Прогноз и фильтрация коррелированных временных рядов

Метенева Антонина Викторовна, гр. 522

Санкт-Петербургский государственный университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель — к.ф.-м.н., доц. Т.М. Товстик
Рецензент — к.ф.-м.н., доц. Н.П. Алексеева

Санкт-Петербург
2014г.

Постановка задачи

Даны временные ряды, состоящие из реальных наблюдений — курсы изменения доллара, евро, золота и серебра. Основная задача — построить прогноз исходных данных. Для этого необходимо:

1. установить взаимосвязь между временными рядами, используя корреляционный анализ;
2. подобрать процессы авторегрессии;
3. используя критерий Дики—Фуллера, исследовать ряды на стационарность;
4. построить прогноз для стационарных рядов;
5. проверить пары нестационарных рядов на коинтегрированность;
6. при помощи фильтрации найти прогноз коинтегрированных нестационарных временных рядов.

Основные определения

Определение

Процесс авторегрессии порядка n *— это случайный процесс X_t, который удовлетворяет разностному уравнению*

$$X_t + a_1 \cdot X_{t-1} + a_2 \cdot X_{t-2} + a_3 \cdot X_{t-3} + \cdots + a_n \cdot X_{t-n} = b_0 \cdot \xi_t, \quad (1)$$

где ξ_t при различных t являются независимыми случайными величинами с $\mathbb{E}\xi_t = 0$, $\mathbb{D}\xi_t = 1$, $\mathbb{E}\xi_t \cdot \xi_s = 0$, при $t \neq s$.

Определение

Процесс авторегрессии (1) ***стационарен****, если все корни характеристического уравнения $q(z) = 1 + a_1 \cdot z + \cdots + a_n \cdot z^n$ по модулю больше 1.*

Основные определения

Определение

Нестационарные ряды X_t, k-тые разности $X_t^{(k)}$ которых являются стационарными, называются ***интегрированными k-го порядка****.*

$$X_t^{(k)} = \begin{cases} X_t^{(k-1)} - X_{t-1}^{(k-1)} & k > 1 \\ X_t - X_{t-1} & k = 1 \end{cases}$$

Определение

Интегрированные ряды первого порядка X_t и Y_t называются ***коинтегрированными****, если существует вектор $\beta = (\beta_1, \beta_2) \neq 0$, для которого $Z_t = \beta_1 \cdot X_t + \beta_2 \cdot Y_t$ — стационарный ряд.*

Исходные данные

Исходными данными послужили временные ряды — курсы доллара, евро, золота, серебра.

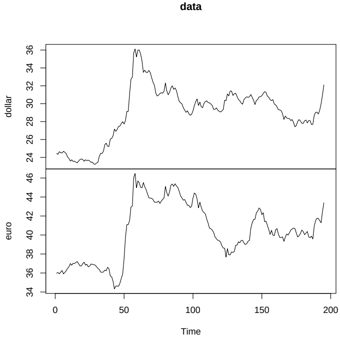

Рис. 1: Данные после усреднения - доллар, евро

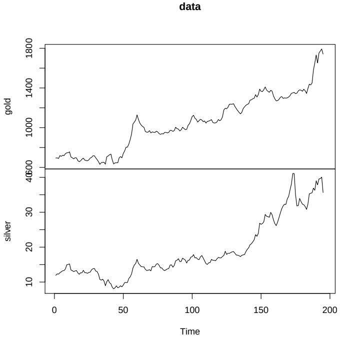

Рис. 2: Данные после усреднения - золото, серебро

Каждый из представленных рядов имеет $N = 196$ наблюдений. Единица деления равна 7 дням.

Коэффициенты корреляции для доллара представлены на Рис. 3, для евро — на Рис. 4.

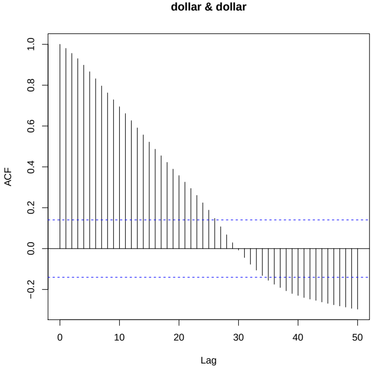

Рис. 3: Доллар

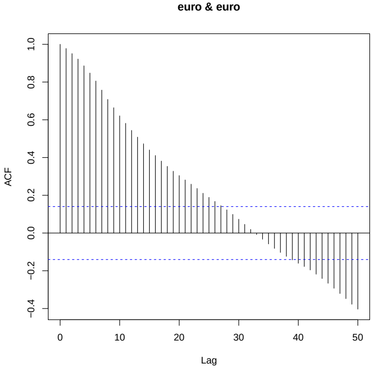

Рис. 4: Евро

Взаимные коэффициенты корреляции

Наибольшая зависимость была выявлена у пар: доллар — евро, золото —серебро, что продемонстрировано на Рис. 5 и Рис. 6. Иную взаимосвязь наблюдаем между долларом и серебром — Рис. 7.

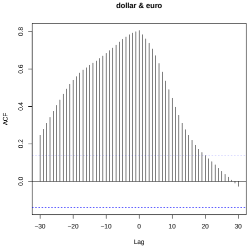

Рис. 5: Значения взаимных выборочных коэффициентов корреляций между долларом и евро

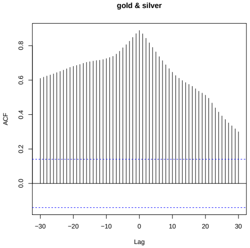

Рис. 6: Значения взаимных выборочных коэффициентов корреляций между золотом и серебром

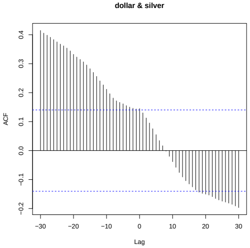

Рис. 7: Значения взаимных выборочных коэффициентов корреляций между долларом и серебром

Усеченные данные. Доллар — евро

Усеченные временные ряды $N = 65$ представлены на Рис. 8. Значения взаимных выборочных коэффициентов корреляции демонстрирует Рис. 9.

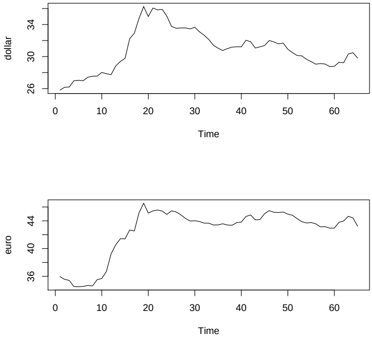

Рис. 8: Усеченные данные — доллар, евро

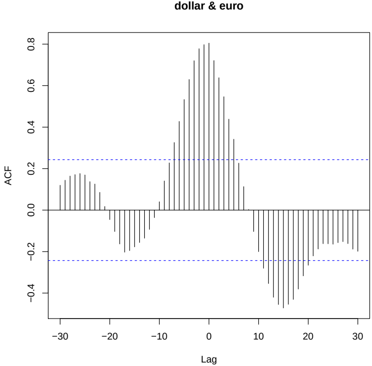

Рис. 9: Значения взаимных выборочных коэффициентов корреляций между долларом и евром

Проверка на наличие единичного корня

Стандартный критерий Дики—Фуллера, проверяет наличие единичного корня у процесса авторегрессии первого порядка $X_t = \widehat{a}_1 \cdot X_{t-1} + \xi_t$.
Основная гипотеза

$$H_0 : a_1 = 1, \tag{2}$$

т.е. характеристический многочлен имеет единичный корень и процесс X_t не является стационарным.
Гипотеза (2) отвергается в пользу альтернативы $H_1 : a_1 < 1$ при значениях

$$\widehat{a}_1 < \widehat{a}_{1\text{крит}}. \tag{3}$$

Оценка $\widehat{a}_{1\text{крит}}$ для критерия Дики—Фуллера

Моделируем процесс $X_t = a \cdot X_{t-1} + \xi_t$ при $a = 1$,
где ξ_t — независимые стандартные нормально распределенные случайные величины, тогда $X_0 = \xi_0$.

$$X_t = \sum_{k=0}^{t} \xi_k, t = 0, \cdots, N. \tag{4}$$

По методу наименьших квадратов находим оценку

$$\widehat{a}_1 = \frac{\sum\limits_{k=1}^{N} X_t X_{t-1}}{\sum\limits_{k=1}^{N} X_{t-1}^2} \tag{5}$$

Реализуем $K = 10000$ раз. Получаем оценки $\widehat{a}_1(j)$, где $j = 1, \cdots, 10000$. Упорядочиваем их в порядке возрастания: $\widehat{a}_1^*(j) < \cdots < \widehat{a}_K^*(j)$.
При 5% уровне значимости получаем $j = 10000 \cdot 0.05 = 500$ и полагаем $\widehat{a}_{1\text{крит}} = \widehat{a}_1^*(500)$.
При $N = 65$: $\widehat{a}_{1\text{крит}} = 0.880$. При $N = 196$: $\widehat{a}_{1\text{крит}} = 0.959$.

Расширенный критерий Дики—Фуллера

Рассмотрим авторегрессию n порядка

$$X_t = a_1 \cdot X_{t-1} + a_2 \cdot X_{t-2} + .. + a_n \cdot X_{t-n} + b_0 \cdot \xi_t. \tag{6}$$

После преобразования получаем:

$$X_t = \rho \cdot X_{t-1} + (\theta_1 \cdot \Delta X_{t-1} + .. + \theta_{n-1} \cdot \Delta X_{t-n+1}) + b_0 \cdot \xi_t, \tag{7}$$

где $\rho = a_1 + a_2 + .. + a_n$, $\theta_j = -(a_{j+1} + .. + a_n)$.
Основная гипотеза принимает вид: $H_0 : \rho = 1$
Полученное уравнение авторегрессии для доллара:

$$X_t = 1.005 \cdot X_{t-1} + 0.096 \cdot X_{t-2} - 0.178 \cdot X_{t-3} + 1.374 \cdot \xi_t. \tag{8}$$

Для доллара имеем $\widehat{\rho} = 0.885$, для объема выборки $N = 65$ критическое значение $\widehat{\rho}_{\text{крит}} = 0.880$, т.е. процесс (8) является нестационарным.
Самый близкий корень к единичному равен 1.083.

Прогноз для процесса авторегрессии

$$X_t = a_1 \cdot X_{t-1} + a_2 \cdot X_{t-2} + a_3 \cdot X_{t-3} + .. + a_n \cdot X_{t-n} + b_0 \cdot \xi_t \qquad (9)$$

имеет вид:

$$\widehat{X}(t,\tau) = \sum_{k=1}^{n} \alpha_{k,\tau} \cdot X_{t+1-k}, \qquad (10)$$

где $\alpha_{k,\tau}$ находятся из (11), $q(z)$— характеристическое уравнение (9).

$$\frac{1}{q(z)} = \sum_{k=0}^{\tau-1} \alpha_k \cdot z^k + z^{\tau-1} \cdot \frac{\alpha_{1,\tau} \cdot z + \alpha_{2,\tau} \cdot z^2 + \cdots + \alpha_{n,\tau} \cdot z^n}{q(z)}. \qquad (11)$$

Таблица 1: Сравнение прогноза $\widehat{X}(t,\tau)$ с наблюдением $X(t+\tau)$ для доллара

τ	$\widehat{X}(t,\tau)$	$X(t+\tau)$	модуль расхождения	ошибка в процентном соотношении, %
1	28,477	29,809	1,332	4,468
2	29,205	30,185	1,080	3,578
3	28,423	29,658	1,235	4,164

Нестационарные ряды. Прогноз

Теорема

Если есть два нестационарных коинтегрированных процесса — $X(t)$ и $Y(t)$, а

$Z^k(t) = X(t) - A \cdot Y(t-k), k > 0$ *— стационарная линейная комбинация,*

тогда прогноз $\widehat{X}(t,\tau)$ нестационарного процесса $X(t)$ на τ шагов вперед равен

$$\widehat{X}(t,\tau) = \widehat{Z}^k(t,\tau) + A \cdot Y(t+\tau-k), \text{где } \tau \leq k. \tag{12}$$

Прогноз нестационарного процесса

Рассмотрим комбинацию усеченных коинтегрированных нестационарных рядов при $k = 3$:

$$Z_t^3 = X_t^d - 0.933 \cdot Y_{t-3}^e, \qquad (13)$$

где X_t^d — доллар, Y_{t-3}^e — евро.
Полученные корни $z_1 = 1.159 - 3.041i$, $z_2 = 1.159 + 3.041i$, $z_3 = 1.205$, следовательно, ряд Z_t^3 стационарен.

$$\text{Прогноз: } \widehat{X}_t^d = \widehat{Z}_t^3 - 0.933 \cdot Y_{t-3}^e. \qquad (14)$$

Таблица 2: Сравнение прогноза $\widehat{X}(t, \tau)$ при $k = 3$ с наблюдением $X(t + \tau)$ для доллара по евро

τ	$\widehat{X}(t, \tau), k = 3$	$X(t + \tau)$	модуль расхождения при $k = 3$	ошибка в процентном соотношении, %
1	$29,783$	$29,809$	$0,026$	$0,087$
2	$29,205$	$30,185$	$0,980$	$3,247$
3	$28,471$	$29,658$	$1,187$	$4,002$

Результаты

- Изучена взаимосвязь исходных экономических временных рядов.
- Проведена проверка на стационарность каждого из рядов двумя способами:
  - решением соответствующих характеристических уравнений,
  - с помощью критерия Дики—Фуллера.
- Представлены методы прогнозирования стационарных и нестационарных временных рядов.
- Для коинтегрированных рядов найдена формула фильтрации, т.е. нахождение прогноза одного нестационарного процесса по другому.