порядком  $m_n$ , увеличивающимся с  $n$ , с меньшей, впрочем, скоростью.

ARMA-процесс с ненулевым средним  $\mu$  может быть записан в виде

$$\Phi(L)(X_t - \mu) = \Theta(L)\varepsilon_t,$$

или, эквивалентно, как

$$\Phi(L)X_t = \alpha + \Theta(L)\varepsilon_t.$$

В последнем уравнении  $\alpha$  может рассматриваться как свободный член в модели линейной регрессии, так как уравнение может быть представлено в виде

$$X_t = \alpha + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \Theta(L)\varepsilon_t.$$

Заметим, что  $\mu$  и  $\alpha$  связаны соотношением

$$\mu = \frac{\alpha}{\Phi(1)} = \frac{\alpha}{1 - \phi_1 - \dots - \phi_p},$$

если многочлен  $\Phi(L)$  обратим.

## 5 Асимптотические результаты

В настоящем разделе мы обсудим ЗБЧ и ЦПТ для линейных процессов  $\{X_t\}$  вида

$$X_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} = C(L)\varepsilon_t, \tag{20}$$

где

$$C(L) = \sum_{j=0}^{\infty} c_j L^j.$$

Соответствующий метод был разработан в работе Phillips& Solo (1992). Используя подход авторов и предполагая, что  $\{\varepsilon_t\}$  — последовательность независимых одинаково распределенных (н.о.р.) случайных величин с нулевым средним и конечной дисперсией, можно вывести ЗБЧ и ЦПТ для серийно коррелированного линейного процесса  $\{X_t\}$ , опираясь только на ЗБЧ и ЦПТ для н.о.р.-последовательностей и некоторое алгебраическое разложение многочлена  $C(L)$ .

Этот метод работает также для более широкого класса последовательностей инноваций  $\{\varepsilon_t\}$ . Например, вместо предположения о н.о.р. последовательности  $\{\varepsilon_t\}$  можно работать со случаем независимости, но неодинаковой распределенности (н.н.о.р.). В этом случае необходимы следующие сильные версии ЗБЧ (СЗБЧ) и ЦПТ.

**Лемма 3 (СЗБЧ для н.н.о.р.-последовательностей)** Пусть  $\{\varepsilon_t\}$  — последовательность независимых случайных величин, таких что  $\sup_t \mathbb{E}|\varepsilon_t|^{1+\delta} < \infty$  для некоторого  $\delta > 0$ . Тогда  $n^{-1} \sum_{t=1}^n \varepsilon_t - n^{-1} \sum_{t=1}^n \mathbb{E}\varepsilon_t \rightarrow_{a.s.} 0$ .

См. White (2001, следствие 3.9).

**Лемма 4 (ЦПТ для н.н.о.р.-последовательностей)** Пусть  $\{\varepsilon_t\}$  — последовательность независимых случайных величин, таких что  $\mathbb{E}\varepsilon_t = 0$  для всех  $t$ ,  $\sup_t \mathbb{E}|\varepsilon_t|^{2+\delta} < \infty$  для некоторого  $\delta > 0$ , и для всех достаточно больших  $n$  выполнено  $n^{-1} \sum_{t=1}^n \mathbb{E}\varepsilon_t^2 > \delta' > 0$ . Тогда  $n^{-1/2} \sum_{t=1}^n \varepsilon_t / (n^{-1} \sum_{t=1}^n \mathbb{E}\varepsilon_t^2)^{1/2} \rightarrow_d N(0, 1)$ .

См. White (2001, теорема 5.10).

С другой стороны, если нельзя предположить, что инновации  $\{\varepsilon_t\}$  независимы, можно опираться на ЗБЧ и ЦПТ для мартингал-разностей.

**Определение 9 (Мартингал-разность)** Пусть  $\{\mathcal{F}_t\}$  — возрастающая последовательность  $\sigma$ -алгебр. Тогда  $\{(u_t, \mathcal{F}_t)\}$  называется мартингалом, если  $\mathbb{E}(u_t | \mathcal{F}_{t-1}) = u_{t-1}$  почти наверное. Последовательность  $\{(\varepsilon_t, \mathcal{F}_t)\}$  называется мартингал-разностью, если  $\mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1}) = 0$  почти наверное.

Если процесс  $\{u_t\}$  задан,  $\sigma$ -алгебра  $\mathcal{F}_t$  часто определяется как  $\sigma(u_t, u_{t-1}, \dots)$ . Пусть  $\{(u_t, \mathcal{F}_t)\}$  — мартингал. Тогда  $\{(u_t - u_{t-1}, \mathcal{F}_t)\}$  — мартингал-разность, так как  $u_t - u_{t-1} = u_t - \mathbb{E}(u_t | \mathcal{F}_{t-1})$ , и, следовательно,  $\mathbb{E}(u_t - u_{t-1} | \mathcal{F}_{t-1}) = 0$ . Легко видеть, что мартингал-разность  $\{\varepsilon_t\}$  серийно некоррелирована:  $\mathbb{E}\varepsilon_t \varepsilon_{t-1} = \mathbb{E}(\varepsilon_{t-1} \mathbb{E}(\varepsilon_t | \mathcal{F}_{t-1})) = 0$ . В то же время величины  $\varepsilon$  могут не быть независимыми.

Если  $\{\varepsilon_t\}$  — мартингал-разность, следующие СЗБЧ и ЦПТ могут быть полезны.

**Лемма 5 (СЗБЧ для мартингал-разностей)** Пусть  $\{(\varepsilon_t, \mathcal{F}_t)\}$  — мартингал-разность, такая что  $\sup_t \mathbb{E}|\varepsilon_t|^{2+\delta} < \infty$  для некоторого  $\delta > 0$ . Тогда  $n^{-1} \sum_{t=1}^n \varepsilon_t \rightarrow_{a.s.} 0$ .

См. White (2001, упражнение 3.77).

**Лемма 6 (ЦПТ для мартингал-разностей)** Пусть  $\{(\varepsilon_t, \mathcal{F}_t)\}$  — мартингал-разность, такая что  $\sup_t \mathbb{E}|\varepsilon_t|^{2+\delta} < \infty$  для некоторого  $\delta > 0$ . Пусть для всех достаточно больших  $n$  выполнено  $n^{-1} \sum_{t=1}^n \mathbb{E}\varepsilon_t^2 > \delta' > 0$ , и пусть  $n^{-1} \sum_{t=1}^n \varepsilon_t^2 - n^{-1} \sum_{t=1}^n \mathbb{E}\varepsilon_t^2 \rightarrow_p 0$ . Тогда  $n^{-1/2} \sum_{t=1}^n \varepsilon_t / (n^{-1} \sum_{t=1}^n \mathbb{E}\varepsilon_t^2)^{1/2} \rightarrow_d N(0, 1)$ .

См. White (2001, следствие 5.26).

**Лемма 7 (ЦПТ для строго стационарных и эргодичных мартингал-разностей)** Пусть  $\{(\varepsilon_t, \mathcal{F}_t)\}$  — строго стационарная и эргодичная мартингал-разность, такая что  $\mathbb{E}\varepsilon_t^2 < \infty$ . Тогда  $n^{-1/2} \sum_{t=1}^n \varepsilon_t \rightarrow_d N(0, \mathbb{E}\varepsilon_t^2)$ .

Заметим, что все рассмотренные альтернативные требования к процессу  $\{\varepsilon_t\}$ , н.о.р., н.н.о.р. или мартингал-разность, сильнее простого БШ. Таким образом, для использования указанного метода необходимо немного усилить МА( $\infty$ )-представление по сравнению с результатом, даваемым разложением Вольда.

## 5.1 БН-разложение

Обсудим сначала алгебраическое разложение многочлена от оператора сдвига на долгосрочную и переходную компоненты. Такое разложение было предложено в работе Beveridge & Nelson (1981).

**Лемма 8** Пусть  $C(L) = \sum_{j=0}^{\infty} c_j L^j$ . Тогда

(a)  $C(L) = C(1) - (1-L)\tilde{C}(L)$ , где  $\tilde{C}(L) = \sum_{j=0}^{\infty} \tilde{c}_j L^j$  с  $\tilde{c}_j = \sum_{h=j+1}^{\infty} c_h$ .

(b) Если  $\sum_{j=1}^{\infty} j^{1/2} |c_j| < \infty$ , то  $\sum_{j=0}^{\infty} \tilde{c}_j^2 < \infty$ .

(c) Если  $\sum_{j=1}^{\infty} j |c_j| < \infty$ , то  $\sum_{j=0}^{\infty} |\tilde{c}_j| < \infty$ .

Здесь  $C(1)$  — долгосрочная компонента, а  $(1-L)\tilde{C}(L)$  считается переходной компонентой по причинам, которые скоро станут ясны, в частности, из уравнений (22) и (23).

**Доказательство.** Для доказательства пункта (a), запишем

$$\begin{aligned} \sum_{j=0}^{\infty} c_j L^j &= \sum_{j=0}^{\infty} c_j - \sum_{j=1}^{\infty} c_j + \left( \sum_{j=1}^{\infty} c_j - \sum_{j=2}^{\infty} c_j \right) L + \left( \sum_{j=2}^{\infty} c_j - \sum_{j=3}^{\infty} c_j \right) L^2 \\ &\quad + \dots + \left( \sum_{j=h}^{\infty} c_j - \sum_{j=h+1}^{\infty} c_j \right) L^h + \dots \end{aligned}$$

Переставляя местами слагаемые, получим

$$\begin{aligned} \sum_{j=0}^{\infty} c_j L^j &= \sum_{j=0}^{\infty} c_j - (1-L) \sum_{j=1}^{\infty} c_j - (1-L) \sum_{j=2}^{\infty} c_j L - \dots - (1-L) \sum_{j=h+1}^{\infty} c_j L^h - \dots \\ &= \sum_{j=0}^{\infty} c_j - (1-L) \sum_{j=0}^{\infty} \left( \sum_{h=j+1}^{\infty} c_h \right) L^j = C(1) - (1-L) \tilde{C}(L). \end{aligned}$$

Для доказательства пункта (b),

$$\begin{aligned} \sum_{j=0}^{\infty} \tilde{c}_j^2 &= \sum_{j=0}^{\infty} \left( \sum_{h=j+1}^{\infty} c_h \right)^2 \leq \sum_{j=0}^{\infty} \left( \sum_{h=j+1}^{\infty} |c_h| \right)^2 = \sum_{j=0}^{\infty} \left( \sum_{h=j+1}^{\infty} |c_h|^{1/2} h^{1/4} |c_h|^{1/2} h^{-1/4} \right)^2 \\ &\leq \sum_{j=0}^{\infty} \left( \sum_{h=j+1}^{\infty} |c_h| h^{1/2} \right) \left( \sum_{h=j+1}^{\infty} |c_h| h^{-1/2} \right) \leq \left( \sum_{h=0}^{\infty} |c_h| h^{1/2} \right) \sum_{j=0}^{\infty} \sum_{h=j+1}^{\infty} |c_h| h^{-1/2}. \end{aligned}$$

Далее, рассмотрим  $\sum_{j=0}^{\infty} \sum_{h=j+1}^{\infty} |c_h| h^{-1/2}$ . Заметим, что  $|c_1|$  входит в сумму только однажды, когда  $j=0$ ,  $|c_2|$  появляется в сумме дважды, когда  $j=0, 1$ , и т.д. Таким образом,  $|c_h|$  входит в сумму для  $j=0, 1, \dots, h-1$ , или  $h$  раз. Следовательно,

$$\sum_{j=0}^{\infty} \sum_{h=j+1}^{\infty} |c_h| h^{-1/2} = \sum_{j=0}^{\infty} |c_j| j^{-1/2} j = \sum_{j=0}^{\infty} |c_j| j^{1/2},$$

и

$$\sum_{j=0}^{\infty} \tilde{c}_j^2 \leq \left( \sum_{j=0}^{\infty} |c_j| j^{1/2} \right)^2.$$

Для доказательства пункта (c),

$$\sum_{j=0}^{\infty} |\tilde{c}_j| = \sum_{j=0}^{\infty} \left| \sum_{h=j+1}^{\infty} c_h \right| \leq \sum_{j=0}^{\infty} \sum_{h=j+1}^{\infty} |c_h| = \sum_{j=0}^{\infty} |c_j| j.$$

■

Заметим, что предположения  $\sum_{j=1}^{\infty} j^{1/2} |c_j| < \infty$  и  $\sum_{j=1}^{\infty} j |c_j| < \infty$  сильнее, чем конечность долгосрочной дисперсии  $\sum_{j=1}^{\infty} |c_j| < \infty$ .

Согласно БН-разложению, если процесс  $\{X_t\}$  линейен, то

$$X_t = C(L) \varepsilon_t = C(1) \varepsilon_t - (1-L) \tilde{C}(L) \varepsilon_t = C(1) \varepsilon_t - (\tilde{\varepsilon}_t - \tilde{\varepsilon}_{t-1}), \quad (21)$$

где

$$\tilde{\varepsilon}_t = \tilde{C}(L) \varepsilon_t.$$

Более того,  $\tilde{\varepsilon}_t$  имеет конечную дисперсию, если  $\sum_{j=0}^{\infty} j^{1/2} |c_j| < \infty$ . Первое слагаемое в правой части уравнения (21),  $C(1)\varepsilon_t$ , называется *долгосрочной компонентой*, а  $\tilde{\varepsilon}_t - \tilde{\varepsilon}_{t-1}$  — *переходной компонентой*.

Аналогичное разложение справедливо в векторном случае. Переходная компонента имеет конечную дисперсию, если  $\sum_{j=1}^{\infty} j^{1/2} \|C_j\| < \infty$ :

$$\begin{aligned} \sum_{j=0}^{\infty} \|\tilde{C}_j\|^2 &= \sum_{j=0}^{\infty} \left\| \sum_{h=j+1}^{\infty} C_h \right\|^2 \leq \sum_{j=0}^{\infty} \left( \sum_{h=j+1}^{\infty} \|C_h\| \right)^2 \\ &= \sum_{j=0}^{\infty} \left( \sum_{h=j+1}^{\infty} \|C_h\|^{1/2} h^{1/4} \|C_h\|^{1/2} h^{-1/4} \right)^2 \leq \left( \sum_{j=0}^{\infty} j^{1/2} \|C_j\| \right)^2. \end{aligned}$$

Условие пункта (с) теоремы 8 преобразуется в  $\sum_{j=1}^{\infty} j \|C_j\| < \infty$  в векторном случае.

## 5.2 ЗБЧ для линейных процессов

В настоящем разделе мы покажем, как получить ЗБЧ для линейного процесса с помощью БН-разложения, следуя подходу Филлипса и Соло.

**Теорема 9 (ЗБЧ для линейных процессов с н.о.р. инновациями)** Пусть  $X_t = C(L)\varepsilon_t$ ,  $\{\varepsilon_t\}$  — последовательность н.о.р. случайных величин с  $\mathbb{E}|\varepsilon_t| < \infty$  и  $\mathbb{E}\varepsilon_t = 0$ , и пусть для  $C(L)$  выполнено  $\sum_{j=1}^{\infty} j |c_j| < \infty$ . Тогда  $n^{-1} \sum_{t=1}^n X_t \rightarrow_p 0$ .

**Доказательство.** Ключевой элемент доказательства — БН-разложение (21). Сначала заметим, что  $\sum_{t=1}^n (\tilde{\varepsilon}_t - \tilde{\varepsilon}_{t-1})$  является так называемой телескопической суммой:

$$\begin{aligned} \sum_{t=1}^n (\tilde{\varepsilon}_t - \tilde{\varepsilon}_{t-1}) &= (\tilde{\varepsilon}_1 - \tilde{\varepsilon}_0) + (\tilde{\varepsilon}_2 - \tilde{\varepsilon}_1) + (\tilde{\varepsilon}_3 - \tilde{\varepsilon}_2) + \dots + (\tilde{\varepsilon}_n - \tilde{\varepsilon}_{n-1}) \\ &= -\tilde{\varepsilon}_0 + \tilde{\varepsilon}_n, \end{aligned} \quad (22)$$

так что

$$n^{-1} \sum_{t=1}^n X_t = C(1) n^{-1} \sum_{t=1}^n \varepsilon_t - n^{-1} (\tilde{\varepsilon}_n - \tilde{\varepsilon}_0). \quad (23)$$

В слагаемом, соответствующем долгосрочной компоненте, происходит накопление шоков, а в переходной компоненте шоки взаимоуничтожаются.

Так как  $\sum_{j=1}^{\infty} j |c_j| < \infty$ , мы получаем, что  $|C(1)| < \infty$ . Следовательно, используя, например, слабый ЗБЧ (СлЗБЧ) для н.о.р.-последовательностей, получаем, что

$$C(1) n^{-1} \sum_{t=1}^n \varepsilon_t \rightarrow_p 0.$$

Осталось показать, что вторым слагаемым (соответствующим переходной компоненте) в уравнении (23) можно асимптотически пренебречь:

$$\mathbb{P} \{n^{-1} |\tilde{\varepsilon}_t| > \delta\} \leq \frac{\mathbb{E} |\tilde{\varepsilon}_t|}{n\delta}$$

и

$$\mathbb{E} |\tilde{\varepsilon}_t| = \mathbb{E} \left| \sum_{j=0}^{\infty} \tilde{c}_j \varepsilon_{t-j} \right| \leq \sum_{j=0}^{\infty} |\tilde{c}_j| \mathbb{E} |\varepsilon_{t-j}| = \mathbb{E} |\varepsilon_0| \sum_{j=0}^{\infty} |\tilde{c}_j| < \infty \quad (24)$$

при условии  $\sum_{j=1}^{\infty} j |c_j| < \infty$ . Следовательно,

$$n^{-1} (\tilde{\varepsilon}_n - \tilde{\varepsilon}_0) \rightarrow_p 0.$$

■

Если же предположить, что  $\mathbb{E}\varepsilon_t^2 < \infty$ , то можно заменить  $\sum_{j=1}^{\infty} j |c_j| < \infty$  на  $\sum_{j=1}^{\infty} j^{1/2} |c_j| < \infty$ , так как

$$P(n^{-1} |\tilde{\varepsilon}_t| > \delta) \leq \frac{\mathbb{E}\tilde{\varepsilon}_t^2}{n^2 \delta^2},$$

и  $\mathbb{E}\tilde{\varepsilon}_t^2 < \infty$  при условии  $\sum_{j=1}^{\infty} j^{1/2} |c_j| < \infty$ .

Можно доказать аналогичный ЗБЧ в предположении, что  $\{\varepsilon_t\}$  — последовательность н.о.р. величин или мартингал-разностей, с помощью соответствующих ЗБЧ (леммы 3 и 5 соответственно). Например, аналогичный результат имеет место, если  $\{\varepsilon_t\}$  н.о.р.,  $\sup_t \mathbb{E} |\varepsilon_t|^{1+\delta} < \infty$  для некоторого  $\delta > 0$  и  $\sum_{j=1}^{\infty} j |c_j| < \infty$ . Если  $\{(\varepsilon_t, \mathcal{F}_t)\}$  — мартингал-разность, то аналогичный результат верен при условиях  $\sup_t \mathbb{E} |\varepsilon_t|^{2+\delta} < \infty$  для некоторого  $\delta > 0$  и  $\sum_{j=1}^{\infty} j^{1/2} |c_j| < \infty$ .

До настоящего момента мы предполагали, что  $\mathbb{E}X_t = 0$ . Можно модифицировать это предположение так, что  $X_t = \mu + C(L)\varepsilon_t$ , где  $\mu$  — среднее процесса  $X_t$ . В этом случае, при тех же условиях имеем  $n^{-1} \sum_{t=1}^n X_t \rightarrow_p \mu$ . Например,  $AR(1)$ -процесс со средним  $\mu$  определяется как  $(1 - \phi L)(X_t - \mu) = \varepsilon_t$ . Если  $|\phi| < 1$ , то выборочное среднее процесса  $X_t$  сходится по вероятности к  $\mu$ , если выполнены соответствующим условия на  $\{\varepsilon_t\}$ .

### 5.3 ЦПТ для линейных процессов

Далее показано, как базовая ЦПТ для н.о.р. случайных величин может быть обобщена на случай линейных процессов с помощью подхода Филлипса–Соло.

**Теорема 10 (ЦПТ для линейных процессов с н.о.р. инновациями)** Пусть  $X_t = C(L)\varepsilon_t$ ,  $\{\varepsilon_t\}$  — последовательность н.о.р. случайных величин с  $\mathbb{E} |\varepsilon_t|^2 = \sigma^2 < \infty$  и  $\mathbb{E}\varepsilon_t = 0$ ,  $C(L)$  удовлетворяет условию  $\sum_{j=1}^{\infty} j^{1/2} |c_j| < \infty$  и  $C(1) \neq 0$ . Тогда  $n^{-1/2} \sum_{t=1}^n X_t \rightarrow_d N(0, \sigma^2 C(1)^2)$ .

**Доказательство.** БН-разложение позволяет записать

$$\begin{aligned} n^{-1/2} \sum_{t=1}^n X_t &= C(1) n^{-1/2} \sum_{t=1}^n \varepsilon_t - n^{-1/2} (\tilde{\varepsilon}_n - \tilde{\varepsilon}_0) = C(1) n^{-1/2} \sum_{t=1}^n \varepsilon_t + o_p(1) \\ &\rightarrow_d C(1) N(0, \sigma^2) \stackrel{d}{=} N(0, \sigma^2 C(1)^2). \end{aligned}$$

Здесь имеет место сходимость по распределению в силу ЦПТ для н.о.р. случайных величин.

■

Использованный подход ясно демонстрирует, почему в случае наличия серийной корреляции асимптотическая дисперсия среднего зависит от долгосрочной дисперсии процесса. Как и ранее, этот подход может быть обобщен на случай, когда  $\{\varepsilon_t\}$  — последовательность н.о.р. величин или мартингал-разностей.

В векторном случае, пусть  $\{\varepsilon_t\}$  — последовательность н.о.р.  $k$ -мерных векторов с  $\mathbb{E}\varepsilon_t = 0$ , и  $\mathbb{E}\varepsilon_t \varepsilon_t' = \Sigma$  — конечная матрица. Пусть  $X_t = C(L)\varepsilon_t$  и  $\sum_{j=0}^{\infty} j^{1/2} \|C_j\| < \infty$ ,  $C(1) \neq 0$ . Так как  $n^{-1/2} \sum_{t=1}^n \varepsilon_t \rightarrow_d N(0, \Sigma)$ , получаем, что

$$n^{-1/2} \sum_{t=1}^n X_t \rightarrow_d N(0, C(1) \Sigma C(1)').$$

#### 5.4 Сходимость выборочных дисперсий

Рассмотрим простую модель линейной регрессии  $Y_t = \beta X_t + U_t$ . МНК-оценка  $\beta$  равна  $\hat{\beta}_n = \sum_{t=1}^n X_t Y_t / \sum_{t=1}^n X_t^2 = \beta + \sum_{t=1}^n X_t U_t / \sum_{t=1}^n X_t^2$ . Таким образом, в работе с асимптотическими свойствами МНК-оценки необходимо иметь дело со вторыми выборочными моментами  $n^{-1} \sum_{t=1}^n X_t^2$ . В настоящем разделе мы обсудим, как можно установить сходимость вторых выборочных моментов в случае, когда процесс  $\{X_t\}$  линеен.

Предположим, что  $\{X_t\}$  — скалярный линейный процесс, удовлетворяющий условиям теоремы 10. Запишем

$$\begin{aligned} X_t^2 &= (C(L)\varepsilon_t)^2 = \left( \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} \right)^2 = \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} c_j c_l \varepsilon_{t-j} \varepsilon_{t-l} \\ &= \sum_{j=0}^{\infty} c_j^2 \varepsilon_{t-j}^2 + 2 \sum_{j=0}^{\infty} \sum_{l>j} c_j c_l \varepsilon_{t-j} \varepsilon_{t-l} \\ &= \sum_{j=0}^{\infty} c_j^2 \varepsilon_{t-j}^2 + 2 \sum_{h=1}^{\infty} \sum_{j=0}^{\infty} c_j c_{j+h} \varepsilon_{t-j} \varepsilon_{t-j-h} \quad (\text{замена } l = j + h, \text{ так что } h = 1, 2, \dots) \\ &= B_0(L)\varepsilon_t^2 + 2 \sum_{h=1}^{\infty} B_h(L) \varepsilon_t \varepsilon_{t-h}, \end{aligned}$$

где для  $h = 0, 1, \dots$  выполнено

$$B_h(L) = \sum_{j=0}^{\infty} b_{h,j} L^j = \sum_{j=0}^{\infty} c_j c_{j+h} L^j.$$

Таким образом,

$$\begin{aligned} B_0(L) &= \sum_{j=0}^{\infty} b_{0,j} L^j = \sum_{j=0}^{\infty} c_j^2 L^j, \\ B_1(L) &= \sum_{j=0}^{\infty} b_{1,j} L^j = \sum_{j=0}^{\infty} c_j c_{j+1} L^j, \end{aligned}$$

и т.д. БН-разложение многочлена  $B_h(L)$  есть

$$B_h(L) = B_h(1) - (1-L)\tilde{B}_h(L), \quad (25)$$

где

$$\tilde{B}_h(L) = \sum_{j=0}^{\infty} \tilde{b}_{h,j} L^j, \quad \tilde{b}_{h,j} = \sum_{l=j+1}^{\infty} b_{h,l} = \sum_{l=j+1}^{\infty} c_l c_{l+h}.$$

БН-разложение многочлена  $B_h(L)$  верно, если  $\sum_{j=1}^{\infty} j^{1/2} |c_j| < \infty$ :

$$\begin{aligned} \sum_{j=0}^{\infty} \tilde{b}_{h,j}^2 &= \sum_{j=0}^{\infty} \left( \sum_{l=j+1}^{\infty} c_l c_{l+h} \right)^2 = \sum_{j=0}^{\infty} \left( \sum_{l=j+1}^{\infty} l^{1/4} c_l c_{l+h} l^{-1/4} \right)^2 \\ &\leq \sum_{j=0}^{\infty} \left( \sum_{l=j+1}^{\infty} l^{1/2} c_l^2 \right) \left( \sum_{l=j+1}^{\infty} c_{l+h}^2 l^{-1/2} \right) \leq \left( \sum_{l=1}^{\infty} l^{1/2} c_l^2 \right) \left( \sum_{j=0}^{\infty} \sum_{l=j+1}^{\infty} c_{l+h}^2 l^{-1/2} \right) \\ &= \left( \sum_{l=1}^{\infty} l^{1/2} c_l^2 \right) \left( \sum_{l=0}^{\infty} c_{l+h}^2 l^{1/2} \right) \leq \left( \sum_{l=1}^{\infty} l^{1/2} c_l^2 \right)^2, \end{aligned}$$

и ряд  $\sum_{l=1}^{\infty} l^{1/2} c_l^2$  сходится, если сходится  $\sum_{l=1}^{\infty} l^{1/2} |c_l|$ :

$$\sum_{l=1}^{\infty} l^{1/2} c_l^2 = \sum_{l=1}^{\infty} l^{1/2} |c_l|^2 \leq \sup_j |c_j| \sum_{l=1}^{\infty} l^{1/2} |c_l| < \infty,$$

где  $\sup_j |c_j| < \infty$ , так как  $\sum_{l=1}^{\infty} l^{1/2} |c_l| < \infty$ , и значит  $|c_l| \rightarrow 0$  при  $l \rightarrow \infty$ .

Таким образом, имеем

$$\begin{aligned} X_t^2 &= B_0(L) \varepsilon_t^2 + 2 \sum_{h=1}^{\infty} B_h(L) \varepsilon_t \varepsilon_{t-h} \\ &= B_0(1) \varepsilon_t^2 + 2 \sum_{h=1}^{\infty} B_h(1) \varepsilon_t \varepsilon_{t-h} - (1-L) \left( \tilde{B}_0(L) \varepsilon_t^2 + 2 \sum_{h=1}^{\infty} \tilde{B}_h(L) \varepsilon_t \varepsilon_{t-h} \right) \\ &= \left( \sum_{j=0}^{\infty} c_j^2 \right) \varepsilon_t^2 + u_t - (1-L) \tilde{v}_t, \end{aligned}$$

где

$$\begin{aligned} u_t &= \varepsilon_t \left( 2 \sum_{h=1}^{\infty} B_h(1) \varepsilon_{t-h} \right), \\ \tilde{v}_t &= \tilde{B}_0(L) \varepsilon_t^2 + 2 \sum_{h=1}^{\infty} \tilde{B}_h(L) \varepsilon_t \varepsilon_{t-h}. \end{aligned}$$

Получаем

$$n^{-1} \sum_{t=1}^n X_t^2 = \left( \sum_{j=0}^{\infty} c_j^2 \right) n^{-1} \sum_{t=1}^n \varepsilon_t^2 + n^{-1} \sum_{t=1}^n u_t - n^{-1} (\tilde{v}_n - \tilde{v}_0).$$

Теперь покажем, что

$$n^{-1} \sum_{t=1}^n u_t \rightarrow_{a.s.} 0.$$

Пусть  $\mathcal{F}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \dots)$ . Тогда  $\{(u_t, \mathcal{F}_t)\}$  — мартингал-разность.

**Лемма 11 (СЗБЧ для мартингал-разностей)** Пусть  $\{(u_t, \mathcal{F}_t)\}$  — мартингал-разность. Если для некоторого  $r \geq 1$  верно  $\sum_{t=1}^{\infty} \mathbb{E} |u_t|^{2r} / t^{1+r} < \infty$ , то  $n^{-1} \sum_{t=1}^n u_t \rightarrow_{a.s.} 0$ .

См. White (2001, теорема 3.76). Проверим, что  $\{u_t\}$  удовлетворяет условию приведенной леммы. Положим  $r = 1$ . Условие выполняется, если  $\sup_t \mathbb{E} u_t^2 < \infty$ , так как  $\sum_{t=1}^{\infty} t^{-2} < \infty$ .

$$\mathbb{E} u_t^2 = 4\sigma^2 \mathbb{E} \left( \sum_{h=1}^{\infty} B_h(1) \varepsilon_{t-h} \right)^2 = 4\sigma^4 \sum_{h=1}^{\infty} B_h(1)^2.$$

Далее,

$$\begin{aligned} \sum_{h=1}^{\infty} B_h(1)^2 &= \sum_{h=1}^{\infty} \left( \sum_{j=0}^{\infty} c_j c_{j+h} \right)^2 \leq \sum_{h=1}^{\infty} \left( \sum_{j=0}^{\infty} c_j^2 \right) \sum_{j=0}^{\infty} c_{j+h}^2 \\ &= \left( \sum_{j=0}^{\infty} c_j^2 \right) \sum_{h=1}^{\infty} \sum_{j=0}^{\infty} c_{j+h}^2 = \left( \sum_{j=0}^{\infty} c_j^2 \right) \sum_{j=1}^{\infty} j c_j^2. \end{aligned}$$

Применяя логику, аналогичную рассуждениям после доказательства теоремы 2 раздела 3.1, получим, что из  $\sum_{j=1}^{\infty} j^{1/2} |c_j| < \infty$  следует  $\sum_{j=1}^{\infty} j c_j^2 < \infty$ . Как и ранее, можно показать, что

$$n^{-1} (\tilde{v}_n - \tilde{v}_0) \rightarrow_p 0,$$

если только  $\sum_{j=1}^{\infty} j^{1/2} |c_j| < \infty$ . Наконец, в силу СлЗБЧ для последовательностей н.о.р. величин

$$n^{-1} \sum_{t=1}^n \varepsilon_t^2 \rightarrow_p \sigma^2.$$

Следовательно,

$$n^{-1} \sum_{t=1}^n X_t^2 \rightarrow_p \sigma^2 \sum_{j=0}^{\infty} c_j^2 = \mathbb{E}X_t^2.$$

## Список литературы

- Andrews, D.W.K. (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21, 930–934.
- Beveridge, S. & C.R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the ‘business cycle’. *Journal of Monetary Economics* 7, 151–174.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press: New York.
- Gourieroux, C. & A. Monfort (1995). *Statistics and Econometric Models*. Cambridge University Press: Cambridge.
- Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press: Princeton.
- Hansen, L.P. & T.J. Sargent (1991). *Rational Expectations Econometrics*. Westview Press: Boulder.
- Phillips, P.C.B. & V. Solo (1992). Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.
- White, H. (2001). *Asymptotic Theory For Econometricians*. Academic Press: San Diego.

## Linear processes: properties and asymptotic results

Vadim Marmer

*University of British Columbia, Vancouver, Canada*

This essay surveys results for linear time series including Wold decomposition, properties of spectral density functions and lag operators, autoregressive moving average models, Beveridge–Nelson decomposition, and Phillips–Solo device for deriving asymptotics.



# Асимптотика почти единичных корней<sup>\*</sup>

Станислав Анатольев<sup>†</sup>

*Российская экономическая школа, Москва, Россия*

Николай Господинов<sup>‡</sup>

*Университет Конкордия, Монреаль, Канада*

В некоторых общепринятых асимптотических теориях предельное распределение претерпевает разрывы, или же асимптотическое распределение неточно приближает истинное конечномерное распределение. В подобных ситуациях оказывается полезен такой инструментарий, как дрейфующие параметризации, когда определенные параметры явным образом зависят от размера выборки. Дрейфующие параметризации используются среди прочего для анализа импульсных откликов и долгосрочного прогнозирования сильнозависимых процессов. Настоящее эссе представляет собой обзор подобных альтернативных асимптотических приближений в контексте временных рядов.

## 1 Введение

Известно, что когда авторегрессионный параметр в AR(1)-модели в точности равен единице, его МНК-оценка асимптотически ведет себя согласно распределению Дики–Фуллера. Это распределение несимметрично и, в частности, смещено и скошено. В то же время, если авторегрессионный параметр строго меньше единицы, асимптотическое распределение гауссово, то есть симметрично и, в частности, несмещено и нескошено. Истинное же распределение МНК-оценки при типичного размера выборке в случае, когда параметр близок к единице, не соответствует нормальному, а обладает всеми признаками распределения Дики–Фуллера. Это означает, что стандартная асимптотика работает в подобных случаях плохо, и хотелось бы иметь какое-то альтернативное, более адекватное, приближение реальности, некое расширение асимптотической теории Дики–Фуллера на случай корней не единичных, но близких к единичным. В дополнение к практическим соображениям, хотелось бы избавиться от скачка в асимптотическом распределении при переходе из зоны стационарности к ситуации с единичным корнем.

В данном эссе обсуждаются альтернативные асимптотические приближения, которые не рассматривают как фиксированный тот параметр, из-за которого наблюдается разрыв асимптотического распределения, а параметризуют его как дрейфующую последовательность, явным образом зависящую от размера выборки. По мере того, как выборка растет, последовательность сдвигается по направлению к границе множества значений параметра и в пределе стремится к ней. Такая искусственная статистическая конструкция гарантирует гладкий переход в асимптотической теории, что соответствует поведению и истинного конечновыборочного распределения. Подобные конструкции в современной эконометрике являются популярным инструментарием, используемым для моделей с сильно автокоррелированными

---

<sup>\*</sup>Цитировать как: Анатольев, Станислав & Николай Господинов (2012) «Асимптотика почти единичных корней», Квантиль, №10, стр. 57–71. Citation: Anatolyev, Stanislav & Nikolay Gospodinov (2012) “Asymptotics of near unit roots,” *Quantile*, No.10, pp. 57–71.

<sup>†</sup>Адрес: 117418, г. Москва, Нахимовский проспект, 47, офис 1721(3). Электронная почта: [sanatoly@nes.ru](mailto:sanatoly@nes.ru)

<sup>‡</sup>Адрес: Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montréal, Québec, H3G 1M8. Электронная почта: [gospodin@alcor.concordia.ca](mailto:gospodin@alcor.concordia.ca)

переменными. Недостатком такого асимптотического подхода является появление дополнительных состоятельно не оцениваемых шумовых параметров, что осложняет процедуру инференции.

Другой важной ситуацией, где введение дрейфующих последовательностей оказывается полезным, является построение долгосрочных прогнозов и функций импульсных откликов. Стандартная асимптотическая теория подразумевает, что горизонт прогнозирования фиксирован, и предел берется только по отношению к размеру выборки. В то же время, когда горизонт прогнозирования составляет соразмерную с длиной выборки величину, стандартная асимптотическая теория не дает разумных приближений для истинных конечновыборочных распределений. Поэтому выгодно параметризовать горизонт прогнозирования, заставив его расти с определенной скоростью по мере того, как размер выборки растет до бесконечности.

Замечательные обзоры некоторых тем, обсуждаемых в данном эссе, можно также найти в Stock (1994, 1997), Stock, Wright & Yogo (2002) и Stock & Watson (2008).

## 2 Параметризация локальности к единице

Рассмотрим AR( $p$ )-процесс

$$y_t^* = \mu_1 + \mu_2 t + y_t, \quad \Gamma(L)y_t = e_t, \quad (1)$$

где  $\Gamma(L) = 1 - \gamma_1 L - \gamma_2 L^2 - \dots - \gamma_p L^p$  с корнями на или вне единичного круга,  $y_{-p+1}, \dots, y_0$  предполагаются фиксированными, а  $e_t$  является последовательностью мартингальных приращений с  $\mathbb{E}[e_t^2] = \sigma^2$  и  $\sup_t \mathbb{E}[e_t^{2+\xi}] < \infty$  при некотором  $\xi > 0$ . Полином  $\Gamma(L)$  можно разложить как (Stock 1991)

$$\Gamma(L) = \psi(L)(1 - \phi L), \quad (2)$$

где  $\phi$  обозначает наибольший корень AR-полинома, а  $\psi(L) = 1 - \sum_{j=1}^{p-1} \psi_j L^j$  — полином от лага, описывающий краткосрочную динамику процесса без корней на или вблизи единичного круга.

Многие экономические временные ряды, включая процентные ставки, безработицу, реальные обменные курсы и подразумеваемую волатильность, характеризуются очень сильной связностью, но в то же время предположение о единичном корне плохо увязывается с экономической теорией. Скорее, речь идет о случае, когда  $\phi$  близок к единице, но не равен в точности единице. Модель можно переписать в виде улучшенной регрессии Дики–Фуллера (ADF)

$$y_t = \rho y_{t-1} + \varphi_1 \Delta y_{t-1} + \dots + \varphi_{p-1} \Delta y_{t-p+1} + e_t, \quad (3)$$

где  $\rho = \gamma_1 + \gamma_2 + \dots + \gamma_p$  и  $\varphi_j = -(\gamma_{j+1} + \dots + \gamma_p)$  для  $j = 1, \dots, p-1$ . Такая форма часто используется на практике для оценивания, тестирования и прогнозирования. Модель (3) обычно оценивается с помощью МНК, а МНК-оценка состоятельна и асимптотически нормальна при  $\rho < 1$ .

Чтобы явно учесть сильную связность процесса, удобно использовать параметризацию локальности к единице (Chan & Wei 1987, Phillips 1987) наибольшего авторегрессионного корня

$$\phi_n = \exp\left(\frac{c}{n}\right) \approx 1 + \frac{c}{n},$$

где  $c$  — фиксированная константа, а  $n$  — размер выборки. Обычно предполагается, что  $c \leq 0$ , причем случай  $c = 0$  соответствует точному единичному корню, но допускаются и слабо взрывные процессы ( $c > 0$ ). В результате данная параметризация включает в себя целый

набор значений  $\phi$ , что особо полезно, когда неясно, каково точное значение наибольшего авторегрессионного корня. Из принятой параметризации следует, что в ADF-представлении процесса (3),  $\rho_n = 1 + c\psi(1)/n$ , где  $\psi(1) = 1 - \psi_1 - \dots - \psi_{p-1}$ .

В рамках локальности к единице пространство параметров моделируется как окрестность единицы, сжимающаяся по мере того, как размер выборки растет. Такая (воображаемая!) статистическая конструкция устраняет разрыв в асимптотическом распределении и предоставляет отличное приближение для конечновыборочного распределения  $\hat{\rho}$  в модели (3). Также она облегчает анализ непрерывных асимптотических пределов, т.к.  $\lim_{n \rightarrow \infty} (\phi_n)^n = \exp(c)$ . Наконец, стоит упомянуть еще одну важную черту принятой параметризации: в отличие от тестирования на единичные корни, где исследователю приходится в конечном счете выбирать между одним из двух типов приближений, асимптотика локальности к единице учитывает неопределенность в значении наибольшего авторегрессионного корня и дает возможность приближения во всем пространстве параметров.

Предельное распределение оценки  $\hat{\rho}_n$  можно получить с помощью функциональной ЦПТ, а именно (опуская для простоты записи зависимость  $\hat{\rho}_n$  от  $n$ ), имеем (Phillips 1987)

$$n(\hat{\rho} - \rho_0) \Rightarrow \psi(1) \frac{\int_0^1 J_c^\tau(s) dW(s)}{\int_0^1 J_c^\tau(s)^2 ds},$$

где  $\Rightarrow$  обозначает слабую сходимость,  $J_c(r)$  — процесс Орнштейна–Уленбека, порождаемый стохастическим дифференциальным уравнением  $dJ_c(r) = cJ_c(r)dr + dW(r)$ ,  $\{W(r) : r \in [0, 1]\}$  — стандартное броуновское движение, и  $J_c^\tau(r) = J_c(r)$  если (1) не содержит детерминистических компонент ( $\mu_1 = 0, \mu_2 = 0$ ),  $J_c^\tau(r) = J_c(r) - \int_0^1 J_c(s)ds$  если (1) включает только константу ( $\mu_1 \neq 0, \mu_2 = 0$ ) и  $J_c^\tau(r) = J_c(r) - \int_0^1 (4 - 6s)J_c(s)ds - r \int_0^1 (12s - 6)J_c(s)ds$  если (1) включает и константу, и линейный тренд ( $\mu_1 \neq 0, \mu_2 \neq 0$ ). Кроме того, нормализованная оценка ( $t$ -статистика) в пределе есть

$$t_\rho = \frac{\hat{\rho} - \rho_0}{\text{se}(\hat{\rho})} \Rightarrow \frac{\int_0^1 J_c^\tau(s) dW(s)}{\sqrt{\int_0^1 J_c^\tau(s)^2 ds}}. \tag{4}$$

Предельные распределения для  $\hat{\rho}$  и  $t_\rho$  — функции параметра локальности к единице, который состоятельно не оцениваем, так как  $\hat{c} - c_0 = O_P(1)$ . Данный факт осложняет инференцию в подобных моделях. В то же время параметризация локальности к единице предоставляет равномерное асимптотическое приближение для  $t_\rho$  для всех  $\rho$  в интервале  $(0, 1]$  (Mikusheva 2007). В частности, если параметр локальности  $c$  уходит на  $-\infty$ , предельное распределение приближается к стандартному нормальному (Phillips 1987). Если  $c = 0$ , предельное распределение схлопывается в распределение Дики–Фуллера. Все промежуточные значения  $c$  непрерывным образом связывают стандартную асимптотику и асимптотику Дики–Фуллера.

Несмотря на то, что параметр локальности  $c$  состоятельно не оценить, существуют методы построения доверительных интервалов для  $c$  и  $\rho$ . Они основаны на обращении асимптотических (Stock 1991), бутстраповских (Hansen 1999) или Монте–Карловских (Andrews 1993) тестов на решетке значений  $c$  (или  $\rho$ ). Mikusheva (2007) показывает, что упомянутые три метода дают асимптотически правильные доверительные интервалы для  $\rho$ .

Покажем, как построить доверительные интервалы для  $\rho$ . Следуя работе Stock (1991), обозначим за  $q_c(\alpha)$   $\alpha$ -квантиль асимптотического распределения в (4). Тогда  $100(1 - \alpha)\%$ -ным доверительным интервалом для  $\rho$  будет  $CI_\rho = \{\rho : q_c(\alpha/2) \leq t_\rho \leq q_c(1 - \alpha/2)\}$ . Асимптотические квантильные функции  $q_c(\alpha/2)$  и  $q_c(1 - \alpha/2)$  обычно получают с помощью симуляций. Для конкретной длины выборки  $n$  эти квантильные функции завязаны на неявное значение  $\rho$ . Статистику  $t_\rho$  тогда можно вычислить для последовательности нулевых гипотез на решетке для  $\rho$ ; пересечения тестовой статистики и квантильных функций дадут концы  $100(1 - \alpha)\%$ -го доверительного интервала. Методы в работах Andrews (1993) и Hansen

(1999) основаны на той же идее, но асимптотические квантильные функции заменяются на их Монте–Карловские или бутстраповские аналоги.

Инструментарий локальности к единице также оказывается полезен при анализе последствий неверной спецификации процессов. Например, перепишем процесс (1)–(2) в виде

$$(1 - \phi L)y_t = a(L)e_t,$$

где  $a(L) = \psi^{-1}(L) = \sum_{i=0}^{\infty} a_i L^i$  и  $a_0 = 1$ ,  $\sum_{i=0}^{\infty} i|a_i| < \infty$ ,  $\sum_{i=0}^{\infty} a_i \neq 0$ , и предполагается отсутствие детерминистической компоненты.

Пусть  $u_t = a(L)e_t$ , тогда  $u_t = (a(1) + (1 - L)a^*(L))e_t$  согласно разложению Бевеиджа–Нельсона, где  $a(1) = \sum_{i=0}^{\infty} a_i$ ,  $a^*(L) = \sum_{i=1}^{\infty} a_i^* L^{i-1}$  и  $a_i^* = -\sum_{j=i}^{\infty} a_j$ . Обозначив  $S_t = \sum_{j=1}^t e_j$  и используя рекурсивную подстановку и суммирование по частям, получим:

$$\begin{aligned} y_t &= \sum_{j=1}^t \phi^{t-j} a(L)e_j \\ &= (a(1) + (1 - L)a^*(L)) S_t + \sum_{j=1}^{t-1} (a(1) + (1 - L)a^*(L)) S_{j-1} (\phi^{t-j+1} - \phi^{t-j}) \\ &= a(1)S_t + v_t + \sum_{j=1}^{t-1} a(1)S_{j-1} (\phi^{t-j+1} - \phi^{t-j}) + \sum_{j=1}^{t-1} v_{j-1} (\phi^{t-j+1} - \phi^{t-j}) \\ &= a(1)S_t + v_t + \frac{c}{n} \sum_{j=1}^{t-1} a(1)\phi^{t-j} S_{j-1} + \frac{c}{n} \sum_{j=1}^{t-1} \phi^{t-j} v_{j-1}, \end{aligned} \quad (5)$$

где  $v_t = a^*(L)e_t$ . Выражение (5) — алгебраическое разложение процесса  $y_t$ . Оно обобщает стандартное разложение Бевеиджа–Нельсона для процессов с единичным корнем как частный случай ( $c = 0$ ). Интересно, что в то время как стандартное разложение Бевеиджа–Нельсона задается компонентами  $a(1)S_t$  (перманентной) и  $v_t$  (преходящей), разложение (5) содержит два дополнительных слагаемых. Четвертая часть

$$\frac{c}{n} \sum_{j=1}^{t-1} \phi^{t-j} v_{j-1}$$

асимптотически пренебрежимо мала, но

$$\frac{c}{n} \sum_{j=1}^{t-1} a(1)\phi^{t-j} S_{j-1}$$

нет, она порядка  $Op(n^{1/2})$ , того же, что и  $a(1)S_t$ . Поэтому (5) содержит два слагаемых перманентной компоненты (т.е., слагаемые порядка  $Op(n^{1/2})$ ), так что упущение одной из них (например, неверное предположение точного единичного корня) приведет к смещенным оценкам и искаженным тестам, а также к смещенным выводам касательно стандартного разложения на тренд и цикл, основанного на предположении о точном единичном корне.

Будем обозначать как  $[\cdot]$  взятие целой части и перейдем к непрерывному времени:  $t = [rn]$  для  $r \in [0, 1]$ . Тогда, деля обе части (5) на  $\sqrt{n}$  и беря предел при  $n \rightarrow \infty$ , получаем:

$$\frac{1}{\sqrt{n}} y_{[rn]} \Rightarrow \sigma^2 a(1) \left( W(r) + c \int_0^r \exp((r-s)c) W(s) ds \right) \equiv J_c(r), \quad (6)$$

где  $J_c(r)$  — процесс Орнштейна–Уленбека, а  $W(r)$  — стандартное броуновское движение. Предельный результат (6) подтверждает, что обе части (5) сходятся к одному и тому же процессу Орнштейна–Уленбека.

Полезно провести аналогичный анализ в многомерных моделях. Пусть теперь  $y_t$  — многомерный процесс

$$(I_m - \Phi L)y_t = A(L)e_t,$$

где  $\Phi = I_m + C/n$ , а  $C = \text{diag}\{c_1, c_2, \dots, c_m\}$ , и  $c_1, c_2, \dots, c_m$  — фиксированные постоянные. Используя ту же технологию, что и раньше, имеем

$$y_t = A(1)S_t + v_t + \frac{C}{n}A(1) \sum_{j=1}^{t-1} \Phi^{t-j} S_{j-1} + \frac{C}{n} \sum_{j=1}^{t-1} \Phi^{t-j} v_{j-1},$$

и наложение ограничений на  $A(1)$  недостаточно, чтобы удалить перманентную компоненту, поскольку имеется другое  $O_P(n^{1/2})$ -слагаемое  $Cn^{-1} \sum_{j=1}^{t-1} \Phi^{t-j} S_{j-1}$ .

Обобщая далее, можно рассмотреть коинтеграционную модель

$$\begin{aligned} x_{1,t} &= \beta x_{2,t} + e_{1,t}, \\ x_{2,t} &= (1 + c/n)x_{2,t-1} + e_{2,t} \end{aligned}$$

рассмотренную в Elliott (1998), и получить похожее представление. В этом случае ограничение  $\beta A(1) = 0$  не уничтожает перманентную компоненту. Как следствие, регрессии редуцированного ранга, которые налагают только это ограничение, неприемлемы и приводят к смещенным оценкам и искаженным тестам, как показал Elliott (1998). Более того, ограничения, наложенные на «неверную» перманентную компоненту, скорей всего негативно повлияют на импульсные отклики и разложения дисперсии, основанные на долгосрочных идентифицирующих ограничениях на  $A(1)$ .

### 3 Инференция и долгосрочное прогнозирование локальных к единице процессов

Отслеживание влияния структурных шоков на динамику эндогенных переменных является одним из главных инструментов анализа экономической политики. Часто интересно тестирование гипотез о величине и форме импульсных откликов при больших горизонтах, например  $\theta_l \equiv \partial y_{t+l} / \partial e_t$  для горизонта  $l$ . Или же интересно измерить, сколько времени требуется невзрывному процессу, чтобы затухнуть на  $100\omega\%$  от первоначального шока, т.е.  $\sup_{l \in L} |\partial y_{t+l} / \partial e_t| \geq 1 - \omega$  для некоторого фиксированного  $\omega \in (0, 1]$ . В обоих случаях ограничения, накладываемые нулевой гипотезой, являются полиномами порядка  $l$  от параметров модели. В модели AR(1) структура затухания монотонна, и  $l = \log(1 - \omega) / \log(\rho)$  или  $l = \lfloor \delta n \rfloor$ , где  $\delta = \log(1 - \omega) / c$  — фиксированная положительная постоянная, если  $c < 0$ . Для AR-моделей более высокого порядка с корнем вблизи единицы Rossi (2005) показала, что  $l = \lfloor \delta n \rfloor + o(1)$ . Поэтому порядок полинома в ограничении увеличивается линейно с размером выборки по мере того, как процесс приближается к таковому с единичным корнем.

По этой причине удобно принять параметризацию  $l = \lfloor \delta n \rfloor$  для некоторого фиксированного  $\delta > 0$ , что ранее использовалось в литературе для анализа импульсных откликов и долгосрочного прогнозирования почти нестационарных процессов (Stock 1996, Phillips 1998, Gospodinov 2002a, Rossi 2005). Данная конструкция позволяет сохранить в асимптотическом приближении неопределенность в оценивании параметров, свойственную конечномерному распределению. В случае же фиксированного горизонта планирования неопределенность в оценивании параметров асимптотически исчезает по мере того как  $n$  уходит в  $\infty$ .

Рассмотрим оценивание параметров модели (3) с  $\rho_n = 1 + c/n$  и наложим ограничение, что  $\theta_l$  равно некоторому конкретному значению  $\theta_{0,l}$ , где  $l = \lfloor \delta n \rfloor$ . Gospodinov (2004) показывает, что оценка  $\rho$  под ограничением сходится с большей скоростью ( $n^{3/2}$ ), чем оценка

без ограничений ( $n$ ). В результате параметр локальности  $c$  можно оценить под ограничением. Данный результат — следствие параметризации горизонта импульсных откликов как функции размера выборки. Из-за этой параметризации параметр почти нестационарной компоненты решает сильно нелинейное полиномиальное ограничение, степень которого растет с размером выборки, что в свою очередь ускоряет сходимость оценки.

С целью иллюстрации этого результата рассмотрим модель AR(2)

$$y_t = \rho y_{t-1} + \varphi \Delta y_{t-1} + e_t.$$

Эту модель можно переписать в виде

$$\begin{pmatrix} y_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} \rho & \varphi \\ \rho - 1 & \varphi \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \begin{pmatrix} e_t \\ e_t \end{pmatrix}$$

или

$$Y_t = \Lambda Y_{t-1} + \bar{e}_t.$$

Тогда  $Y_{t+l} = \Lambda^{l+1} Y_{t-1} + \Lambda^l \bar{e}_t + \Lambda^{l-1} \bar{e}_{t+1} + \dots + \bar{e}_{t+l}$ , а импульсный отклик при горизонте  $l$  равен

$$\theta_l \equiv \frac{\partial y_{t+l}}{\partial e_t} = (1, 0) \Lambda^l \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Поскольку  $\rho - 1 = c/n = o_P(1)$ ,

$$\begin{aligned} & (1, 0) \begin{pmatrix} \rho & \varphi \\ o_P(1) & \varphi \end{pmatrix}^l \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= (1, 0) \begin{pmatrix} \rho^l + o_P(1) & \rho^{l-1} \varphi + \rho^{l-2} \varphi^2 + \dots + \varphi^l + o_P(1) \\ o_P(1) & \varphi^l + o_P(1) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \rho^l + \rho^{l-1} \varphi + \rho^{l-2} \varphi^2 + \dots + \varphi^l + o_P(1) \\ &= \rho^l \left( 1 + \frac{\varphi}{\rho} + \dots + \left( \frac{\varphi}{\rho} \right)^l \right) + o_P(1). \end{aligned}$$

Далее, если  $l = \lfloor \delta n \rfloor$ , по мере того как  $n \rightarrow \infty$ , имеем:  $\rho^l = \exp(c\delta)$ ,  $1 + \varphi/\rho + \dots + (\varphi/\rho)^l = \rho/(\rho - \varphi) = (1 - \varphi)^{-1} + o_P(1)$  и  $\theta_l \rightarrow (1 - \varphi)^{-1} \exp(c\delta)$  (Rossi 2004).

Обозначим как  $\tilde{c}$  и  $\tilde{\varphi}$  ограниченные МНК-оценки параметров  $c$  и  $\varphi$  при ограничении  $\theta_l = \theta_{0,l}$ . Из выражений выше следует, что ограниченная оценка  $\tilde{c}$  является нелинейной функцией  $\tilde{\varphi}$ , а именно

$$\tilde{c} = \frac{\log(\theta_{0,l}(1 - \tilde{\varphi}))}{\delta} = g(\tilde{\varphi}).$$

Разложение по Тейлору первого порядка  $\tilde{c}$  вокруг  $c_0$  дает

$$\sqrt{n}(\tilde{c} - c_0) = \frac{\partial g(\varphi)}{\partial \varphi} \sqrt{n}(\tilde{\varphi} - \varphi_0) + o_P(1),$$

т.е. оценка  $\tilde{c}$  вынуждена сходиться с той же скоростью, что и оценка  $\tilde{\varphi}$ . Gospodinov (2004) развивает далее процедуру построения асимптотических поточечных доверительных коридоров для импульсных откликов на основе обращения  $\mathcal{LR}$ -теста.

Похожая технология используется в контексте долгосрочного условного прогнозирования локальных к единице процессов. Проиллюстрируем основную идею на примере модели AR(1) с нулевым средним

$$y_t = \rho y_{t-1} + e_t, \quad t = 1, 2, \dots, n,$$

Положим, нам интересны прогнозы будущих значений  $y_{n+l}$ , условные на последнем наблюдении  $y_n$ , когда  $\rho$  находится вблизи единицы параметризовано как  $\rho_n = 1 + c/n$ , а  $l$  сопоставимо с размером выборки и параметризовано как  $l = \lfloor \lambda n \rfloor$ . В авторегрессионных моделях зависимость между данными, используемыми для оценивания, и данными, используемыми для прогнозирования, порождает некоторые трудности в оценивании эффекта неопределенности в оценке параметров и их свойств на прогнозы (Phillips 1979). В отличие от стационарного случая, зависимость между  $\hat{\rho}$  и  $y_n$  асимптотически не исчезает, если корни AR-процесса на или возле единичного круга. Это видно из совместной асимптотики МНК-оценки  $\rho$  и последнего наблюдения (Phillips 1987):

$$\left( n(\hat{\rho} - \rho_0), n^{-1/2} y_n \right) \Rightarrow \left( \frac{1}{2} \frac{J_c(1)^2 - J_c(0)^2 - 2c \int_0^1 J_c(s)^2 ds - 1}{\int_0^1 J_c(s)^2 ds}, J_c(1) \right).$$

Истинным значением будущего значения  $y$  через  $l$ -шагов будет

$$y_{n+l} = \rho_0^l y_n + \sum_{j=0}^{l-1} \rho_0^j e_{n+l-j},$$

истинным условным ожиданием —  $\bar{y}_{n+l|n} = \mathbb{E}[y_{n+l}|y_n] = \rho_0^l y_n$ , а доступным условным ожиданием —  $\hat{y}_{n+l|n} = \hat{\rho}^l y_n$ , где  $\hat{\rho}$  — МНК-оценка. Целью является приближение распределения для  $\hat{y}_{n+k|n} - \bar{y}_{n+k|n}$  и ошибок прогноза  $\hat{y}_{n+k|n} - y_{n+k}$ , условно на значении  $y_n$ . Поскольку  $y_n$  имеет порядок  $O_P(n^{1/2})$  и расходится при  $n \rightarrow \infty$ , более разумно уславливать на масштабированном конечном значении  $n^{-1/2} y_n = x$ .

Условное асимптотическое представление нормализованной разности  $\hat{y}_{n+k|n} - \bar{y}_{n+k|n}$  следующее (Gospodinov 2002a):

$$\begin{aligned} & \left( n^{-1/2} (\hat{y}_{n+k|n} - \bar{y}_{n+k|n}) | x \right) \\ & \Rightarrow \exp(\lambda c) \left( \exp \left( \frac{\lambda x^2 - J_c(0)^2 - 2c \int_0^1 J_{c|x}(s)^2 ds - 1}{\int_0^1 J_{c|x}(s)^2 ds} \right) - 1 \right) x, \end{aligned}$$

где  $J_{c|x}(r)$  — траектория, согласующаяся с масштабированным конечным значением  $x$ . Условное асимптотическое представление нормализованных ошибок прогнозов

$$n^{-1/2} (\hat{y}_{n+k|n} - y_{n+k})$$

включает дополнительную компоненту  $N(0, (\exp(2c\lambda) - 1)/2c)$  (Gospodinov 2002a).

Предельные распределения зависят от состоятельно не оцениваемого  $c$ . Gospodinov (2002a) предлагает процедуру условного решетчатого бутстрапа и доказывает ее асимптотическую обоснованность при параметризациях локальности к единице и длинного горизонта.

## 4 Почти необратимое бегущее среднее и локальное к нулю отношение сигнала к шуму

Параметризации локальности к единице и к нулю также играют важную роль в моделировании почти необратимых процессов и процессов с почти постоянными локальными уровнями.

Чтобы понять основную идею, рассмотрим стохастический процесс  $\{y_t\}_{t=1}^n$ , порожденный моделью бегущего среднего первого порядка MA(1):

$$y_t = e_t - \theta e_{t-1},$$

где  $|\theta| \leq 1$  и  $e_t \sim iid(0, \sigma^2)$  с  $\mathbb{E}[e_t^4] < \infty$ .

Пусть нас интересует параметр  $\theta$ , а за  $\eta$  обозначим (возможно бесконечномерный) вектор шумовых параметров, полностью описывающих распределение  $e$ . Если  $|\theta|$  строго меньше единицы, процесс  $y_t$  обратим, и оценка максимального правдоподобия (ММП) параметра  $\theta$  асимптотически нормально распределена со средним  $\theta$  и дисперсией  $(1 - \theta^2)/n$ . Когда  $\theta$  находится вблизи единичного круга, гауссовское распределение является не очень качественным приближением истинного распределения ММП-оценки. Более того, когда истинный MA-параметр находится в окрестности единицы, в конечных выборках оценка принимает значения в точности на границе множества обратимости с положительной вероятностью.

Распределение ММП-оценки параметра  $\theta$  в присутствии единичного MA-корня нестандартно (см. Davis & Dunsmuir 1996). Параметризовав MA-параметр как локальный к единице, т.е. в форме  $\theta_n = 1 + c/n$ , где  $c \leq 0$  — фиксированная постоянная, мы получим конструкцию, полезную для анализа предельного поведения ММП-оценки и предоставляющую плавный переход от приближения нормальным распределением к приближению в случае необратимого MA-процесса.

Гибкая перепараметризация модели MA(1) задается процессом с постоянными локальными уровнями:

$$\begin{aligned} x_t &= \alpha_t + u_t, \\ \alpha_t &= \alpha_{t-1} + \tau \xi_t, \end{aligned}$$

где  $\alpha_t$  ненаблюдаемый меняющийся во времени параметр,  $u_t$  и  $\xi_t$  взаимно некоррелированные белые шумы, а  $\tau$  представляет из себя отношение сигнала к шуму. Взяв приращения и определив  $\Delta x_t = y_t$ , получим:

$$y_t = \tau \xi_t + \Delta u_t.$$

Легко показать, что здесь автокорреляционная структура та же, что и в ограниченной модели MA(1)

$$\Delta x_t = e_t - \theta e_{t-1}$$

с ограничением  $0 \leq \theta \leq 1$ . На самом деле существует взаимно однозначное отображение между параметрами двух представлений  $\tau$  и  $\theta$ , а именно,  $\tau = \sqrt{(1 - \theta)^2/\theta}$  и  $\theta = (\tau^2 + 2 - \sqrt{\tau^4 + 4\tau^2})/2$ , монотонные по  $\theta$  и  $\tau$ , соответственно. Отсюда следует, что тестирование на единичный MA-корень  $\theta = 1$  эквивалентно тестированию на постоянство  $\alpha$ , т.е. нулевой гипотезы  $H_0 : \tau = 0$  против альтернативы  $H_1 : \tau > 0$ . Параметризация локальности к единице MA-параметра  $\theta_n = 1 + c/n$  в ограниченной модели MA(1) соответствует параметризации локальности к нулю отношения сигнала к шуму  $\tau_n = \lambda/n$  в модели с постоянными локальными уровнями. Gospodinov (2002b) предлагает бутстраповскую процедуру построения доверительного интервала и несмещенного (в смысле медианы) оценивания  $\theta$  и  $\tau$  и доказывает ее асимптотическую обоснованность при параметризациях локальности к единице и к нулю.



## 5 Почти коинтегрированные временные ряды с локальной к нулю дисперсией коинтеграционной ошибки

Введенные выше дрейфующие параметризации оказываются полезными и в анализе многомерных процессов. Рассмотрим треугольную систему

$$\begin{aligned} y_t &= \beta x_t + u_{y,t}, \\ x_t &= u_{x,t} \end{aligned} \tag{7}$$

и

$$\begin{pmatrix} (1 - \rho L)u_{y,t} \\ (1 - L)u_{x,t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix},$$

где  $(\varepsilon_{y,t}, \varepsilon_{x,t})'$  — вектор iid-случайных величин с нулевым средним; для простоты мы предполагаем отсутствие динамики более высоких порядков и детерминистических компонент. Заметим, что коинтеграция соответствует случаю  $\rho = 1$ , а отсутствие таковой — случаю  $-1 < \rho < 1$ . Предположим теперь, что  $\mathbb{V}(\varepsilon_{y,t}) \ll \mathbb{V}(\varepsilon_{x,t}) = 1$ . Такую конструкцию можно мотивировать в контексте регрессий форвардных премий, которые описывают динамику спотовых и одномесячных форвардных обменных курсов и их разницы, которая и называется форвардной премией. Существует объемная литература по тестированию на коинтеграцию между спотовыми и форвардными обменными курсами. Из поведения их совместной динамики во времени видно, что два ряда очень тесно следуют друг за другом и не показывают намерения сильно отклоняться друг от друга. В то же время, сильная устойчивость, т.е. наличие почти единичного корня в поведении ошибок коинтеграции, приводит к тому, что статистические тесты часто отвергают гипотезу коинтеграции, что входит в противоречие с экономической интуицией. Анализ совместного процесса, однако, показывает, что компонента, связанная с единичным корнем, настолько маленькая, что не дает переменным отклоняться друг от друга в течение достаточно долгих периодов. Каким образом можно смоделировать и проанализировать такую систему? Чтобы отразить наблюдаемые свойства данных, вспомним и скомбинируем уже встречавшиеся дрейфующие параметризации.

Чтобы смоделировать сильную связность и малую дисперсию ошибок коинтеграции, т.е. форвардной премии, параметризуем  $u_{y,t}$  как угасший процесс с почти единичным корнем. Это подразумевает двойную локализацию:

$$\rho_n = 1 + \frac{c}{n}$$

для некоторой фиксированной постоянной  $c \leq 0$  и

$$\varepsilon_{y,t} = \tau v_t, \quad \tau = \frac{\lambda}{\sqrt{n}}$$

для некоторых фиксированной постоянной  $\lambda > 0$  и процесса  $v_t \sim iid(0, 1)$ . Обозначим за  $\{(V_1(r), V_2(r))' : r \in [0, 1]\}$  двумерное броуновское движение с коэффициентом корреляции  $\delta$ . Заметим, что хотя процесс  $u_{y,t}$  почти интегрирован, он ограничен:

$$u_{y,t} = \frac{\lambda}{\sqrt{n}} \sum_{i=1}^t \left(1 + \frac{c}{n}\right)^{t-i} v_i \Rightarrow \lambda J_c(r),$$

где  $J_c(r)$  — процесс Орнштейна–Уленбека, порожденный стохастическим дифференциальным уравнением  $dJ_c(r) = cJ_c(r)dr + dV_2(r)$ .

Теперь исследуем последствия такой параметризации для свойств оценок модели коинтеграции. Используя подход контролирующей переменной, получим, что эффективную оценку  $\beta$  можно получить из регрессии

$$y_t = \beta x_t + \omega \Delta x_t + e_t,$$

где  $\omega$  — регрессионный коэффициент в уравнении  $u_{y,t}$  на  $\varepsilon_{x,t}$ , а  $e_t$  — ошибки этой регрессии. МНК-оценка  $\beta$  в этой регрессии эквивалентна ММП-оценке (Phillips 1991). Оценка  $\hat{\beta}$  асимптотически распределена как (Gospodinov 2009)

$$\sqrt{n} (\hat{\beta} - \beta_0) \Rightarrow \lambda \frac{\int_0^1 J_c(s) V_1(s) ds}{\int_0^1 V_1(s)^2 ds}.$$

Эта оценка состоятельна, но ее скорость сходимости ниже, чем  $n$ . Кроме того, общепринятая  $t$ -статистика для гипотезы  $H_0 : \beta = \beta_0$  расходится со скоростью  $n^{1/2}$ , как в ложных регрессиях.

Векторное представление коррекции ошибки (VEC) модели (7) следующее:

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = (\rho - 1) \begin{pmatrix} 1 & -\beta \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix},$$

а одиночное уравнение условной VEC-модели выглядит как

$$\Delta y_t = \gamma u_{y,t-1} + \bar{\omega} \Delta x_t + \varepsilon_{y,t}, \quad (8)$$

где  $\gamma = \rho - 1$ , а нулевая гипотеза коинтеграции — это  $H_0 : \gamma = 0$ . Обозначим через  $\tilde{\gamma}$  МНК-оценку  $\gamma$  в (8). Взятие пределов выявляет, что асимптотическое поведение оценки  $\tilde{\gamma}$  в условной VEC-модели такое же, как в Hansen (1995) и Zivot (2000):

$$n(\tilde{\gamma} - \gamma_0) \Rightarrow \delta \frac{\int_0^1 J_c(s) dW_1(s)}{\int_0^1 J_c(s)^2 ds} + (1 - \delta)^{1/2} \frac{\int_0^1 J_c(s) dV_2(s)}{\int_0^1 J_c(s)^2 ds}$$

и

$$t_{\tilde{\gamma}} \Rightarrow \delta \bar{z} + \sqrt{1 - \delta^2} \frac{\int_0^1 J_c(s) dV_2(s)}{\left(\int_0^1 J_c(s)^2 ds\right)^{1/2}}$$

где  $\bar{z}$  — стандартная нормальная случайная величина, а  $W_1$  — стандартное броуновское движение, независимое от  $V_2$ .

Часто VEC-модель определяется как (например, для прогнозирования)

$$\Delta y_t = \gamma u_{y,t-1} + \xi_t, \quad (9)$$

где  $\gamma = \rho - 1$  и  $\xi_t = \bar{\omega} \Delta x_t + \varepsilon_{y,t}$ . Обозначим  $\hat{\gamma}$  МНК-оценку  $\gamma$  в (9). В этом случае

$$\sqrt{n} (\hat{\gamma} - \gamma_0) \Rightarrow \frac{1}{\lambda} \left( \delta \frac{\int_0^1 J_c(s) dV_2(s)}{\int_0^1 J_c(s)^2 ds} + (1 - \delta)^{1/2} \frac{\int_0^1 J_c(s) dW_1(s)}{\int_0^1 J_c(s)^2 ds} \right)$$

и

$$t_{\hat{\gamma}} \Rightarrow \delta \frac{\int_0^1 J_c(s) dV_2(s)}{\left(\int_0^1 J_c(s)^2 ds\right)^{1/2}} + \sqrt{1 - \delta^2} \bar{z}.$$

Разница между моделью (9) и условной моделью (8) в более медленной сходимости оценки. Кроме того, в отличие от условной VEC-модели, параметр локальности  $\lambda$  из параметризации отношения сигнала к шуму входит в предельное распределение оценки. Поскольку  $\lambda$  фигурирует в знаменателе, его близкие к нулю значения приводят к сильной изменчивости оценки. Наконец, когда  $\delta \rightarrow 0$ , распределение  $t$ -статистики в безусловной VEC-модели асимптотически приближается к стандартному нормальному, в то время как в условной VEC-модели оно приближается к распределению типа Дики–Фуллера, и наоборот, когда  $\delta \rightarrow 1$ . Симуляции свидетельствуют, что смещение оценки  $\beta$  в условной VEC-модели может быть большим и отрицательным, когда корреляция между ошибками  $\delta$  близка к единице, в то время как оценка в условной VEC-модели остается несмещенной. Более того, дисперсия оценки в безусловной модели может многократно превышать дисперсию оценки в условной (Gospodinov 2009).

## 6 Локальность к единице и почти идентификация

Существует тесная связь между локальностью к единице и слабой идентификацией в структурных моделях временных рядов. Рассмотрим двумерный векторный авторегрессионный процесс для  $\tilde{y}_t = (y_{1,t}, y_{2,t})'$  порядка  $p + 1$

$$\Psi(L)(I_2 - \Phi L)\tilde{y}_t = u_t,$$

где матрица  $\Phi$  содержит наибольшие корни системы и

$$\Psi(L) = I_2 - \sum_{i=1}^p \Psi_i L^i = \begin{bmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{bmatrix}$$

является полиномом  $p$ -го порядка с корнями вне единичного круга. Предполагается, что ошибки  $u_t$  являются двумерной мартингал-разностью, причем  $\mathbb{E}[u_t u_t' | u_{t-1}, u_{t-2}, \dots] = \Sigma > 0$  и  $\sup_t \mathbb{E}[\|u_t\|^{2+\xi}] < \infty$  для некоторого  $\xi > 0$ , а стартовые значения предполагаются фиксированными.

Матрица наибольших корней  $\Phi$  параметризуется как

$$\Phi_n = \begin{bmatrix} 1 & 0 \\ 0 & 1 + c/n \end{bmatrix} = I_2 + \frac{C}{n},$$

где

$$C = \begin{bmatrix} 0 & 0 \\ 0 & c \end{bmatrix},$$

и  $c \leq 0$  — фиксированная постоянная. Внедиагональные элементы  $\Phi_n$  приравниваются к нулю, чтобы исключить случай, когда какой-либо из процессов (почти)  $I(2)$  (Elliott 1998, Phillips 1988). Структура этой матрицы также означает, что процессы некоинтегрированы.

Выгодно наложить наличие точного единичного корня на первую переменную, так что тогда  $\Delta y_{1,t}$  будет стационарным процессом. Пусть  $y_t = (\Delta y_{1,t}, y_{2,t})'$  и

$$D(L) = \Psi(L) \begin{bmatrix} 1 & 0 \\ 0 & 1 - (1 + c/n)L \end{bmatrix}.$$

Тогда приведенная форма VAR-модели выглядит как

$$D(L)y_t = u_t, \tag{10}$$

или

$$y_t = D_1 y_{t-1} + \dots + D_{p+1} y_{t-p-1} + u_t.$$

Умножение слева обеих частей (10) на матрицу

$$B_0 = \begin{bmatrix} 1 & -b_{12}^{(0)} \\ -b_{21}^{(0)} & 1 \end{bmatrix}$$

дает

$$B(L)y_t = \varepsilon_t,$$

где  $B(L) = B_0 D(L)$ , а  $\varepsilon_t = B_0 u_t$  обозначают структурные шоки  $(\varepsilon_{1,t}, \varepsilon_{2,t})'$ , которые согласно обычной практике предполагаются ортогональными с дисперсиями  $\sigma_1^2$  и  $\sigma_2^2$ . Долгосрочное идентифицирующее ограничение об отсутствии долгосрочного влияния шоков  $\varepsilon_2$  на  $y_1$  накладывает нижнетреугольную структуру на матрицу бегущего среднего  $D(1)^{-1} B_0^{-1}$ . При этом идентифицирующем ограничении матрица долгосрочных мультипликаторов в структурной модели,  $B(1) = B_0 D(1)$ , также нижнетреугольна.

В демонстрационных целях упростим модель, сделав предположение  $\Psi(L) = I_2$ . Приведенную форму этой модели первого порядка можно переписать как

$$\begin{aligned} \Delta y_{1,t} &= u_{1,t}, \\ y_{2,t} &= \left(1 + \frac{c}{n}\right) y_{2,t-1} + u_{2,t}. \end{aligned} \quad (11)$$

Структурная форма модели выглядит так:

$$\begin{aligned} \Delta y_{1,t} &= b_{12}^{(0)} y_{2,t} + b_{12}^{(1)} y_{2,t-1} + \varepsilon_{1,t}, \\ y_{2,t} &= b_{21}^{(0)} \Delta y_{1,t} + b_{22}^{(1)} y_{2,t-1} + \varepsilon_{2,t}, \end{aligned} \quad (12)$$

где  $b_{12}^{(1)} = -b_{12}^{(0)}(1 + c/n)$  и  $b_{22}^{(1)} = 1 + c/n$ .

Чтобы идентифицировать структурные параметры модели (12) из приведенной формы (11), мы налагаем долгосрочное ограничение о нижнетреугольной структуре  $B(1)$ . Точнее, добавляя и вычитая  $b_{12}^{(0)} y_{2,t-1}$  из первого уравнения (12) и полагая долгосрочный мультипликатор  $b_{12}^{(0)} + b_{12}^{(1)}$  нулю, мы преобразуем первое уравнение структурной модели к

$$\Delta y_{1,t} = b_{12}^{(0)} \Delta y_{2,t} + \varepsilon_{1,t}. \quad (13)$$

Поскольку  $\Delta y_{2,t}$  эндогенно, неизвестный параметр  $b_{12}^{(0)}$  можно оценить инструментальными переменными, используя  $y_{2,t-1}$  в качестве инструмента. Соотношение между эндогенной переменной и инструментом задается вторым уравнением приведенной формы VAR (11)

$$\Delta y_{2,t} = \frac{c}{n} y_{2,t-1} + u_{2,t}. \quad (14)$$

Заметим что параметризация локальности к единице автоматически приводит к локальной к нулю корреляции между эндогенной переменной и инструментом, что напоминает аналитический инструментальный слабых инструментов (Staiger & Stock 1997, Паган 2007), но здесь корреляция между эндогенной переменной и инструментом сходится к нулю со скоростью  $n$  из-за почти нестационарности.

Подставляя (13) и (14) в инструментальную оценку  $b_{12}^{(0)}$ , равную

$$\hat{b}_{12}^{(0)} = \frac{\sum_{t=2}^n y_{2,t-1} \Delta y_{1,t}}{\sum_{t=2}^n y_{2,t-1} \Delta y_{2,t}},$$

получим

$$\hat{b}_{12}^{(0)} - b_{12}^{(0)} = \frac{n^{-1} \sum_{t=2}^n y_{2,t-1} \varepsilon_{1,t}}{cn^{-2} \sum_{t=2}^n y_{2,t-1}^2 + n^{-1} \sum_{t=2}^n y_{2,t-1} u_{2,t}}.$$

Предельное распределение инструментальной оценки  $b_{12}^{(0)}$  задается как (Gospodinov 2010)

$$\hat{b}_{12}^{(0)} - b_{12}^{(0)} \Rightarrow \frac{d_2 \delta \int_0^1 J_c(s) dV_1(s) + \sqrt{1 - \delta^2} \int_0^1 J_c(s) dV_2(s)}{d_1 \left( c \int_0^1 J_c(s)^2 ds + \int_0^1 J_c(s) dV_1(s) \right)},$$

где  $V_1(r)$  и  $V_2(r)$  — независимые стандартные броуновские движения,

$$J_c(r) = \exp(cr) \int_0^r \exp(-cs) dV_1(s)$$

— процесс Орнштейна–Уленбека,

$$\delta^2 = \frac{(b_{21}^{(0)})^2 \sigma_1^2}{(b_{21}^{(0)})^2 \sigma_1^2 + \sigma_2^2}$$

— квадрат коэффициента корреляции между  $\varepsilon_{1,t}$  и  $u_{2,t}$ , и  $d_2/d_1$  — отношение стандартных отклонений  $\varepsilon_{1,t}$  и  $u_{2,t}$ , соответственно.

Этот результат показывает, что инструментальная оценка  $b_{12}^{(0)}$  несостоятельна. Причина несостоятельности в том, что  $y_{2,t-1}$  является слабым инструментом в том смысле, что несет очень мало информации об эндогенной переменной  $\Delta y_{2,t}$ . В свою очередь, слабость инструмента возникает из-за персистентной природы процесса  $y_{2,t}$ , моделируемого как имеющего локальный к единице корень. Более того, предельное распределение  $\hat{b}_{12}^{(0)} - b_{12}^{(0)}$  нестандартно. Во-первых, числитель предельного распределения — смесь гауссовой случайной величины и функционала от процесса Орнштейна–Уленбека с весами, определяющимися корреляцией между первым структурным шоком и шоком приведенной формы второго уравнения. Во-вторых, предельное распределение — отношение двух случайных величин, ибо знаменатель тоже случайная величина, содержащая функционалы процесса Орнштейна–Уленбека.

Полезно взглянуть на ситуацию, когда  $y_{2,t}$  приближается к процессу с точным единичным корнем ( $c \rightarrow 0$ ). В этом случае распределение  $\hat{b}_{12}^{(0)} - b_{12}^{(0)}$  имеет предел

$$\frac{d_2}{d_1} \left( \delta + \sqrt{1 - \delta^2} \frac{\int_0^1 V_1(s) dV_2(s)}{\int_0^1 V_1(s) dV_1(s)} \right).$$

Более того, если  $\delta = 1$ , предельное распределение  $\hat{b}_{12}^{(0)} - b_{12}^{(0)}$  имеет точечную вероятностную массу в точке  $d_2/d_1$ . В случае другой крайности  $\delta = 0$  распределение  $\hat{b}_{12}^{(0)} - b_{12}^{(0)}$  сходится к отношению стандартного нормального распределения

$$\frac{\int_0^1 V_1(s) dV_2(s)}{\left( \int_0^1 V_1^2(s) ds \right)^{1/2}}$$

и распределения Дики–Фуллера

$$\frac{\int_0^1 V_1(s) dV_1(s)}{\left( \int_0^1 V_1^2(s) ds \right)^{1/2}}.$$

Асимптотические распределения оценок остальных структурных параметров и импульсные отклики можно получить аналогично (Gospodinov 2010). Оказывается, что проблема слабого инструмента, которая лежит в основе несостоятельности  $\hat{b}_{12}^{(0)}$ , также портит оценивание  $b_{21}^{(0)}$  и импульсные отклики, делая их оценки несостоятельными. Подробности и улучшенную процедуру инференции можно найти в Gospodinov (2010).

## Список литературы

- Паган, А. (2007). Слабые инструменты. *Квантиль* 2, 71–81.
- Andrews, D.W.K. (1993). Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica* 61, 139–165.
- Chan, N.H. & C.Z. Wei (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050–1063.
- Davis, R.A. & W.T.M. Dunsmuir (1996). Maximum likelihood estimation for MA(1) processes with a root on or near the unit circle. *Econometric Theory* 12, 1–29.
- Elliott, G. (1998). On the robustness of cointegration methods when regressors almost have unit roots. *Econometrica* 66, 149–158.
- Gospodinov, N. (2002a). Median unbiased forecasts for highly persistent autoregressive processes. *Journal of Econometrics* 111, 85–101.
- Gospodinov, N. (2002b). Bootstrap-based inference in models with a nearly noninvertible moving average component. *Journal of Business & Economic Statistics* 20, 254–268.
- Gospodinov, N. (2004). Asymptotic confidence intervals for impulse responses of near-integrated processes. *Econometrics Journal* 7, 505–527.
- Gospodinov, N. (2009). A new look at the forward premium puzzle. *Journal of Financial Econometrics* 7, 312–338.
- Gospodinov, N. (2010). Inference in nearly nonstationary SVAR models with long-run identifying restrictions. *Journal of Business & Economic Statistics* 28, 1–12.
- Hansen, B.E. (1995). Rethinking the univariate approach to unit root tests: how to use covariates to increase power. *Econometric Theory* 11, 1148–1171.
- Hansen, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.
- Hansen, B.E. (1999). The grid bootstrap and the autoregressive model. *Review of Economics and Statistics* 81, 594–607.
- Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica* 75, 1411–1452.
- Phillips, P.C.B. (1979). The sampling distribution of forecasts from a first-order autoregression. *Journal of Econometrics* 9, 241–261.
- Phillips, P.C.B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.
- Phillips, P.C.B. (1988). Regression theory for near-integrated time series. *Econometrica* 56, 1021–1043.
- Phillips, P.C.B. (1991). Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.
- Phillips, P.C.B. (1998). Impulse response and forecast error variance asymptotics in nonstationary VARs. *Journal of Econometrics* 83, 21–56.
- Rossi, B. (2005). Confidence intervals for half-life deviations from purchasing power parity. *Journal of Business & Economic Statistics* 23, 432–442.
- Stock, J.H. (1991). Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics* 28, 435–459.
- Stock, J.H. (1994). Unit roots, structural breaks and trends. Глава 46 в *Handbook of Econometrics* под редакцией Engle, R. & D. McFadden, том 4, стр. 2739–2841. North-Holland: Amsterdam.
- Stock, J.H. (1996). VAR, error-correction and pretest forecasts at long horizons. *Oxford Bulletin of Economics and Statistics* 58, 685–701.
- Stock, J. H. (1997). Cointegration, long-run comovements and long horizon forecasting. Глава в *Advances in Econometrics: Proceedings of the Seventh World Congress of the Econometric Society* под редакцией Krepс, D. & K.F. Willis, 34–60. Cambridge University Press: Cambridge.
- Stock, J.H. & M.W. Watson (2008). What's new in econometrics – time series. *NBER Summer Institute*.
- Stock, J.H., J.H. Wright & M. Yogo (2002). A Survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20, 518–529.

## Asymptotics of near unit roots

**Stanislav Anatolyev**

*New Economic School, Moscow, Russia*

**Nikolay Gospodinov**

*Concordia University, Montréal, Canada*

Sometimes the conventional asymptotic theory yields that the limiting distribution changes discontinuously, or that the asymptotic distribution does not approximate accurately the actual finite-sample distribution. In such situations one finds useful an asymptotic tool of drifting parameterizations where certain parameters are allowed to depend explicitly on the sample size. It proves useful, among other things, for impulse response analysis and forecasting of strongly dependent processes at long horizons. This essay provides a review of these alternative asymptotic approximations in the context of time series models.

