

Квантиль

*международный эконометрический журнал
на русском языке*

**№9
июль 2011 г.**

СОДЕРЖАНИЕ ВЫПУСКА

Эконометрический ликбез: временные ряды

Цыплаков Александр. Введение в моделирование в пространстве состояний	1
Хейфец Игорь. Тестирование распределений	25

Задачи и решения

Решения 8.1, 8.2, 8.3	35
-----------------------	----

Статьи: финансовая эконометрика

Балаев Алексей. Моделирование многомерных параметрических плотностей финансовых доходностей	39
Колоколов Алексей. Хеджирование фьючерсами: многомерные GARCH с динамическими условными корреляциями	61

Квантиль

№9, июль 2011 г.

Сайт в Интернете: <http://quantile.ru>

Адрес электронной почты: quantile@quantile.ru

Доступ к журналу бесплатный и неограниченный

РЕДАКТОР

Станислав Анатольев

Российская Экономическая Школа (Москва, Россия)

РЕДАКЦИОННЫЙ СОВЕТ

Виктория Зинде-Уолш

Университет МакГилл (Монреаль, Канада)

Рустам Ибрагимов

Гарвардский университет (Кэмбридж, США)

Анна Микушева

Массачусетский технологический институт (Кэмбридж, США)

Алексей Онацкий

Кэмбриджский университет (Кэмбридж, Великобритания)

Константин Тюрин

Университет штата Индиана (Блумингтон, США)

Александр Цыплаков

Новосибирский государственный университет (Новосибирск, Россия)

Виктор Черножуков

Массачусетский технологический институт (Кэмбридж, США)

К СВЕДЕНИЮ АВТОРОВ

Рукописи для публикации в разделе «Статьи» принимаются в электронном виде по адресу submit@quantile.ru. Работы могут принадлежать любой прикладной сфере экономической науки. Главным требованием является интенсивное использование адекватных эконометрических методов. Рукопись должна быть написана на русском (для русскоязычных авторов) или на английском (для остальных авторов) языке в формате *Microsoft Word* или (предпочтительнее) *LaTeX*, и по объему не превышать 30 страниц формата А4 с двойным междустрочным интервалом. Работы подвергаются контролю качества членами редакционного совета и независимыми референтами. Перспективная работа может быть при необходимости возвращена автору на доработку. Редакция также приглашает к сотрудничеству экспертов по эконометрике, готовых внести вклад в методологические рубрики журнала.

При публикации статьи или методологического эссе в журнале «Квантиль» передача авторских прав не происходит ни полностью, ни частично.

Решения задач из рубрики «Задачи и решения», а также новые задачи можно присылать по адресу ps@quantile.ru.

Эконометрический ликбез: временные ряды

Введение в моделирование в пространстве состояний*

Александр Цыплаков[†]

Новосибирский государственный университет, Новосибирск, Россия

Многие модели временных рядов, и прежде всего это относится к различным моделям с ненаблюдаемыми компонентами, можно представить в так называемой форме пространства состояний. Модель пространства состояний — это мощный инструмент, который позволяет применить к исходной модели широкий спектр стандартных процедур, включая оценивание и прогнозирование. Статья предлагает обзор этого универсального класса моделей и соответствующих процедур.

1 Введение

Модель пространства состояний задается двумя условными распределениями:¹

$$F(y_t | a_{\leq t}, y_{< t}) = F(y_t | a_t), \quad (1)$$

$$F(a_t | a_{< t}, y_{< t}) = F(a_t | a_{t-1}). \quad (2)$$

Через y_t обозначен вектор $k_t \times 1$ наблюдаемых данных в момент t . В основе наблюдений y_t лежит динамика некоторой «системы». Вектор a_t ($m_t \times 1$) — это значение переменной состояния системы в момент t . Предполагается, что состояние не наблюдаемо в явном виде. Распределение состояния системы в момент t определяется состоянием системы в предыдущий момент (a_{t-1}), что задается формулой (2), т. е. динамика системы носит марковский характер. Распределение (1) определяет, как наблюдаемая величина y_t зависит от ненаблюдаемого состояния a_t . Если модель относится к конечному ряду, то $t = 1, \dots, T$ и при $t = 1$ вместо (2) следует задать начальное распределение $F(a_1)$. На Рис. 1 такая модель изображена в виде условной схемы.

Многие методы, относящиеся к моделям вида (1)–(2), приложимы и к более общим моделям, в которых условные распределения a_t и y_t зависят от лагов y_t , т. е. имеют вид $F(y_t | a_t, y_{< t})$ и $F(a_t | a_{t-1}, y_{< t})$.

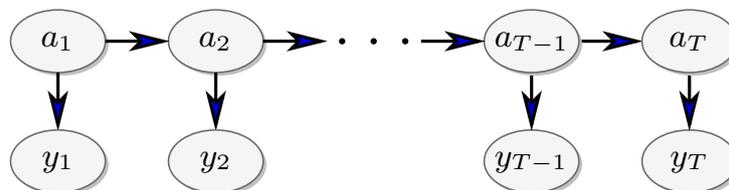


Рис. 1: Схема модели пространства состояний

*Цитировать как: Цыплаков, Александр (2011) «Введение в моделирование в пространстве состояний», Квантиль, №9, стр. 1–24. Citation: Tsyplakov, Alexander (2011) “An introduction to state space modeling,” *Quantile*, No.9, pp. 1–24.

[†]Адрес: 630090, г. Новосибирск, Весенний проезд, 6–44. Электронная почта: tsy@academ.org

¹Здесь использованы следующие обозначения для отрезков ряда $\{z_t\}$: $z_{< t} = (z_1, \dots, z_{t-1})$ и $z_{\leq t} = (z_1, \dots, z_t)$. Если ряд бесконечный, то $z_{\leq t} = (z_t, z_{t-1}, \dots)$ и аналогично для $z_{< t}$.

Наиболее часто используется (благодаря простоте вычислений) *линейная модель пространства состояний*. Она записывается в виде двух уравнений:

$$y_t = R_t^y + R_t^{ya} a_t + \epsilon_t^y, \quad (3)$$

$$a_t = R_t^a + R_t^{aa} a_{t-1} + \epsilon_t^a. \quad (4)$$

Здесь (4) называется *уравнением перехода* (*transition equation*), а (3) — *уравнением измерения* (*measurement equation*). Когда модель применяется к конечному временному ряду, предполагается, что индекс t пробегает от 1 до T и что при $t = 1$ член $R_t^{aa} a_{t-1}$ отсутствует.

В базовой формулировке ошибки ϵ_t^y , ϵ_t^a независимы как между собой, так и по времени. Ошибки ϵ_t^y , ϵ_t^a имеют нулевые математические ожидания и ковариационные матрицы Ω_t^y , Ω_t^a . В гауссовской линейной модели пространства состояний распределения переменных являются нормальными:

$$y_t | a_{\leq t}, y_{< t} \sim \mathcal{N}(R_t^y + R_t^{ya} a_t, \Omega_t^y), \quad (5)$$

$$a_t | a_{< t}, y_{< t} \sim \mathcal{N}(R_t^a + R_t^{aa} a_{t-1}, \Omega_t^a). \quad (6)$$

Часто в литературе уравнение перехода встречается в сдвинутой вперед по времени форме:

$$a_{t+1} = R_t^a + R_t^{aa} a_t + \epsilon_t^a.$$

Каждая из форм имеет свои удобства и недостатки, но при независимости ошибок ϵ_t^y , ϵ_t^a они эквивалентны между собой. Однако в варианте модели пространства состояний с коррелированными ошибками ϵ_t^y , ϵ_t^a такой сдвиг уже принципиально меняет модель.

Вводные обсуждения моделирования в пространстве состояний можно найти в Comman-deur & Koopman (2007), Durbin (2004), Harvey & De Rossi (2007), О'Коннелл (1990). Для дальнейшего чтения можно порекомендовать монографии Durbin & Koopman (2001), Kim & Nelson (1999), Carré, Moulines & Rydén (2007) и сборник Harvey & Proietti (2005).

2 Некоторые примеры

2.1 Регрессия с меняющимися коэффициентами

В обычной модели линейной регрессии, примененной к временным рядам y_t , x_t ,

$$y_t = x_t^\top \beta_t + \epsilon_t$$

предполагается, что коэффициенты не меняются по времени: $\beta_t = \beta$ ($m \times 1$). Можно модифицировать эту модель, предположив, что коэффициенты подчинены независимым процессам случайного блуждания:

$$\beta_{tj} = \beta_{t-1,j} + \eta_{tj}, \quad j = 1, \dots, m,$$

приращения которых независимы и имеют дисперсии $\mathbb{V}(\eta_{tj}) = \sigma_j^2$. Такую модель можно представить как модель пространства состояний, если предположить, что $a_t = \beta_t$, $R_t^y = 0$, $R_t^{ya} = x_t^\top$, $\Omega_t^y = \sigma_\epsilon^2$, $R_t^a = 0_m^\top$, $R_t^{aa} = I_m$, $\Omega_t^a = \text{diag}\{\sigma_j^2\}$.

2.2 Модель ARMA

Модель ARMA(p, q)

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q},$$

можно записать в форме пространства состояний. Сначала ARMA(p, q) представляется как ARMA($M, M - 1$), где $M = \max(p, q + 1)$. «Недостающие» коэффициенты просто считаются

равным и нулю: $\phi_j = 0$ при $j > p$ и $\theta_j = 0$ при $j > q$. Тогда модель переводится в пространство состояний с помощью следующих переобозначений:

$$a_{t1} = y_t, \quad a_{tj} = \sum_{i=j}^M [\phi_i y_{t+j-i-1} + \theta_{i-1} \epsilon_{t+j-i}],$$

$$R_t^y = 0, \quad R_t^{ya} = (1, 0, \dots, 0), \quad \Omega_t^y = 0,$$

$$R_t^a = 0_M, \quad R_t^{aa} = \begin{pmatrix} \phi_1 & & & \\ & \ddots & & \\ & & I_{M-1} & \\ \phi_{M-1} & & & \\ \phi_M & & & 0_{M-1}^\top \end{pmatrix}, \quad \Omega_t^a = \sigma_\epsilon^2 s s^\top, \quad s = (1, \theta_1, \dots, \theta_{M-1})^\top.$$

В частном случае ARMA(2, 1) этот метод дает следующую модель пространства состояний ($z_t = \phi_2 y_{t-1} + \theta_1 \epsilon_t$):

$$\begin{aligned} y_t &= (1, 0) \begin{pmatrix} y_t \\ z_t \end{pmatrix} + 0, \\ \begin{pmatrix} y_t \\ z_t \end{pmatrix} &= \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ \theta_1 \epsilon_t \end{pmatrix}. \end{aligned} \tag{7}$$

2.3 Структурные модели временных рядов

В так называемых структурных моделях временных рядов наблюдаемый временной ряд представляется как сумма отдельных ненаблюдаемых компонент, каждая из которых имеет некоторую содержательную интерпретацию: тренд, сезонность, шум и т. д. См. напр. Harvey (1989) и Harvey & Shephard (1993). Следует отметить, что название «структурные модели» излишне нагружено. Не предполагается, что модель отражает структурные взаимосвязи между экономическими показателями, а декомпозиция ряда основана на чисто технических характеристиках компонент, таких как представление в виде случайного блуждания или белого шума, и т. п.

Простейшая подобного рода модель — *модель локального уровня*. Она служит обобщением модели $y_t = \mu + \epsilon_t$, где μ — константа, ϵ_t — белый шум. Идея состоит в том, что μ может со временем постепенно меняться. Модель локального уровня имеет следующий вид:

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, \\ \mu_t &= \mu_{t-1} + \xi_t. \end{aligned} \tag{8}$$

Здесь ϵ_t и ξ_t — два независимых ряда гауссовских случайных возмущений, имеющих дисперсии σ_ϵ^2 и σ_ξ^2 соответственно, а μ_t — меняющийся во времени ненаблюдаемый локальный уровень. Здесь μ_t представляет собой случайное блуждание, поэтому подобный процесс также называют случайным блужданием с шумом. В обозначениях модели (3)–(4) имеем $a_t = \mu_t$, $R_t^y = 0$, $R_t^{ya} = 1$, $\Omega_t^y = \sigma_\epsilon^2$, $R_t^a = 0$, $R_t^{aa} = 1$, $\Omega_t^a = \sigma_\xi^2$.

Ту же модель можно было бы представить в виде

$$\Delta y_t = \epsilon_t - \epsilon_{t-1} + \xi_t,$$

что можно свести к модели ARIMA(0,1,1). Но в таком виде в модели уже не будет явным образом фигурировать меняющийся уровень, поэтому она потеряет свою «структурность» и станет моделью в приведенной форме.

Модель локального линейного тренда служит таким же обобщением линейного тренда, как модель локального уровня (8) обобщением константы:

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, \\ \mu_t &= \mu_{t-1} + \nu_{t-1} + \xi_t, \\ \nu_t &= \nu_{t-1} + \zeta_t. \end{aligned} \quad (9)$$

Переменная μ_t отвечает за тренд, а ν_t — за наклон этого тренда. Здесь $a_t = (\mu_t, \nu_t)^\top$, $R_t^y = 0$, $R_t^{y^a} = (1, 0)$, $\Omega_t^y = \sigma_\epsilon^2$, $R_t^a = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $R_t^{aa} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $\Omega_t^a = \text{diag}(\sigma_\xi^2, \sigma_\zeta^2)$.

Компонента γ_t , представляющая сезонность на основе фиктивных переменных, задается следующим уравнением:

$$\gamma_t = -\gamma_{t-1} - \dots - \gamma_{t-s+1} + \omega_t, \quad (10)$$

где s — количество сезонов. Введя переменные $\delta_{tj} = \gamma_{t-j+1}$, процесс (10) можно переписать в следующем виде:

$$\begin{aligned} \delta_{t1} &= -\delta_{t-1,1} - \dots - \delta_{t-1,s-1} + \omega_t, \\ \delta_{tj} &= \delta_{t-1,j-1}, \quad j = 2, \dots, s, \end{aligned}$$

что позволяет его включить как компоненту в линейную модель пространства состояний. Заметим, что (10) — это обобщение обычной детерминированной сезонной компоненты

$$\gamma_t = \delta_1 D_{t1} + \dots + \delta_s D_{ts},$$

где D_{tj} — фиктивная переменная для j -го сезона, с ограничением, что сезонные эффекты в сумме равны нулю: $\delta_1 + \dots + \delta_s = 0$. Детерминированный случай получается при $\sigma_\omega^2 = 0$.

Тригонометрическая сезонная компонента представляет собой сумму элементарных компонент

$$\gamma_t = \sum_{j=1}^{\lfloor s/2 \rfloor} \gamma_{tj}, \quad (11)$$

где

$$\begin{aligned} \gamma_{tj} &= \cos(\lambda_j) \gamma_{t-1,j} + \sin(\lambda_j) \gamma_{t-1,j}^* + \omega_{tj}, \quad j = 1, \dots, \lfloor s/2 \rfloor, \\ \gamma_{tj}^* &= -\sin(\lambda_j) \gamma_{t-1,j} + \cos(\lambda_j) \gamma_{t-1,j}^* + \omega_{tj}^*, \quad j = 1, \dots, \lfloor (s-1)/2 \rfloor, \end{aligned} \quad (12)$$

соответствуют j -й сезонной частоте $\lambda_j = 2\pi j/s$. Дисперсия ошибки обычно предполагается одинаковой для каждой подкомпоненты: $\mathbb{V}(\omega_{tj}) = \mathbb{V}(\omega_{tj}^*) = \sigma_\omega^2$.

Стохастический цикл — это компонента ψ_t , которая вместе с парной вспомогательной компонентой ψ_t^* задается формулами

$$\begin{aligned} \psi_t &= \rho \cos(\lambda_c) \psi_{t-1} + \rho \sin(\lambda_c) \psi_{t-1}^* + \kappa_t, \\ \psi_t^* &= -\rho \sin(\lambda_c) \psi_{t-1} + \rho \cos(\lambda_c) \psi_{t-1}^* + \kappa_t^*. \end{aligned}$$

Здесь $\rho \in (0; 1]$ — это коэффициент затухания, λ_c — частота цикла в радианах. κ_t и κ_t^* — случайные ошибки с одинаковой дисперсией σ_κ^2 . При $\rho = 1$ и $\sigma_\kappa^2 = 0$ получаем обычный детерминированный цикл:

$$\psi_t = \psi_0 \cos(\lambda_c t) + \psi_0^* \sin(\lambda_c t).$$

Различные компоненты, которые здесь описаны, представляют собой как бы «конструктор», из которого можно собрать общую модель временного ряда. Простой пример приведен ниже в пункте 3.3.3.

2.4 Выделение общего стохастического тренда

Так называемые динамические факторные модели, являющиеся обобщением обычного статистического факторного анализа, тесно связаны с моделями в пространстве состояний. С помощью факторных моделей можно выделять ненаблюдаемые факторы, являющиеся общими для целого ряда наблюдаемых рядов, строить различные индексы. Например, в Chang, Miller & Park (2009) рассматривается выделение из нескольких рядов общего стохастического тренда. Стохастический тренд моделируется как случайное блуждание $\mu_t = \mu_{t-1} + \xi_t$, а вектор наблюдаемых рядов задается как

$$y_t = \beta\mu_t + \epsilon_t,$$

где β — вектор факторных нагрузок. Чтобы модель была идентифицирована, следует задать какую-либо нормировку параметров, например, положить $\sigma_\xi^2 = 1$. В качестве одного из примеров Chang, Miller & Park (2009) взяли построение индекса фондового рынка по 30 рядам, составляющим индекс DJIA.

2.5 Моделирование временной структуры процентных ставок

В известной модели Нельсона—Сигеля (Nelson & Siegel (1987)) кривая доходности (*yield curve*), которая описывает зависимость доходности по ценным бумагам с фиксированной процентной ставкой от срока до погашения m , имеет следующий вид:

$$R(m) = \beta_0 + \beta_1 (1 - \exp(-m/\tau)) \frac{\tau}{m} + \beta_2 \left((1 - \exp(-m/\tau)) \frac{\tau}{m} - \exp(-m/\tau) \right).$$

При фиксированном значении τ эта функция линейна по параметрам $\beta_0, \beta_1, \beta_2$. Модель можно рассматривать как регрессию и получать соответствующие оценки, имея данные о доходностях бескупонных облигаций с разными сроками погашения. Однако для ее использования требуется, чтобы параметры не менялись. Это можно считать обоснованным предположением, только если оценивание кривой идет по данным из достаточно короткого временного промежутка, так что параметры остаются практически неизменными.

В динамических вариантах модели Нельсона—Сигеля коэффициенты $\beta_0, \beta_1, \beta_2$ меняются во времени в соответствии с векторной авторегрессией первого порядка, а коэффициент τ не меняется (Diebold & Li (2006), Diebold, Rudebusch & Aruoba (2006)). Получается линейная модель пространства состояний, в которой вектор состояния имеет вид $a_t = (\beta_{t0}, \beta_{t1}, \beta_{t2})^\top$. Для модели несложно получить оценки максимального правдоподобия фиксированных параметров, сглаженные оценки меняющихся коэффициентов и прогнозы. Удобно также, что не требуется дополнительных модификаций для данных с меняющейся во времени (например, из-за пропуска наблюдений) размерностью. Близкая модель, также имеющая вид линейной модели пространства состояний, обсуждается в Christensen, Diebold & Rudebusch (2011).

3 Инференция относительно состояния системы

Нас может интересовать состояние системы в момент t при том, что история наблюдений до момента s , т. е. y_1, \dots, y_s , известна. Инференция относительно a_t на основе y_1, \dots, y_s при $t > s$ называется предсказанием, при $t = s$ — фильтрацией, а при $t < s$ — сглаживанием. Рассмотрим сглаживание с фиксированным интервалом (см. определение ниже) и фильтрацию. Но прежде обсудим другой важный вопрос — каким образом можно задать параметры распределения вектора состояния в момент $t = 1$.

3.1 Инициализация модели пространства состояний

Если формулировка модели пространства состояний при $t > 1$ обычно непосредственно следует из формулировки исходной эконометрической модели, то $t = 1$ следует рассматривать как особый случай. Существует несколько различных подходов, которые позволяют задать начальные условия, т. е. матрицы R_1^a и Ω_1^a . Конечно, если значение переменной состояния в момент $t = 1$ точно известно, то следует просто положить $R_1^a = a_1$ и $\Omega_1^a = 0$. Но такое бывает редко.

Возможно, что начальное состояние точно не известно, но известно (например, из какого-то предыдущего опыта), что оно распределено нормально с некоторыми средним и ковариационной матрицей. Тогда для инициализации модели просто используются эти значения. Это уже фактически байесовский взгляд на модели пространства состояний. При таком подходе задаваемое распределение будет априорным распределением, а калмановскую фильтрацию, обсуждаемую ниже, можно рассматривать как байесовское обучение.

Частный случай — это так называемое расплывчатое априорное распределение: в качестве Ω_1^a берется «очень большая» матрица, что отражает большую степень неопределенности относительно исходного состояния. Часто берут $\Omega_1^a = cI$ при больших значениях c . При $c \rightarrow \infty$ получаем полностью расплывчатое распределение («про начальное состояние нет никакой информации»). Конечно, стандартные алгоритмы нельзя использовать непосредственно в предельном случае $c \rightarrow \infty$ и требуется искать им замену. Даже просто использование слишком больших значений c может привести к переполнению регистров при компьютерных расчетах.² Среднее R_1^a для расплывчатого априорного распределения выбирается более или менее произвольным образом, поскольку его влияние будет несущественным при достаточно сильной «расплывчатости».

Близкий подход состоит в том, чтобы включить начальное состояние a_1 в число неизвестных параметров и оценивать его. При этом $R_1^a = a_1$ и $\Omega_1^a = 0$. Такой подход имеет тот очевидный недостаток, что требуется оценивать большее число параметров, что замедляет оценивание.

Еще один подход годится для случайных процессов, задаваемых инвариантной по времени моделью пространства состояний:

$$y_t = R^y + R^{ya}a_t + \epsilon_t^y, \quad (13)$$

$$a_t = R^a + R^{aa}a_{t-1} + \epsilon_t^a, \quad (14)$$

где индекс t принимает значения от минус до плюс бесконечности, $\mathbb{V}(\epsilon_t^y) = \Omega^y$, $\mathbb{V}(\epsilon_t^a) = \Omega^a$. В предположении стационарности³ процесса a_t имеем

$$\begin{aligned} \mathbb{E}(a_t) &= R^a + R^{aa}\mathbb{E}(a_t), \\ \mathbb{V}(a_t) &= R^{aa}\mathbb{V}(a_t)(R^{aa})^\top + \Omega^a, \end{aligned}$$

откуда

$$R_1^a = \mathbb{E}(a_t) = (I - R^{aa})^{-1}R^a$$

и⁴

$$\text{vec}(\Omega_1^a) = \text{vec}(\mathbb{V}(a_t)) = (I - R^{aa} \otimes R^{aa})^{-1} \text{vec}(\Omega^a).$$

²Алгоритмы, призванные решить эти вычислительные проблемы, описаны, например, в Durbin & Koopman (2001).

³Условие стационарности процесса a_t состоит в том, что все корни характеристического уравнения $\det(I - zR^{aa}) = 0$ лежат за пределами единичной окружности на комплексной плоскости, или, что эквивалентно, собственные значения матрицы R^{aa} лежат в пределах единичной окружности.

⁴Здесь $\text{vec}(A)$ — вектор из стоящих друг под другом столбцов матрицы A , а \otimes — произведение Кронекера.

В случае диагональных матриц R^{aa} и Ω^a можно каждую компоненту вектора состояний a_t рассматривать отдельно, как независимый стационарный процесс AR(1), и составлять матрицы R_1^a и Ω_1^a из соответствующих средних и дисперсий.

3.2 Фильтр Калмана

Для модели временного ряда, представленной в форме пространства состояний, существует целый ряд стандартных алгоритмов. Самый известный из них — фильтр Калмана. Фильтр назван по имени венгерско-американского инженера-электрика Р. Калмана. Калмановская фильтрация широко используется для различных технических задач, например, для радиолокационного обнаружения, управления космическими аппаратами и т. п.

Фильтр Калмана настолько тесно связан с моделями в пространстве состояний, что его достаточно часто принимают за статистическую модель данных. Следует понимать, что модель — это модель пространства состояний, а фильтр Калмана — это всего лишь один из вычислительных алгоритмов, связанных с данной моделью (хотя и очень важный).

Непосредственная задача калмановской фильтрации состоит в том, чтобы получить оценки ненаблюдаемого состояния системы и сопутствующих величин, используя только информацию, содержащуюся в наблюдениях, полученных до момента t включительно, т. е. y_1, \dots, y_t . Таким образом, для получения оценок a_t не используется информация из будущего. Это рекуррентный алгоритм, последовательно вычисляющий подобные оценки начиная от $t = 1$ и заканчивая $t = T$. При этом предполагается, что все коэффициенты и ковариационные матрицы линейной модели пространства состояний (3)–(4) известны.

Применение фильтра Калмана для оценивания ненаблюдаемого состояния системы в реальном времени особенно важно в технических приложениях. Например, система видеонаблюдения за некоторым объектом может использовать калмановскую фильтрацию, чтобы предсказывать положение объекта в следующий момент времени и сохранять направленность видеокамеры на объект. В экономических приложениях такое непосредственное пошаговое использование фильтра Калмана довольно редко. В основном фильтр Калмана используется как промежуточный алгоритм, служащий частью других алгоритмов.

Обозначим через \mathbb{E}_t оператор математического ожидания, условного относительно информации, доступной в момент t . Это $y_{\leq t}$ — наблюдения до момента t включительно. Соответствующую условную ковариационную матрицу (в одномерном случае — дисперсию) обозначим через \mathbb{V}_t , а условную кросс-ковариационную матрицу двух векторов — через \mathbb{C}_t . Введем также следующие обозначения:

$$a_{t|\tau} = \mathbb{E}_\tau(a_t), \quad P_{t|\tau} = \mathbb{V}_\tau(a_t), \quad y_{t|\tau} = \mathbb{E}_\tau(y_t).$$

Чтобы вывести рекуррентные формулы фильтра Калмана, мы рассмотрим совместное распределение y_t и a_t , условное относительно информации, доступной до момента $t - 1$ включительно ($y_{<t}$). Это некоторое многомерное нормальное распределение:

$$\begin{pmatrix} y_t \\ a_t \end{pmatrix} | y_{<t} \sim \mathcal{N} \left(\begin{pmatrix} y_{t|t-1} \\ a_{t|t-1} \end{pmatrix}, \begin{pmatrix} F_t & G_t \\ G_t^\top & P_{t|t-1} \end{pmatrix} \right),$$

где

$$F_t = \mathbb{V}_{t-1}(y_t), \quad G_t = \mathbb{C}_{t-1}(y_t, a_t).$$

Таблица 1: Алгоритм калмановской фильтрации

Инициализация: $a_{1|0} = R_1^a$, $P_{1|0} = \Omega_1^a$.

$$a_{t|t-1} = R_t^a + R_t^{aa} a_{t-1|t-1}, \quad (15)$$

$$P_{t|t-1} = R_t^{aa} P_{t-1|t-1} (R_t^{aa})^\top + \Omega_t^a, \quad (16)$$

$$y_{t|t-1} = R_t^y + R_t^{ya} a_{t|t-1}, \quad (17)$$

$$G_t = R_t^{ya} P_{t|t-1}, \quad (18)$$

$$F_t = G_t (R_t^{ya})^\top + \Omega_t^y, \quad (19)$$

$$v_t = y_t - y_{t|t-1}, \quad (20)$$

$$K_t = G_t^\top F_t^{-1}, \quad (21)$$

$$a_{t|t} = a_{t|t-1} + K_t v_t, \quad (22)$$

$$P_{t|t} = P_{t|t-1} - K_t G_t. \quad (23)$$

Параметры этого распределения можно выразить следующим образом:

$$a_{t|t-1} = \mathbb{E}_{t-1}(a_t) = \mathbb{E}_{t-1}(R_t^a + R_t^{aa} a_{t-1} + \epsilon_t^a) = R_t^a + R_t^{aa} a_{t-1|t-1},$$

$$P_{t|t-1} = \mathbb{V}_{t-1}(a_t) = \mathbb{V}_{t-1}(R_t^{aa} a_{t-1}) + \mathbb{V}_{t-1}(\epsilon_t^a) = R_t^{aa} P_{t-1|t-1} (R_t^{aa})^\top + \Omega_t^a,$$

$$y_{t|t-1} = \mathbb{E}_{t-1}(y_t) = \mathbb{E}_{t-1}(R_t^y + R_t^{ya} a_t + \epsilon_t^y) = R_t^y + R_t^{ya} a_{t|t-1},$$

$$F_t = \mathbb{V}_{t-1}(y_t) = \mathbb{V}_{t-1}(R_t^{ya} a_t) + \mathbb{V}_{t-1}(\epsilon_t^y) = R_t^{ya} P_{t|t-1} (R_t^{ya})^\top + \Omega_t^y,$$

$$G_t = \mathbb{C}_{t-1}(y_t, a_t) = \mathbb{C}_{t-1}(R_t^{ya} a_t, a_t) + \mathbb{C}_{t-1}(\epsilon_t^y, a_t) = R_t^{ya} P_{t|t-1}.$$

Далее рассмотрим распределение a_t , условное относительно как y_t , так и y_1, \dots, y_{t-1} . Это тоже некоторое многомерное нормальное распределение:

$$a_t | y_1, \dots, y_t \sim \mathcal{N}(a_{t|t}, P_{t|t}).$$

Чтобы рассчитать его параметры, следует применить правило для условного распределения из многомерного нормального распределения:

$$a_{t|t} = a_{t|t-1} + G_t^\top F_t^{-1} (y_t - y_{t|t-1})$$

и

$$P_{t|t} = P_{t|t-1} - G_t^\top F_t^{-1} G_t.$$

Данные уравнения можно записать более компактно, введя обозначение

$$K_t = G_t^\top F_t^{-1}.$$

Эта матрица называется *калмановским коэффициентом усиления (Kalman gain)*.

Теперь пришло время собрать все рекуррентные соотношения фильтра Калмана в одном месте; см. Таблицу 1.

Дадим интерпретацию этих уравнений. Прежде всего отметим, что если векторы p и q имеют совместно многомерное нормальное распределение, то

$$\mathbb{V}(p|q) = \mathbb{V}[p - \mathbb{E}(p|q)].$$

Условное математическое ожидание $\mathbb{E}(p|q)$ — это наилучший в среднеквадратическом смысле прогноз величины p среди прогнозов, использующих только информацию о q , а $\mathbb{V}[p - \mathbb{E}(p|q)]$ — это ковариационная матрица ошибки такого прогноза, или, что то же самое, (матричный) средний квадрат ошибки прогноза.

В момент $t - 1$ имеется информация о y_1, \dots, y_{t-1} . Исходя из этой информации можно получить с помощью фильтра Калмана наилучшую оценку текущего состояния системы, т. е. $a_{t-1|t-1}$. Ковариационная матрица ошибки, связанной с данной оценкой, равна $P_{t-1|t-1}$. По (15) и (16) можно получить прогноз состояния системы на один шаг вперед, т. е. $a_{t|t-1}$, и ковариационную матрицу ошибки такого прогноза $P_{t|t-1}$. Далее, на основе $a_{t|t-1}$ и $P_{t|t-1}$ по (17) и (19) можно получить прогноз изучаемой переменной на один шаг вперед, т. е. $y_{t|t-1}$, и ковариационную матрицу ошибки такого прогноза F_t .

Когда наступает момент времени t , появляется дополнительное наблюдение y_t . Становится известной ошибка одношагового прогноза $v_t = y_t - y_{t|t-1}$. Эту ошибку называют *инновацией*, поскольку она несет в себе новую информацию о процессе y_t . Полученная информация позволяет по (22) и (23) скорректировать прогноз $a_{t|t-1}$ и рассчитать уточненную оценку текущего состояния системы ($a_{t|t}$). Ковариационная матрица ошибки, связанной с данной оценкой равна $P_{t|t}$.

Величина корректировки $K_t v_t$ в уравнении (22) зависит от ошибки прогноза и коэффициента усиления. Значение матрицы K_t определяет, в какой степени ошибка прогноза v_t влияет на коррекцию оценки состояния системы. В частности, если K_t близка к нулю, то оценка состояния практически не меняется. Это бывает, когда шум в текущем наблюдении велик (что измеряется ковариационной матрицей Ω_t^y) по сравнению с точностью текущей оценки состояния $P_{t|t-1}$.

Укажем также два важных свойства:

- Ковариационные матрицы и калмановские коэффициенты усиления не зависят от наблюдений за y_t . Вообще говоря, их можно рассчитать отдельно, до вычисления $a_{t|t-1}$, $y_{t|t-1}$ и $a_{t|t}$.
- Алгоритм инвариантен к умножению всех ковариационных матриц на одно и то же положительное число. Так если, например, в модели локального уровня разделить дисперсии σ_ϵ^2 и σ_ξ^2 на σ_ϵ^2 , то вычисленные $a_{t|t-1}$, $y_{t|t-1}$ и $a_{t|t}$ не поменяются.

3.3 Сглаживание

3.3.1 Сглаживающая схема Калмана

Как уже говорилось, само по себе пошаговое (онлайновое) оценивание ненаблюдаемого состояния редко бывает полезным для экономиста. Если интерес представляет оценка ненаблюдаемого состояния в конкретный момент времени t , то предпочтительнее использовать для получения такой оценки всю имеющуюся информацию y_1, \dots, y_T , а не только информацию, доступную в момент $t - 1$ или t , как при калмановской фильтрации. Задачу получения подобных оценок называют *сглаживанием*.

Заметим, что вообще говоря различают три разновидности сглаживания: сглаживание в фиксированной точке $a_{\tau|t}$, сглаживание с фиксированным лагом $a_{t-L|t}$ и сглаживание на фиксированном интервале $a_{t|T}$. (Предполагается, что τ , L и T фиксированы, а t меняется.) Здесь речь идет о сглаживании на фиксированном интервале.

Наиболее известный инструмент для сглаживания — это *сглаживающая схема Калмана* (*калмановский сглаживатель*, *Kalman smoother*). Калмановское сглаживание осуществляется в два прохода по наблюдениям. Первый проход осуществляется в прямом направлении по времени ($t = 1, \dots, T$) и совпадает с фильтром Калмана. Второй проход осуществляется также рекуррентно, но уже в обратном направлении по времени ($t = T, \dots, 1$) и результатом

Таблица 2: Алгоритм калмановского сглаживания

Проход в прямом направлении: применяется фильтр Калмана (Таблица 1).

Проход в обратном направлении ($t = T, \dots, 1$):

$$\begin{aligned} J_t &= P_{t|t}(R_{t+1}^{aa})^\top P_{t+1|t}^{-1}, \\ a_{t|T} &= a_{t|t} + J_t(a_{t+1|T} - a_{t+1|t}), \\ P_{t|T} &= P_{t|t} + J_t(P_{t+1|T} - P_{t+1|t})J_t^\top. \end{aligned}$$

его являются сглаженные оценки вектора состояния $a_{t|T}$. Ковариационные матрицы ошибок таких оценок $P_{t|T}$ можно получить также рекуррентно во время второго прохода. Алгоритм показан в Таблице 2, вывод приведен в Приложении А. Заметим, что инициализация данного рекуррентного алгоритма не представляет сложности, поскольку величины $a_{t|T}$ и $P_{t|T}$ для $t = T$ получаются на этапе фильтрации.

3.3.2 Пример использования сглаживания: удаление тренда с помощью фильтра Ходрика—Прескотта

Одним из самых широко используемых методов выделения тренда и сглаживания для макроэкономических временных рядов является так называемый *фильтр Ходрика—Прескотта* (Hodrick & Prescott (1997)).

Оценка тренда $\hat{\mu}_t$ для ряда y_t получается в результате решения следующей оптимизационной задачи:

$$\sum_{t=1}^T (y_t - \mu_t)^2 + \lambda \sum_{t=2}^{T-1} [(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1})]^2 \rightarrow \min_{\{\mu_t\}} \quad (24)$$

Первое слагаемое отвечает за точность подгонки, а второе — за степень гладкости тренда. Для обычного детерминированного линейного тренда второе слагаемое будет нулевым. Параметр $\lambda > 0$ регулирует гладкость оценки тренда.⁵ При малых значениях λ оценка $\hat{\mu}_t$ практически совпадет с наблюдаемым рядом y_t , а при очень больших приближается к обычному линейному тренду, подобранному методом наименьших квадратов по точкам (t, y_t) . Вычитая из фактического ряда y_t оценку тренда $\hat{\mu}_t$ получим так называемый output gap, часто используемый в прикладном макроэкономическом моделировании.

Известно (см. напр. Harvey & Jaeger (1993)), что фильтр Ходрика—Прескотта можно вывести из модели локального линейного тренда (9), если принять σ_ξ^2 за ноль:

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, \\ \mu_t &= \mu_{t-1} + \nu_t + 0, \\ \nu_t &= \nu_{t-1} + \zeta_t. \end{aligned} \quad (25)$$

Параметр λ фильтра Ходрика—Прескотта соответствует отношению дисперсий $\sigma_\epsilon^2/\sigma_\zeta^2$. Можно увидеть аналогию целевой функции задачи (24) со взвешенной суммой квадратов

$$\sum_{t=1}^T \epsilon_t^2/\sigma_\epsilon^2 + \sum_{t=3}^T \zeta_t^2/\sigma_\zeta^2.$$

⁵Благодаря исходной статье Hodrick & Prescott (1997) для случая квартальных данных (нездоровую) популярность в прикладных исследованиях приобрело значение $\lambda = 1600$.

Оценка тренда $\hat{\mu}_t$ получается как результат применения калмановского сглаживателя к модели (25). Важно только отношение дисперсий, поэтому при расчетах можно просто положить $\sigma_\epsilon^2 = \lambda$ и $\sigma_\zeta^2 = 1$.⁶

Близкий по смыслу подход используется для вывода формул сглаживания с помощью кубических сплайнов, только модель исходно формулируется в непрерывном времени (см. Wecker & Ansley (1983)). При этом не обязательно, чтобы данные наблюдались через равные интервалы времени.

3.3.3 Пример использования сглаживания: корректировка на сезонность

Одно из возможных применений сглаженных оценок — это корректировка на сезонность. Для проведения такой корректировки надо представить временной ряд в виде «структурной модели», включающей аддитивную сезонную компоненту, получить оценку этой компоненты при помощи сглаживания, а затем вычесть оценку сезонной компоненты из корректируемого ряда.

Рассмотрим подобную корректировку на примере помесечных данных о производстве электроэнергии в России за период 1991:01—2011:01⁷ (Рис. 2).

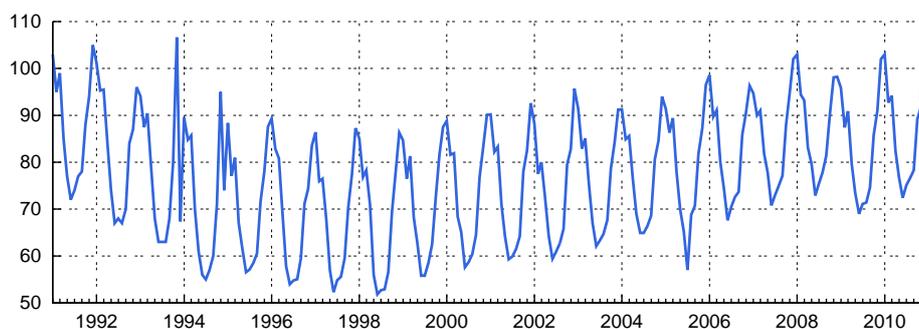


Рис. 2: Потребление электроэнергии в России, млрд кВт·ч.

Предположим, что потребление электроэнергии можно описать структурной моделью, включающей локальный уровень ($\mu_t = \mu_{t-1} + \xi_t$) и тригонометрическую сезонность (12) при $s = 12$:

$$y_t = \mu_t + \sum_{j=1}^6 \gamma_{tj} + \epsilon_t.$$

Сначала для этой модели следует получить оценки максимального правдоподобия параметров (об этом см. ниже). Затем в результате сглаживания надо вычислить значения $\gamma_{tj|T}$. Скорректированное на сезонность потребление электроэнергии будет равно

$$y_t - \sum_{j=1}^6 \gamma_{tj|T}.$$

Оценка на Рис. 3 была вычислена⁸ при следующих значениях параметров: $\sigma_\epsilon = 1$, $\sigma_\xi = 0,5$, $\sigma_\omega = 0,05$. Для начального состояния использовалось расплывчатое распределение с нуле-

⁶Заметим попутно, что можно получить односторонний аналог фильтра Ходрика—Прескотта, если применить к указанной модели калмановскую фильтрацию вместо калмановского сглаживания. Это позволяет использовать соответствующий показатель output gap для прогнозирования, поскольку не требуется «подсматривать в будущее».

⁷Данные с <http://e3.prime-tass.ru/macro/index.aspx>.

⁸Вычисления производились на языке Ox (см. Doornik (2009)). Бесплатную версию Ox Console можно загрузить со страницы <http://www.doornik.com/download.html>. Программы доступны по запросу у автора.

вым средним и дисперсией 10^{10} для локального тренда и 10^6 для подкомпонент сезонной компоненты.

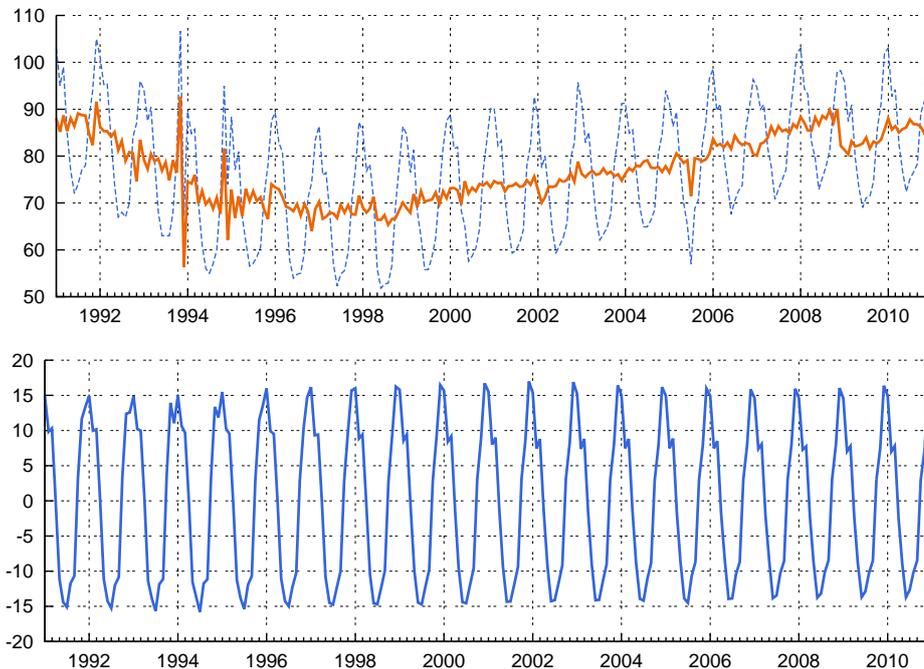


Рис. 3: Потребление электроэнергии в России, скорректированное на сезонность, и соответствующая сглаженная сезонная составляющая.

4 Прогнозирование и пропущенные наблюдения

4.1 Пропущенные наблюдения

Использование фильтра Калмана позволяет без особых проблем учитывать пропущенные наблюдения. Это связано с тем, что в формулировке модели пространства состояний, а также в формулах для фильтрации и сглаживания матрицы могут иметь переменную размерность. Если данные для момента t отсутствуют, то можно действовать, как если бы размерность наблюдаемой переменной y_t была равной нулю ($k_t = 0$). При этом получим $a_{t|t} = a_{t|t-1}$ и $P_{t|t} = P_{t|t-1}$.

Оценку для пропущенного значения можно получить на основе сглаженной оценки состояния:

$$y_{t|T} = \mathbb{E}_T(R_t^y + R_t^{y^a} a_t + \epsilon_t^y) = R_t^y + R_t^{y^a} a_{t|T}.$$

Ковариационная матрица ошибки такой оценки равна⁹

$$\mathbb{V}(y_t - y_{t|T}) = \mathbb{V}[R_t^{y^a}(a_t - a_{t|T}) + \epsilon_t^y] = R_t^{y^a} P_{t|T} (R_t^{y^a})^\top + \Omega_t^y.$$

В использованном выше ряду потребления электроэнергии ноябрю 2008 г. соответствует неправдоподобное значение 8.1. В предыдущих расчетах оно было заменено на 98.1. Точечная оценка для этого наблюдения равна 92.509, а интервальная — [85.927; 99.090] (точечная плюс/минус две стандартные ошибки). Для получения этой оценки параметры модели были оценены методом максимального правдоподобия с учетом пропуска.

⁹Конечно, в этих формулах уже не предполагается, что $k_t = 0$. В то же время $y_{\leq T}$ не включает y_t , так что ϵ_t^y не коррелирует с $a_{t|T}$.

4.2 Прогнозирование

Фактически прогнозирование представляет собой использование фильтра Калмана для «пропущенных» значений y_{T+1}, y_{T+2} и т. д. Новая информация при этом не используется ($a_{t|t} = a_{t|t-1}$ и $P_{t|t} = P_{t|t-1}$ при $t > T$), а оценки будущих значений наблюдаемой переменной y_t получаются как одношаговые прогнозы по обычным формулам фильтра Калмана, т. е. $y_{t|T} = y_{t|t-1}$ при $t > T$. Ковариационные матрицы ошибок прогноза — это матрицы F_t из фильтра Калмана; они могут использоваться, в частности, для расчета интервальных прогнозов.

На Рис. 4 приведен прогноз потребления электроэнергии на период 2010:01–2011:01. Изображен точечный прогноз, интервальный (точечный плюс/минус две стандартные ошибки) и фактические значения ряда. Для получения прогнозов параметры модели были оценены методом максимального правдоподобия по укороченной выборке 1991:01–2010:01.

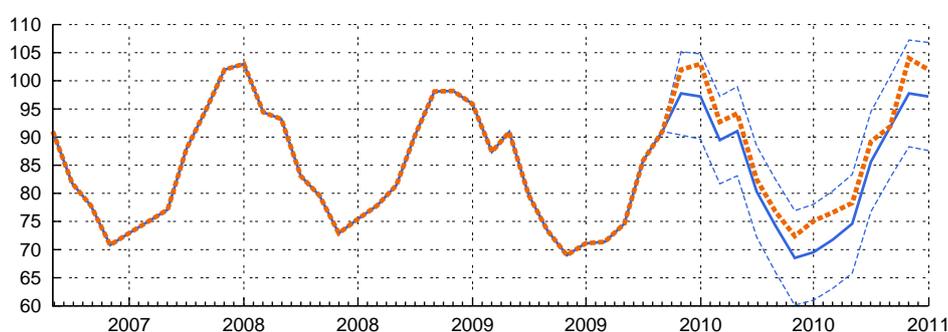


Рис. 4: Прогноз потребления электроэнергии в России.

5 Оценки параметров и диагностика

5.1 Оценки параметров

В предшествующем рассмотрении мы исходили из того, что матрицы коэффициентов модели пространства состояний нам известны. Но на практике они неизвестны и требуется прежде получить некоторые их оценки. Предполагается, что все матрицы коэффициентов и ковариационные матрицы в модели (3)–(4) могут зависеть от вектора неизвестных параметров θ : $R_t^y = R_t^y(\theta)$, $R_t^{ya} = R_t^{ya}(\theta)$, $\Omega_t^y = \Omega_t^y(\theta)$, $R_t^a = R_t^a(\theta)$, $R_t^{aa} = R_t^{aa}(\theta)$, $\Omega_t^a = \Omega_t^a(\theta)$. Таким образом, задача состоит в том, чтобы подобрать вектор θ . Например, в модели ARMA(2,1) (7) требуется найти оценку вектора $\theta = (\phi_1, \phi_2, \theta_1, \sigma_\epsilon^2)$.

Самый распространенный метод оценивания параметров гауссовской линейной модели пространства состояний — это метод максимального правдоподобия. Функция правдоподобия может быть вычислена на основе инноваций v_t и их ковариационных матриц F_t , которые являются «продуктами работы» фильтра Калмана. Функцию правдоподобия (которая по определению является плотностью распределения вектора $y = (y_1, \dots, y_T)$) можно разложить на произведение вкладов отдельных наблюдений

$$f(y) = \prod_{t=1}^T f(y_t | y_{<t}).$$

В случае гауссовской линейной модели пространства состояний предполагается, что ошибки имеют совместное нормальное распределение, так что вклады в функцию правдоподобия представляют собой условные нормальные распределения:

$$y_t | y_{<t} \sim \mathcal{N}(y_{t|t-1}, F_t).$$

Отсюда по формуле плотности многомерного нормального распределения

$$\ln f(y_t | y_{<t}) = -\frac{k_t}{2} \ln(2\pi) - \frac{1}{2} \ln |F_t| - \frac{1}{2} (y_t - y_{t|t-1})^\top F_t^{-1} (y_t - y_{t|t-1}).$$

Можно заметить, что $y_t - y_{t|t-1}$ — это инновации (ошибки одношаговых прогнозов v_t). Таким образом, вклад отдельного наблюдения в логарифмическую функцию правдоподобия есть

$$\ell_t(\theta) = -\frac{k_t}{2} \ln(2\pi) - \frac{1}{2} \ln |F_t(\theta)| - \frac{1}{2} v_t(\theta)^\top F_t^{-1}(\theta) v_t(\theta), \quad (26)$$

где в явном виде отражена зависимость от параметров. В целом логарифмическая функция правдоподобия — это сумма таких вкладов:

$$\ell(\theta) = \sum_{t=1}^T \ell_t(\theta).$$

Тогда

$$\hat{\theta} \in \arg \max_{\theta} \ell(\theta)$$

является оценкой максимального правдоподобия для гауссовской линейной модели пространства состояний.

В общем случае для максимизации $\ell(\theta)$ используют численные методы. Для градиентных методов оптимизации требуются производные целевой функции по параметрам. Здесь можно использовать численные производные или же требующие более сложных выкладок аналитические производные.

Опишем идею получения аналитических производных. Чтобы вычислить $\partial \ell_t(\theta) / \partial \theta_j$, где θ_j — j -й элемент вектора параметров, следует применить стандартные правила матричного дифференцирования к (26). Кроме того, требуется знать $\partial F_t(\theta) / \partial \theta_j$ и $\partial v_t(\theta) / \partial \theta_j$. Эти производные можно вычислить рекуррентно, применяя правила матричного дифференцирования к уравнениям фильтра Калмана (Таблица 1) — дифференцируя каждое уравнение по θ_j . Аналогично можно получить и вторые производные, только выкладки будут еще более громоздкими.

Несколько упростить вычисление аналитических производных позволяет следующий прием. Пусть $f(y, a | \theta)$ — плотность распределения полных данных, где полные данные включают не только наблюдаемые данные $y = (y_1, \dots, y_T)$, но значение состояния $a = (a_1, \dots, a_T)$. Введем функцию

$$Q(\theta; \theta') = \mathbb{E}[\ln f(y, a | \theta) | y, \theta'].$$

Ожидание здесь берется по распределению $a | y, \theta'$, соответствующему сглаживанию с параметрами θ' . Можно показать, что

$$\frac{\partial \ell(\theta)}{\partial \theta} = \left. \frac{\partial Q(\theta; \theta')}{\partial \theta} \right|_{\theta'=\theta}.$$

Дифференцировать $Q(\theta; \theta')$ обычно проще, чем функцию правдоподобия $\ell(\theta)$, выраженную через $v_t(\theta)$ и $F_t(\theta)$. Для многих классов моделей можно в явном виде решить уравнение $\partial Q(\theta; \theta') / \partial \theta = 0$ относительно θ , что дает рекуррентный алгоритм получения оценок максимального правдоподобия. При текущем приближении θ' производится сглаживание. Затем по найденным сглаженным величинам находится следующее приближение θ'' как решение уравнений $\partial Q(\theta; \theta') / \partial \theta = 0$. Процесс повторяется до схождения. Такой алгоритм известен как EM (см. Shumway & Stoffer (1982), Watson & Engle (1983)).

Существуют рекуррентные формулы для расчета не только производных, но и информационной матрицы (см. напр. Cavanaugh & Shumway (1996)). Они позволяют использовать для максимизации функции правдоподобия метод скоринга.

Проверка гипотез в модели пространства состояний может быть основана на асимптотической нормальности и стандартном принципе Вальда. Асимптотические оценки ковариационной матрицы находятся также стандартными для метода максимального правдоподобия способами. Например, они могут быть основаны на матрице вторых производных логарифмической функции правдоподобия или на информационной матрице.

Нужно учитывать, что принцип Вальда «не работает» на краях пространства допустимых параметров. В частности, такое приближение нельзя использовать для проверки гипотезы о равенстве дисперсии ошибки нулю. Также в некоторых случаях одна или несколько переменных состояния могут подчиняться процессам авторегрессии первого порядка с почти единичными характеристическими корнями; тогда нормальное асимптотическое приближение тоже мало пригодно.

5.2 Диагностика модели

Для диагностики удобно использовать нормированные инновации, полученные на основе фильтра Калмана. Они вычисляются как

$$z_t = L_t^{-1}v_t,$$

где L_t — какой-либо квадратный корень¹⁰ из F_t . В случае гауссовской модели эти величины независимы между собой, одинаково нормально распределены, имеют нулевые математические ожидания и единичные ковариационные матрицы. Все эти свойства можно проверять с помощью соответствующих статистик. Например, в случае одномерного y_t статистика $\sum_{t=1}^T z_t/\sqrt{T}$ имеет стандартное нормальное распределение. Значимое отличие этой статистики от нуля позволяет говорить о неадекватности модели. Таким же образом можно выявлять толстые хвосты, автокорреляцию, гетероскедастичность и другие нарушения. Следует однако помнить, что распределения статистик, вообще говоря, зависят от того факта, что параметры модели неизвестны и вместо них использованы оценки.¹¹

При анализе нормированных инноваций не всегда можно определить источник нарушений. Например, распределение с толстыми хвостами может проистекать из толщины хвостов как ошибок ϵ_t^y , так и ошибок ϵ_t^a . Частично эту проблему можно решить, анализируя соответствующие сглаженные остатки:

$$\begin{aligned}\epsilon_{t|T}^y &= y_t - R_t^y - R_t^{ya} a_{t|T}, \\ \epsilon_{t|T}^a &= a_{t|T} - R_t^a - R_t^{aa} a_{t-1|T}.\end{aligned}$$

Можно, в частности, с помощью нормированных остатков диагностировать наличие выбросов и структурных сдвигов. Следует только помнить, что сглаженные остатки могут быть заметно автокоррелированными.

Более подробное обсуждение диагностики см. в Harvey & Koopman (1992) и Durbin & Koopman (2001). См. также de Jong & Penzer (1998), Penzer (2006).

¹⁰То есть такая матрица, что $F_t = L_t L_t^T$.

¹¹Как пример можно привести Q-статистику для автокорреляции в случае, когда параметры модели ARMA оцениваются. Она распределена асимптотически как хи-квадрат, но с меньшим числом степеней свободы, чем если бы использовались известные истинные параметры.

6 Нелинейные и негауссовские модели пространства состояний

6.1 Для чего нужны нелинейные и негауссовские модели пространства состояний?

Область применения гауссовских линейных моделей пространства состояний широка, но для многих приложений предположения линейности и нормальности являются слишком ограниченными.

В частности, ошибки в модели пространства состояний могут иметь толстые хвосты, как у распределения Стьюдента. Стандартные алгоритмы, такие как фильтр Калмана могут быть не робастными по отношению к отдельным большим выбросам или скачкам. Более того, существуют модели пространства состояний с переменными, которые не похожи на нормально распределенные по самой своей природе, например, с ограниченными, счетными или даже качественными зависимыми переменными.

Ранее в качестве линейной модели пространства состояний мы рассмотрели модель линейной регрессии с меняющимися коэффициентами. Но подобное динамическое расширение можно применить и к любой другой статической модели. Так в динамической модели Нельсона—Сигеля можно предположить, что параметр τ и дисперсия ошибки не являются постоянными, а это сразу делает модель существенно нелинейной. Да и в той же линейной регрессии можно было бы предположить, что не только коэффициенты при переменных, но и дисперсия ошибки меняется во времени.

Одной из простейших нелинейных негауссовских моделей является модель стохастической волатильности (см. напр. обзор Цыплаков (2010)). Стохастическая волатильность призвана уловить меняющуюся дисперсию и может быть использована как стандартная компонента в самых разных динамических моделях. Так Stock & Watson (2007) использовали модель локального линейного тренда, дополненную стохастической волатильностью в ошибках, для прогнозирования инфляции.

Приведем еще один, более сложный, пример использования нелинейной негауссовской модели. В Creal, Koopman & Zivot (2010) рассматривается метод построения индикатора экономического цикла с помощью многомерной модели пространства состояний, основанной на разложении на тренд и цикл. В модели с помощью стохастической волатильности учитывается изменение дисперсии во времени. Распределение случайных возмущений моделируется с помощью смеси нормальных распределений. Такой подход обеспечивает робастность получаемых оценок цикла к изменению дисперсии, выбросам и структурным сдвигам в тренде.

Нелинейные и негауссовские модели пространства состояний — это очень обширная тема, причем данная область исследований быстро развивается и все больший набор различных моделей становится доступным для рядовых прикладных исследований. В одном обзоре практически невозможно дать об этом достаточно полное и сбалансированное представление. Мы затронем только некоторые аспекты.

Методы для нелинейных негауссовских моделей пространства состояний включают такие задачи как фильтрация, сглаживание, прогнозирование и оценивание. С точки зрения экономических приложений фильтрация в основном нужна для получения диагностических статистик, а наибольший интерес представляют оценивание и прогнозирование. В технических приложениях больший упор делается на фильтрацию (модификации фильтра Калмана, корпускулярные фильтры). Все указанные задачи в случае моделей общего вида сталкиваются с отсутствием явных аналитических формул и необходимостью численного интегрирования. В зависимости от области приложения используются как грубые аппроксимации, не требующие интегрирования (например, расширенный фильтр Калмана), так и более точные, в основе которых лежит интегрирование методом Монте-Карло. Мы начнем описание различных подходов со случаев, когда отказ от линейности и нормальности распределения не приводит к необходимости полного отказа от стандартных методов, таких как фильтр Калмана,

и закончим методами Монте-Карло. Последний пункт посвящен дискретным переменным в пространстве состояний.

6.2 Условно линейные модели

При выводе соотношений фильтра Калмана все выкладки для момента t были сделаны условно относительно предыстории $y_{<t}$. Поэтому все матрицы коэффициентов могут зависеть не только от вектора неизвестных параметров θ , но и от $y_{<t}$: $R_t^y = R_t^y(\theta, y_{<t})$, $R_t^{ya} = R_t^{ya}(\theta, y_{<t})$, $\Omega_t^y = \Omega_t^y(\theta, y_{<t})$, $R_t^a = R_t^a(\theta, y_{<t})$, $R_t^{aa} = R_t^{aa}(\theta, y_{<t})$, $\Omega_t^a = \Omega_t^a(\theta, y_{<t})$, — от этого рассуждения не потеряют корректности. При этом модель будет только условно линейная в том смысле, что уравнения (3)–(4) для момента t линейны по y_t , a_t , a_{t-1} и ошибкам, если зафиксировать зависимость от $y_{<t}$.

Для гауссовской условно линейной модели функция правдоподобия по прежнему вычисляется по формуле (26), что позволяет проводить оценку модели. Сглаживание можно также осуществлять с помощью калмановского сглаживателя.

Одно из применений условно линейных моделей — это авторегрессии с меняющимися коэффициентами:

$$y_t = \beta_{0t} + \beta_{1t}y_{t-1} + \dots + \beta_{pt}y_{t-p} + \epsilon_t$$

Можно подходить к подобной авторегрессии как к регрессии с меняющимися коэффициентами (см. параграф 2.1), положив $x_t = (1, y_{t-1}, \dots, y_{t-p})^\top$.

Можно через зависимость от $y_{<t}$ моделировать и меняющуюся во времени дисперсию. В Evans (1991) для моделирования динамики темпа инфляции использована авторегрессия с меняющимися коэффициентами, в которой дисперсия ошибки ϵ_t описывалась моделью типа GARCH. Поскольку ошибки ϵ_{t-1} не являются наблюдаемыми, Эванс заменил ϵ_{t-1}^2 на квадрат ошибки предсказания из фильтра Калмана, т. е. v_t^2 . Конечно, такой прием может вызвать нарекания с содержательной стороны. В работе Harvey, Ruiz & Sentana (1992) предложены аппроксимации, призванные помочь совместить модели пространства состояний с «почти настоящими» моделями GARCH.

6.3 Линейная модель с негауссовскими ошибками

Те же формулы фильтра Калмана могут быть использованы в случае негауссовской линейной модели в пространстве состояний, когда ошибки ϵ_t^y , ϵ_t^a не обязательно имеют нормальное распределение. Чтобы это продемонстрировать, нужно только оператор условного математического ожидания \mathbb{E}_t заменить на соответствующий оператор линейной проекции (теоретической линейной регрессии), оператор условной ковариационной матрицы \mathbb{V}_t — на оператор ковариационной матрицы ошибки проекции, а оператор условной кросс-ковариационной матрицы \mathbb{C}_t — на оператор кросс-ковариационной матрицы ошибок проекции двух переменных.

В случае негауссовской модели (при соответствующих предположениях) можно, максимизируя гауссовскую функцию правдоподобия, получать состоятельные и асимптотически нормальные оценки. Эти оценки называют оценками квази максимального правдоподобия (КМП). Вклад t -го наблюдения в логарифм квази функции правдоподобия задается формулой (26), где $F_t(\theta)$ и $v_t(\theta)$ вычисляются по формулам фильтра Калмана. Вклад t -го наблюдения в гауссовский скор-вектор имеет вид

$$\frac{\partial \ell_{Qt}}{\partial \theta_j} = -\frac{1}{2} \operatorname{tr} \left(F_t^{-1} \frac{\partial F_t}{\partial \theta_j} \right) + \frac{1}{2} v_t^\top F_t^{-1} \frac{\partial F_t}{\partial \theta_j} F_t^{-1} v_t + v_t^\top F_t^{-1} \frac{\partial y_{t|t-1}}{\partial \theta_j},$$

где индекс Q показывает, что это не настоящая, а квази логарифмическая функция правдоподобия. При истинных значениях параметров математическое ожидание второго слагаемого

правой части равно первому слагаемому со знаком минус, а математическое ожидание третьего слагаемого равно нулю, поскольку v_t и $\partial y_t|_{t-1}/\partial \theta_j$ некоррелированы. В целом

$$\mathbb{E}\nabla_{\theta}\ell_{Q_t}(\theta) = 0.$$

Отсюда видно, что оценки КМП $\hat{\theta}_Q$, являющиеся решением уравнения

$$\frac{1}{T} \sum_{t=1}^T \nabla_{\theta}\ell_{Q_t}(\hat{\theta}_Q) = 0,$$

являются оценками обобщенного метода моментов. Таким образом, они являются состоятельными и асимптотически нормальными при выполнении ряда предположений.

При соответствующих предположениях асимптотическое распределение $\hat{\theta}_Q$ имеет вид

$$\sqrt{T}(\hat{\theta}_Q - \theta) \xrightarrow{d} \mathcal{N}(0, A^{-1}BA^{-1}),$$

где

$$A = \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}_{\theta}[\nabla_{\theta}^2 \ell_Q(\theta)], \quad B = \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{V}_{\theta}[\nabla_{\theta} \ell_Q(\theta)].$$

Состоятельность и асимптотическое распределение оценок квази-максимального правдоподобия рассмотрены, например, в Watson (1989). Общая теория КМП-оценителей для динамических моделей обсуждается в White (1984). В Papanastassiou (2006) рассматривается метод расчета оценки для ковариационной матрицы оценок КМП.

Хотя обычные методы остаются применимыми в случае негауссовских моделей, интерес представляют и альтернативные методы, в явном виде учитывающие негауссовскую природу модели. В частности, если принять во внимание, что ошибки имеют толстые хвосты, то можно повысить устойчивость моделей пространства состояний к выбросам и скачкам. Для моделей, включающих дискретные переменные, линейное гауссовское моделирование может быть очень сильно неадекватным или вообще неприменимым.

6.4 Приближения для нелинейных и негауссовских моделей

6.4.1 Расширенный фильтр Калмана

Расширенный фильтр Калмана (*extended Kalman filter*, *EKF*, см. напр. Anderson & Moore (1979), Ch. 8, О'Коннелл (1990)) применяется к нелинейной модели пространства состояний следующего вида:

$$y_t = h_t(a_t) + \epsilon_t^y, \tag{27}$$

$$a_t = g_t(a_{t-1}) + \epsilon_t^a. \tag{28}$$

Ошибки ϵ_t^y , ϵ_t^a независимы как между собой, так и по времени, имеют нулевые математические ожидания и ковариационные матрицы Ω_t^y , Ω_t^a . Идея расширенного фильтра Калмана состоит в линеаризации нелинейных функций h_t и g_t в точке, соответствующей текущей оценке вектора состояния: g_t линеаризуется в точке $a_{t-1|t-1}$, а h_t линеаризуется в точке $a_{t|t-1}$. Получаем следующие приближения:

$$y_t = h_t(a_{t|t-1}) + \nabla h_t(a_{t|t-1})(a_t - a_{t|t-1}) + \epsilon_t^y, \tag{29}$$

$$a_t = g_t(a_{t-1|t-1}) + \nabla g_t(a_{t-1|t-1})(a_{t-1} - a_{t-1|t-1}) + \epsilon_t^a. \tag{30}$$

Здесь ∇h_t , ∇g_t — матрицы Якоби для нелинейных вектор-функций h_t и g_t .

Формулы алгоритма фильтра Калмана при этом остаются практически неизменными, просто при расчетах используются следующие матрицы коэффициентов:

$$R_t^y = h_t(a_{t|t-1}) - \nabla h_t(a_{t|t-1})a_{t|t-1}, \quad R_t^{ya} = \nabla h_t(a_{t|t-1}),$$

$$R_t^a = g_t(a_{t-1|t-1}) - \nabla g_t(a_{t-1|t-1})a_{t-1|t-1}, \quad R_t^{aa} = \nabla g_t(a_{t-1|t-1}).$$

К сожалению, для расширенного фильтра Калмана при нелинейности модели в общем случае не выполнены все те «хорошие» свойства, которые характерны для обычного фильтра Калмана. В частности, максимизация «функции правдоподобия», рассчитанной по обычным формулам, не приводит к состоятельным оценкам.

Для технических приложений бывает нужна онлайн-фильтрация, при которой невозможно оценивать параметры при появлении каждого нового наблюдения. В качестве альтернативы можно дополнить вектор состояния a_t вектором (как бы) меняющихся во времени параметров θ_t и применять расширенный фильтр Калмана к этому обобщенному вектору состояния. Это так называемый *сдвоенный (dual) расширенный фильтр Калмана* (Wan & Nelson (2001)), являющийся частным случаем рекуррентных методов оценивания (Льюнг (1991), гл. 11). Такую процедуру можно применять и в случае линейной модели пространства состояний. При этом модель станет нелинейной, поскольку она практически всегда нелинейна по параметрам θ . Конечно, при таком подходе проблема оценивания переходит на другой уровень: требуется выбрать (фиксированные) параметры для модели изменения параметров θ_t .

6.4.2 Нелинейная фильтрация без производных

Получивший в последние годы распространение *непахучий*¹² фильтр Калмана (*unscented Kalman filter*, UKF, см. Julier & Uhlmann (1997), Wan & van der Merwe (2001)) призван заменить расширенный фильтр Калмана. Для его работы не требуется взятие производных, и даже не требуется, чтобы функции h_t и g_t были дифференцируемыми. Он не сложнее расширенного фильтра Калмана и при этом опыт показывает, что он проводит фильтрацию более качественно.

В основе *непахучего* фильтра Калмана лежит *непахучее преобразование (unscented transform)*. Пусть есть нелинейное преобразование $y = f(x)$, где $x \in \mathbb{R}^n$. Предполагается, что x является случайной величиной со средним μ_x и ковариационной матрицей Σ_x . Требуется аппроксимировать среднее μ_y и ковариационную матрицу Σ_y величины y . В окрестности μ_x берется набор из $2n + 1$ определенных сигма-точек $x^{(i)}$ с соответствующими весами $W^{(i)}$, так чтобы они отражали параметры μ_x и Σ_x . Сигма-точки пропускаются через нелинейное преобразование и получается новый набор точек $f(x^{(i)})$, по которым вычисляются соответствующие μ_y и Σ_y .

Похожий метод состоит в аппроксимации нелинейной функции с помощью конечных разностей. Это фильтр Калмана на основе центральных разностей (*central difference Kalman filter*, Nørgaard, Poulsen & Ravn (1998)). В целом данное семейство методов фильтрации без взятия производных называется *калмановской фильтрацией на основе сигма-точек*.

6.4.3 Модели обобщенного авторегрессионного скор-процесса

Сложности, возникающие при работе с нелинейными моделями пространства состояний заставляют искать более «дешевые» альтернативы. В работе Creal, Koopman & Lucas (2010) предложен общий подход, позволяющий модели с меняющимися во времени ненаблюдаемыми факторами превращать в модели особого вида, с которыми работать гораздо проще. Это

¹²Название не несет смысловой нагрузки и просто выбрано как запоминающееся.

не модели пространства состояний, но близкие к ним по смыслу модели. Авторы назвали их моделями обобщенного авторегрессионного скор-процесса (*generalized autoregressive score*, GAS). В этот класс входит широко известная модель GARCH и другие модели, встречающиеся в литературе, в частности, модель авторегрессионной условной длительности Engle & Russell (1998).

6.4.4 Приближение Лапласа

В общем случае для нелинейных и негауссовских моделей нет явных формул для вычисления функции правдоподобия. Сравнительно несложный метод приближенного численного интегрирования, который применим для достаточно широкого класса моделей — это приближение Лапласа (*Laplace's approximation*, см. напр. Davis & Rodriguez-Yam (2005), Shimada & Tsukuda (2005)). В основе лежит разложение до членов второго порядка логарифмической функции правдоподобия полных данных $\ln(f(y, a))$ в точке $\hat{a}(y)$, которая определяется как аргмаксимум $f(y, a)$ по a . (Точка $\hat{a} = \hat{a}(y)$ является также модой условного распределения $f(a|y)$). Пусть $\ln(g(y, a))$ — такое квадратичное приближение. Тогда $g(y, a)$ можно проинтегрировать аналитически по a и получить приближенную функцию правдоподобия

$$f_{LA}(y) = g(y, a)/g(a).$$

Здесь плотность $g(a)$ соответствует многомерному нормальному распределению и получается приравниванием первых и вторых производных $\ln(g(a))$ в точке \hat{a} к производным $\ln(f(y, a))$ по a . Минимизация $f_{LA}(y)$ по параметрам θ дает оценки параметров, которые, впрочем, не являются в общем случае состоятельными. Хорошее приближение получается, когда маргинальное распределение ряда a является многомерным нормальным, например, имеет линейную марковскую структуру (6). Многомерное нормальное распределение, соответствующее $g(a)$, можно использовать для целей (приближенных) фильтрации, сглаживания и прогнозирования.

6.5 Методы Монте-Карло

6.5.1 Последовательные методы Монте-Карло

Последовательные методы Монте-Карло (*sequential Monte Carlo*, см. Doucet, de Freitas & Gordon (2001), Creal (2009)) — это семейство методов, которое используется для фильтрации в нелинейных негауссовских моделях пространства состояний общего вида. Особенно большой популярностью пользуется термин «корпускулярные фильтры» (*particle filters*, см. Gordon, Salmond & Smith (1993), Doucet & Johansen (2009)). Условные распределения при фильтрации аппроксимируются взвешенной выборкой, элементы которой называют корпускулами. Из корпускул для момента времени t генерируют корпускулы для момента $t + 1$ и пересчитывают веса. Характеристики распределения $a_t | y_{\leq t}$ для каждого t можно оценить с помощью соответствующих выборочных моментов.

6.5.2 Методы Монте-Карло на основе приближения Лапласа

Классические и байесовские методы Монте-Карло на основе приближения Лапласа предложены в Shephard & Pitt (1997), Durbin & Koopman (1997), Durbin & Koopman (2000). См. также Durbin & Koopman (2001), Jungbacker & Koopman (2007). В отличие от последовательных методов Монте-Карло в данном подходе все значения переменной состояния $a = (a_1, \dots, a_T)$ генерируются совместно, по распределению $g(a)$, соответствующему приближению Лапласа для условного распределения $f(a|y)$. Сгенерированные выборки можно использовать как основу для метода выборки по значимости или для метода марковских цепей Монте-Карло.

6.6 Включение дискретных переменных в пространство состояний

В скрытых марковских моделях переменная состояния a_t принимает конечное число значений $\{1, \dots, S_t\}$. Это марковская цепь, задаваемая матрицей перехода $\Pi_t = \{\pi_{tij}\}$. Здесь $\pi_{tij} = \Pr(a_t = j | a_{t-1} = i)$ — вероятность перехода из состояния i в состояние j . Подобные модели давно изучаются в инженерной литературе, особенно в литературе по распознаванию речи. См. напр. Rabiner (1989), а также Cappé, Moulines & Rydén (2007), Bhar & Namori (2004).

Существуют и модели, в которых часть компонент переменной состояния дискретна, а часть непрерывна. Наиболее удобно работать с моделями, которые условно относительно дискретных переменных являются линейными гауссовскими. Усложнением являются модели, в которых уравнения для y_t и непрерывной части состояния a_t являются линейными гауссовскими условно относительно дискретных переменных и наблюдаемой предыстории $y_{<t}$.

С помощью дискретных переменных можно моделировать переключения между несколькими состояниями («режимами»), асимметричные распределения и распределения с толстыми хвостами в виде смеси гауссовских распределений и т. п. Включение дискретных переменных в пространство состояний в эконометрической литературе получило распространение благодаря статье Hamilton (1989), где введены модели с марковским переключением режима (*Markov switching models*). В Kim & Nelson (1999) рассмотрены различные виды моделей с переключением режима в контексте моделей пространства состояний. См. также Bhar & Namori (2004).

Список литературы

- О’Коннелл, П. Э. (1990). Калмановская фильтрация. Глава в *Справочник по прикладной статистике* (под редакцией Э. Ллойд & У. Ледерман), М.: Финансы и статистика, том 2, глава 20.
- Льюнг, Л. (1991). *Идентификация систем: Теория для пользователя*. М.: Наука.
- Цыплаков, А. (2010). Сделать тайное явным: искусство моделирования с помощью стохастической волатильности. *Квантиль* 8, 69–122.
- Anderson, B. D. O. & J. B. Moore (1979). *Optimal Filtering*. Dover Publications. (2005 reprint).
- Bhar, R. & S. Namori (2004). *Hidden Markov Models: Applications to Financial Economics*. Kluwer Publishers.
- Cappé, O., E. Moulines & T. Rydén (2007). *Inference in Hidden Markov Models*. Springer, 2-е издание.
- Cavanaugh, J. E. & R. H. Shumway (1996). On computing the expected Fisher information matrix for state-space model parameters. *Statistics & Probability Letters* 26, 347–355.
- Chang, Y., J. I. Miller & J. Y. Park (2009). Extracting a common stochastic trend: Theory with some applications. *Journal of Econometrics* 150, 231–247.
- Christensen, J. H., F. X. Diebold & G. D. Rudebusch (2011). The affine arbitrage-free class of Nelson-Siegel term structure models. *Journal of Econometrics*, forthcoming.
- Commandeur, J. J. F. & S. J. Koopman (2007). *An Introduction to State Space Time Series Analysis*. Oxford University Press.
- Creal, D. (2009). A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews*, forthcoming.
- Creal, D., S. J. Koopman & A. Lucas (2010). Generalized autoregressive score models with applications. Unpublished.
- Creal, D., S. J. Koopman & E. Zivot (2010). Extracting a robust US business cycle using a time-varying multivariate model-based bandpass filter. *Journal of Applied Econometrics* 25, 695–719.
- Davis, R. A. & G. Rodriguez-Yam (2005). Estimation for state-space models based on a likelihood approximation. *Statistica Sinica* 15, 381–406.

- de Jong, P. & J. Penzer (1998). Diagnosing shocks in time series. *Journal of the American Statistical Association* 93, 796–806.
- Diebold, F. X. & C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics* 130, 337–364.
- Diebold, F. X., G. D. Rudebusch & B. Aruoba (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics* 131, 309–338.
- Doornik, J. A. (2009). *Ox 6 — An Object-Oriented Matrix Programming Language*. Timberlake Consultants Ltd.
- Doucet, A., N. de Freitas & N. Gordon, редакторы (2001). *Sequential Monte Carlo Methods in Practice*. Springer.
- Doucet, A. & A. M. Johansen (2009). A tutorial on particle filtering and smoothing: Fifteen years later. Глава в *The Oxford Handbook of Nonlinear Filtering* (под редакцией D. Crisan & B. Rozovsky), Oxford University Press. Forthcoming.
- Durbin, J. (2004). Introduction to state space time series analysis. Глава в *State Space and Unobserved Component Models* (под редакцией A. Harvey, S. J. Koopman & N. Shephard), Cambridge University Press, 3–25.
- Durbin, J. & S. J. Koopman (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space models. *Biometrika* 84, 669–684.
- Durbin, J. & S. J. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society Series B* 62, 3–56.
- Durbin, J. & S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.
- Engle, R. F. & J. R. Russell (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* 66, 1127–1162.
- Evans, M. (1991). Discovering the link between inflation rates and inflation uncertainty. *Journal of Money, Credit and Banking* 23, 169–184.
- Gordon, N. J., D. J. Salmond & A. F. M. Smith (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F* 140, 107–113.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.
- Harvey, A. & G. De Rossi (2007). Signal extraction. Глава в *Palgrave Handbook of Econometrics* (под редакцией T. C. Mills & K. Patterson), Palgrave Macmillan, том 1: Econometric Theory, глава 27, 970–1000.
- Harvey, A., E. Ruiz & E. Sentana (1992). Unobserved component time series models with ARCH disturbances. *Journal of Econometrics* 52, 129–157.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- Harvey, A. C. & A. Jaeger (1993). Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics* 8, 231–247.
- Harvey, A. C. & S. J. Koopman (1992). Diagnostic checking of unobserved components time series models. *Journal of Business and Economic Statistics* 10, 377–389.
- Harvey, A. C. & T. Proietti, редакторы (2005). *Readings in Unobserved Components Models*. Advanced Texts in Econometrics. Oxford University Press.
- Harvey, A. C. & N. Shephard (1993). Structural time series models. Глава в *Handbook of Statistics, Volume 11* (под редакцией G. S. Maddala, C. R. Rao & H. D. Vinod), Amsterdam: Elsevier Science Publishers B V.
- Hodrick, R. J. & E. C. Prescott (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking* 29, 1–16.
- Julier, S. J. & J. K. Uhlmann (1997). A new extension of the Kalman filter to nonlinear systems. Глава в *Proc. of AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls*. Orlando, Florida.

- Jungbacker, B. & S. J. Koopman (2007). Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika* 94, 827–839.
- Kim, C. & C. R. Nelson (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. MIT Press.
- Nelson, C. R. & A. F. Siegel (1987). Parsimonious modeling of yield curves. *Journal of Business* 60, 473–489.
- Nørgaard, M., N. K. Poulsen & O. Ravn (1998). Advances in derivative-free state estimation for nonlinear systems. Technical Report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU. Revised edition, Oct 2004. <http://www2.imm.dtu.dk/pubdb/p.php?2706>
- Papanastassiou, D. (2006). Computing the covariance matrix of QML estimators for a state space model. *Statistics & Probability Letters* 76, 1001–1006.
- Penzer, J. (2006). Diagnosing seasonal shifts in time series using state space models. *Statistical Methodology* 3, 193–210.
- Rabiner, L. A. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77, 257–286.
- Shephard, N. & M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84, 653–668.
- Shimada, J. & Y. Tsukuda (2005). Estimation of stochastic volatility models: An approximation to the nonlinear state space representation. *Communications in Statistics — Simulation and Computation* 34, 429–450.
- Shumway, R. H. & D. S. Stoffer (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis* 3, 253–264.
- Stock, J. H. & M. W. Watson (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking* 39, 3–33.
- Wan, E. A. & A. T. Nelson (2001). Dual extended Kalman filter methods. Глава в *Kalman Filtering and Neural Networks* (под редакцией S. Haykin), Wiley, глава 5, 123–173.
- Wan, E. A. & R. van der Merwe (2001). The unscented Kalman filter. Глава в *Kalman Filtering and Neural Networks* (под редакцией S. Haykin), Wiley, глава 7, 221–280.
- Watson, M. W. (1989). Recursive solution methods for dynamic linear rational expectations models. *Journal of Econometrics* 41, 65–89.
- Watson, M. W. & R. Engle (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC, and varying coefficient regression models. *Journal of Econometrics* 23, 385–400.
- Wecker, W. E. & C. F. Ansley (1983). The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association* 78, 81–89.
- White, H. L. (1984). Maximum likelihood estimation of misspecified dynamic models. Глава в *Misspecification Analysis* (под редакцией T. K. Dijkstra), New York: Springer, 1–19.

Приложение А. Вывод калмановского сглаживателя

Для вывода формул калмановского сглаживателя рассмотрим совместное распределение a_t и a_{t+1} , условное относительно $y_{\leq t}$:

$$\begin{pmatrix} a_t \\ a_{t+1} \end{pmatrix} | y_{\leq t} \sim \mathcal{N} \left(\begin{pmatrix} a_{t|t} \\ a_{t+1|t} \end{pmatrix}, \begin{pmatrix} P_{t|t} & C_t \\ C_t^\top & P_{t+1|t} \end{pmatrix} \right),$$

где

$$C_t = \mathbb{C}_t(a_t, a_{t+1}) = \mathbb{C}_t(a_t, R_{t+1}^{aa} a_t) + \mathbb{C}_t(a_t, \epsilon_{t+1}^a) = P_{t|t} (R_{t+1}^{aa})^\top.$$

По свойствам многомерного нормального распределения

$$\mathbb{E}(a_t | a_{t+1}, y_{\leq t}) = a_{t|t} + J_t(a_{t+1} - a_{t+1|t})$$

и

$$\mathbb{V}(a_t | a_{t+1}, y_{\leq t}) = P_{t|t} - J_t C_t^\top = P_{t|t} - J_t P_{t+1|t} J_t^\top,$$

где

$$J_t = C_t P_{t+1|t}^{-1} = P_{t|t} (R_{t+1}^{aa})^\top P_{t+1|t}^{-1}.$$

Введем обозначение $z_{>t}$ для отрезка ряда $\{z_t\}$, начинающегося с момента t : (z_{t+1}, \dots, z_T) .

Заметим, что $\epsilon_{>t+1}^a, \epsilon_{>t}^y$ независимы от $a_t, a_{t+1}, y_{\leq t}$, и что $a_{t+1}, y_{\leq t}, \epsilon_{>t+1}^a, \epsilon_{>t}^y$ взаимно однозначно связаны с $a_{>t}, y_{\leq T}$. Отсюда следует¹³, что

$$\mathbb{E}(a_t | a_{t+1}, y_{\leq t}) = \mathbb{E}(a_t | a_{t+1}, y_{\leq t}, \epsilon_{>t+1}^a, \epsilon_{>t}^y) = \mathbb{E}(a_t | a_{>t}, y_{\leq T}),$$

и поэтому

$$\mathbb{E}[\mathbb{E}(a_t | a_{t+1}, y_{\leq t}) | y_{\leq T}] = \mathbb{E}[\mathbb{E}(a_t | a_{>t}, y_{\leq T}) | y_{\leq T}] = \mathbb{E}(a_t | y_{\leq T}) = a_{t|T}.$$

Таким образом,

$$a_{t|T} = \mathbb{E}[a_{t|t} + J_t(a_{t+1} - a_{t+1|t}) | y_{\leq T}] = a_{t|t} + J_t(a_{t+1|T} - a_{t+1|t}).$$

Вычислим теперь ковариационную матрицу. Воспользуемся тем, что

$$\mathbb{V}(a_t | a_{t+1}, y_{\leq t}) = \mathbb{V}(a_t | a_{>t}, y_{\leq T}).$$

Отсюда с использованием формулы полной дисперсии¹⁴ получим

$$\begin{aligned} P_{t|T} &= \mathbb{V}_T(a_t) = \mathbb{V}(a_t | y_{\leq T}) = \mathbb{E}_T[\mathbb{V}(a_t | a_{>t}, y_{\leq T})] + \mathbb{V}_T[\mathbb{E}(a_t | a_{>t}, y_{\leq T})] \\ &= \mathbb{E}_T[\mathbb{V}(a_t | a_{t+1}, y_{\leq t})] + \mathbb{V}_T[\mathbb{E}(a_t | a_{t+1}, y_{\leq t})] \\ &= \mathbb{E}_T[P_{t|t} - J_t P_{t+1|t} J_t^\top] + \mathbb{V}_T[a_{t|t} + J_t(a_{t+1} - a_{t+1|t})] \\ &= P_{t|t} - J_t P_{t+1|t} J_t^\top + J_t \mathbb{V}_T[a_{t+1}] J_t^\top = P_{t|t} - J_t P_{t+1|t} J_t^\top + J_t P_{t+1|T} J_t^\top \end{aligned}$$

или

$$P_{t|T} = P_{t|t} + J_t(P_{t+1|T} - P_{t+1|t})J_t^\top.$$

An introduction to state space modeling

Alexander Tsyplakov

Novosibirsk State University, Novosibirsk, Russia

Many time series models, primarily various models with unobservable components, can be represented in a so called state space form. A state space model is a powerful tool that allows one to apply to the original model a wide range of standard procedures including estimation and forecasting. This essay provides a survey of this universal class of models and related procedures.

¹³Если величина r независима от p и q , то $\mathbb{E}(p|q, r) = \mathbb{E}(p|q)$.

¹⁴Здесь используется условный вариант этой формулы: $\mathbb{V}(p|q) = \mathbb{E}[\mathbb{V}(p|q, r)|q] + \mathbb{V}[\mathbb{E}(p|q, r)|q]$.

Тестирование распределений*

Игорь Хейфец[†]

Российская экономическая школа, Москва, Россия

В эссе рассмотрены методы тестирования предположений о форме функций распределений при помощи теории эмпирических процессов. На примерах показаны трудности при применении таких методов — эффект от оценивания параметра, зависимость асимптотики от распределения — и способы их преодоления — мартингальное преобразование, бутстрап. Тесты применимы, в частности, к GARCH- и диффузионным моделям.

1 Введение

Методы оценивания и прогнозирования эконометрических моделей основываются на некоторых предположениях о моделях. Если эти предположения неверны, мы не можем быть уверены в полученных результатах. Оценки могут уже не быть эффективными или даже стать несостоятельными. Например, оценка методом максимального правдоподобия (ММП) при неверном предположении, что распределение случайных величин (СВ) является нормальным, перестает быть эффективной, а при неверном предположении, что распределение является t_5 (т.е. распределением Стьюдента с 5 степенями свободы), перестает быть даже состоятельной. Другие примеры последствий плохой спецификации см. в Крил (2008). Поэтому важно научиться тестировать предпосылки, на которые опираются модели.

В данном эссе мы рассмотрим некоторые методы тестирования предположений о функциональной форме функций распределений (ФР), называемыми для краткости критериями согласия (англ. goodness-of-fit). Проверку другого типа предпосылок, таких как о функциональной форме среднего и дисперсии, форме связей между переменными (например, линейности), симметричности плотностей, независимости или некоррелированности ошибок с регрессорами и т.п., мы напрямую здесь рассматривать не будем. На практике знать функциональную форму ФР важно не только для применения различных методов оценки и прогнозирования динамики курсов акций, индексов, обменных курсов, процентных ставок, но также и при непосредственном моделировании распределения, например, распределения неравенства доходов среди населения или распределения доходностей портфелей.

2 Безусловные критерии согласия

Пусть у нас имеется выборка независимых СВ X_1, X_2, \dots, X_n длины n . Мы хотим протестировать простую гипотезу, что X_i имеют распределение $F(x)$:

$$H_0 : X_i \sim F \quad \forall i = 1, \dots, n, \tag{1}$$

против любой альтернативы. Определим *эмпирическую* (или выборочную) ФР X_i как

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i \leq x\},$$

где $\mathbb{I}\{\cdot\}$ — индикатор-функция. Если гипотеза верна, то $F_n(x)$ должна быть близка к предполагаемой $F(x)$.

*Цитировать как: Хейфец, Игорь (2011) «Тестирование распределений», Квантиль, №9, стр. 25–34. Citation: Kheifets, Igor (2011) “Goodness-of-fit testing,” Quantile, No.9, pp. 25–34.

[†]Адрес: 117418, г. Москва, Нахимовский проспект, 47, офис 2114. Электронная почта: ikheifets@nes.ru

Можно определить «близость» двух функций как

$$D_n = \sup_x |F_n(x) - F(x)|, \quad (2)$$

т. е. взять sup-метрику (не вдаваясь в подробности, как определить пространства функций, чтобы это была метрика в строгом смысле). По теореме Гливленко–Кантелли $D_n \xrightarrow{п.н.} 0$ (почти наверное) при $n \rightarrow \infty$. Далее,

$$\sqrt{n}D_n \xrightarrow{d} \sup_x |B(F(x))|$$

при $n \rightarrow \infty$, где $B(\cdot)$ — Броуновский мост. Если $F(x)$ непрерывна, то предельное распределение $\sqrt{n}D_n$ не зависит от $F(x)$, совпадает с распределением $\sup_{x \in [0,1]} |B(x)|$ и равно

$$K(x) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 x^2} = \frac{\sqrt{2\pi}}{x} \sum_{i=1}^{\infty} e^{-(2i-1)^2 \pi^2 / (8x^2)} \quad (3)$$

(Kolmogoroff, 1933), что называется *распределением Колмогорова*. *Свобода от распределения* (англ. distribution free) — важное свойство статистики, позволяющее получить критические значения для любого n вне зависимости от $F(x)$. Этот тест называют *критерием Колмогорова* или *критерием Колмогорова–Смирнова*. Если взять ω^2 -метрику

$$\omega_n^2 = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x),$$

то получится *тест Крамера–фон Мизеса* (также известного как *критерий ω^2* или *критерий Смирнова*). В зависимости от метрики свойства критериев могут меняться. Мощность будет зависеть также от типа альтернативы, и сказать, какой критерий оптимален в общем случае, нельзя. Однако, такие критерии *состоятельны против всех альтернатив* (от лат. *omnibus* — «для всех»), т. е. любое отклонение от истинного распределения будет замечено при достаточно большой выборке, в противоположность критериям, которые хорошо работают против одних альтернатив и не работают против остальных.

Заметим, что если гипотеза (1) верна, и $F(x)$ непрерывна, то СВ

$$U_i = F(X_i) \sim U[0, 1] \quad (4)$$

равномерно распределены на $[0, 1]$. На этом факте основан альтернативный способ вывода статистики Колмогорова. Рассмотрим *эмпирический процесс*

$$V_n(u) = \sqrt{n} [F_n^U(u) - u], \quad u \in [0, 1], \quad (5)$$

где $F_n^U(u)$ — эмпирическая ФР U_i . Этот процесс сравнивает эмпирическую и предполагаемую ФР для U_i . При нулевой гипотезе эта разность должна быть маленькой для всех $u \in [0, 1]$. Если взять sup-метрику, то мы получим статистику $\sqrt{n}D_n$ для равномерных распределений.

Для многомерных случайных величин ситуация более сложная. Пусть у нас имеется выборка независимых случайных векторов размерности K

$$X_1, X_2, \dots, X_n,$$

где $X_i = (X_{i1}, X_{i2}, \dots, X_{iK})$, $i = 1, \dots, n$, и мы хотим протестировать гипотезу (1), что X_i имеют совместное распределение $F(x)$, $x = (x_1, x_2, \dots, x_K) \in R^K$. Пусть $F_n(x)$ обозначает совместную эмпирическую ФР

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \prod_{k=1}^K \mathbb{I}\{X_{ik} \leq x_k\}.$$

Если гипотеза верна, то непараметрическая оценка ФР $F_n(x)$ должна быть близка к $F(x)$. Для $K > 1$ распределение sup- и ω^2 -статистик зависит от $F(x)$, то есть нарушена свобода от распределения.

Rosenblatt (1952) предложил следующее преобразование (*преобразование Розенблатта*):

$$U_{ik} = \Pr \{X_{ik} | X_{ik-1}, \dots, X_{i1}\} \quad \forall i = 1, \dots, n, k = 1, \dots, K. \quad (6)$$

Если $X_i \sim F$, то $U_i = (U_{i1}, U_{i2}, \dots, U_{iK}) \sim U[0, 1]^K$ равномерно распределены на единичном кубе размерности K . Этим можно воспользоваться для построения теста. Иногда условные вероятности в (6) известны в явном виде. Например, их легко получить для многомерного нормального распределения.

На практике вместо проверки конкретной ФР необходимо научиться проверять принадлежность к параметрическому семейству ФР $\{F(x, \theta), \theta \in \Theta\}$. Например, ММП дает состоятельные и асимптотически эффективные оценки для всего семейства нормальных данных. Рассмотрим *сложную гипотезу*

$$H_0 : \exists \theta_0 \quad X_i \sim F(x, \theta_0) \quad \forall i = 1, \dots, n. \quad (7)$$

Если гипотеза (7) верна, то непараметрическая оценка $F_n(x)$ должна быть близка к параметрической оценке $F(x, \hat{\theta})$, где $\hat{\theta}$ — \sqrt{n} -состоятельная оценка истинного параметра θ_0 .

Вернемся к одномерному случаю. Так как $F(x, \hat{\theta})$ не совпадает с $F(x, \theta_0)$, $\hat{U}_i = F(X_i, \hat{\theta})$ не распределены равномерно, в отличие от U_i в (4). Эмпирический процесс

$$\hat{V}_n(u) = \sqrt{n} \left[F_n^{\hat{U}_i}(u) - u \right], \quad u \in [0, 1], \quad (8)$$

и статистика $\hat{D}_n = \sup_{u \in [0, 1]} |\hat{V}_n(u)|$ имеет при каждом новом значении параметра распределение, отличное от D_n , т.е. мы наблюдаем *эффект от оценивания параметров*, и нельзя использовать критические значения (3) (Durbin, 1973). Используя теорию слабой сходимости эмпирических процессов (см. Billingsley, 1968), обычно можно вывести, как изменится процесс, построенный на оценках на месте истинных параметров. Например, для (8) можно показать, что равномерно по $u \in [0, 1]$

$$\hat{V}_n(u) = V_n(u) - g(u)\sqrt{n}(\hat{\theta} - \theta_0) + o_p(1), \quad (9)$$

где $g(u)$ — некоторая функция, зависящая от производной $F(x, \theta)$ по параметру (Bai, 2003). Обычно используют один из двух способов получения критических значений. Первый способ заключается в преобразовании $\hat{V}_n(u)$, убирающем дополнительный член в разложении (9). См. например, мартингальное преобразование, описанное в Khmaladze (1981) и примененное в Bai (2003), Delgado & Stute (2008) и др. Распределение преобразованного процесса и статистики можно затабулировать, но сами преобразования зависят от исходных моделей. Второй способ заключается в бутстрапировании \hat{D}_n , см. Andrews (1997) и Corradi & Swanson (2005). Далее мы рассмотрим подробнее оба этих способа.

3 Тестирование условных распределений

3.1 Обобщенный критерий Колмогорова и параметрический бутстрап

Andrews (1997) обобщил критерий Колмогорова для тестирования условных распределений. Пусть дана выборка независимых одинаково распределенных (НОР) случайных векторов $Z_i = (Y_i', X_i')' \in R^{V+K}$. Рассмотрим гипотезу

$$H_0 : \exists \theta_0 \quad Y_i | X_i \sim F(y|x, \theta_0) \quad \forall i = 1, \dots, n. \quad (10)$$

Пусть $H_n(z)$ и $G_n(x)$ — эмпирические ФР Z_i и X_i соответственно. Рассмотрим полупараметрическую оценку совместной ФР Z_i , основанную на гипотетической условной $F(y|x, \theta_0)$ и эмпирической маржинальной $G_n(x)$:

$$F_n(z, \theta) = \frac{1}{n} \sum_{i=1}^n F(y|X_i, \theta) \mathbb{I}\{X_i \leq x\}, \quad z = (y, x)' \in R^{V+K}.$$

Сравнивая полупараметрическую (опирающуюся на \sqrt{n} -состоятельную оценку $\hat{\theta}$) и эмпирическую ФР, получаем статистику

$$CK_n = \sqrt{n} \max_{i \leq n} |H_n(Z_i) - F_n(Z_i, \hat{\theta})|.$$

В отличие от статистики Колмогорова (2), здесь достаточно брать максимум по конечному числу элементов выборки, что значительно сокращает время вычислений. Распределение статистики подвержено эффекту от оценивания параметров. Критические значения можно получить с помощью параметрического бутстрапа.

На примере статистики CK_n опишем подробнее алгоритм получения критических значений с помощью бутстрапа.¹ Задача бутстрапа состоит в приближении распределения статистики для СВ предполагаемым распределением вне зависимости, совпадает ли оно с истинным или нет. Эмпирическое распределение H_n для этих целей не подходит, так как не генерирует предполагаемое распределение, если гипотеза (10) не верна. Естественным кандидатом для генерирования Y_i является гипотетическое условное распределение $F(y|x, \theta_0)$, или $F(y|x, \hat{\theta})$, если параметр неизвестен. Регрессоры X_i можно сгенерировать из G_n или брать одни и те же исходные данные для каждой симуляции (*бутstrap с фиксированными регрессорами*). Основываясь на симуляциях Монте Карло, Andrews (1997) сделал вывод, что бутstrap с фиксированными регрессорами работает лучше. Распределение СВ, получаемых таким бутстрапом, известно с точностью до конечномерного параметра θ , поэтому процедура называется параметрическим бутстрапом.

В итоге параметрический бутstrap выглядит следующим образом:

1. По данным $Z_i = (Y'_i, X'_i)'$ получаем оценку параметра $\hat{\theta}$ и статистику CK_n .
2. С помощью $F(y|x, \hat{\theta})$, условно на X_i , генерируем Y_{bi} .
3. По $Z_{bi} = (Y'_{bi}, X'_i)'$ получаем оценку параметра $\hat{\theta}^b$ и статистику CK_n^b .
4. Повторяем шаги 2–3 для $b = 1, \dots, B$. Считываем p -значение статистики CK_n , равное пропорции CK_n^b , больших CK_n .

Кодирование такого алгоритма не составляет труда. Однако, время исполнения программы может оказаться ощутимым. Для теоретического обоснования применимости бутстрапа Andrews (1997) вывел для своей статистики разложение типа (9) и показал, что оно асимптотически одинаково как при нулевой гипотезе, так и при использовании сгенерированных бутстрапом данных.

3.2 Динамические модели и мартингалное преобразование

Ослабим теперь предположение, что Z_i являются НОР. А именно, предположим, что их распределение может меняться с i , и Y_i могут быть зависимы. Для простоты пусть $V = 1$. Маргинальные распределения не учитывают ни кросс-, ни временных зависимостей, поэтому при работе с кросс-секциями и временными рядами моделируют условные распределения.

¹Обзоры бутстрапа см. в журнале *Квантиль* №3 (2007).

Пусть Ω_i — вся информация, доступная нам «сегодня», т.е. минимальная σ -алгебра, содержащая $Z_1, Z_2, \dots, Z_{i-1}, X_i$. Рассмотрим гипотезу

$$H_0 : \exists \theta_0 \quad Y_i | \Omega_i \sim F_i(y | \cdot, \theta_0) \quad \forall i = 1, \dots, n. \quad (11)$$

Эту гипотезу можно переформулировать в терминах оценок прогноза ФР, см. Diebold, Gunther & Tay (1998), в противоположность оценкам точечного и интервального прогнозов.

С помощью гипотезы (11) можно протестировать большое количество динамических моделей. Приведем пример одного класса таких моделей. Пусть

$$Y_i = \mu(i, \theta) + \sigma(i, \theta)\varepsilon_i, \quad (12)$$

где $\mu(i, \theta)$ и $\sigma(i, \theta)$ Ω_i -измеримы, и ε_i — НОР с ФР $F_{\varepsilon_i}(\varepsilon, \theta)$. Этот класс включает в себя популярные модели гетероскедастичности. Скажем, модель AR(1)-GARCH(1,1)-N может быть представлена в виде

$$\begin{aligned} \mu(i, \theta) &= \omega_0 + \omega_1 Y_{i-1}, \\ \sigma^2(i, \theta) &= \alpha_0 + \beta_1 \sigma^2(i-1, \theta) + \alpha_1 (Y_{i-1} - \omega_0 - \omega_1 Y_{i-2})^2, \end{aligned}$$

$F_{\varepsilon_i}(\varepsilon, \theta) \sim N(\mu_\varepsilon, \sigma_\varepsilon^2)$, и $\theta = (\omega_0, \omega_1, \alpha_0, \alpha_1, \beta_1, \mu_\varepsilon, \sigma_\varepsilon)$. В то же время, не все параметрические модели можно записать в форме условного распределения. Например, GARCH-модели с не специфицированными ФР инноваций, GARCH-модели с некоррелированными (но, возможно, зависимыми) инновациями, SV-модели.² Для модели (12) легко выписать соответствующее условное распределение

$$F_i(y | \Omega_i, \theta) = F_{\varepsilon_i} \left(\frac{y - \mu(i, \theta)}{\sigma(i, \theta)}, \theta \right).$$

Если гипотеза (11) верна, а $F(y | \cdot, \theta_0)$ непрерывна по y , то

$$U_i = F_i(Y_i | \Omega_i, \theta_0) \sim \text{НОР } U[0, 1], \quad (13)$$

т.е. независимы и равномерно распределены на $[0, 1]$. Diebold, Gunther & Tay (1998) предложили визуальные методы проверки (13), а именно, с помощью гистограмм и коррелограмм первых четырех центральных моментов. Эти методы дают хорошее наглядное представление о спецификации, однако не могут учесть эффект от оценивания параметров. Действительно, при проверке сложной гипотезы мы не знаем истинное значение параметра θ_0 . Пусть у нас есть \sqrt{n} -состоятельная оценка $\hat{\theta}$. Тогда $\hat{U}_i = F_i(Y_i | \Omega_i, \hat{\theta})$, так же как и в безусловном случае, не обязаны быть равномерными и независимыми даже при правильной спецификации.

Формальное правило проверки (13) можно осуществить с помощью теории эмпирических процессов. Используя \hat{U}_i , построим эмпирическую ФР $F_n^{\hat{U}}(u)$ и процесс $\hat{V}_n(u)$, см. (8). Bai (2003) вывел для этого случая аналог разложения (9) и предложил использовать мартингальное преобразование (Khmaladze, 1981) для получения критических значений с учетом эффекта от оценивания параметра. А именно, при дополнительных предположении о непрерывной дифференцируемости ФР по параметру, разложение (9) имеет место с

$$g(u) = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left. \frac{\partial F_i(x | \Omega_i, \theta_0)}{\partial \theta} \right|_{x=F_i^{-1}(u | \Omega_i, \theta_0)}, \quad (14)$$

что позволяет увидеть в явном виде и обработать эффект от оценивания параметров. Например, мартингальное преобразование выглядит следующим образом (Bai, 2003):

$$\hat{W}_n(u) = \hat{V}_n(u) - \int_0^u \left(\bar{g}(s)' C^{-1}(s) \int_s^1 \bar{g}(\tau) d\hat{V}_n(\tau) \right) ds,$$

²См. обзор по GARCH моделям в России (2010), а по SV моделям — в Цыплаков (2010).

где $\bar{g}(u) = (1, \dot{g}(u)')'$, $\dot{g}(u)$ — производная $g(u)$ и $C(u) = \int_u^1 \bar{g}(\tau)\bar{g}(\tau)'d\tau$. Процесс $\hat{W}_n(u)$ слабо сходится к Броуновскому процессу, поэтому, опираясь на теорему о непрерывном преобразовании, распределение статистик на основе преобразованного процесса легко просимулировать и затабулировать. Сложность применения данного метода заключается в том, что выражение для $g(u)$ в (14) надо вычислять явно для каждой ФР. Примеры таких аналитических формул для GARCH(1,1), ARCH(p,q) и нелинейных динамических регрессий приведены в Bai (2003).

Теперь покажем, что тесты на базе V_n обнаруживают *локальные \sqrt{n} -альтернативы*, то есть распределения, приближающиеся к истинным со скоростью \sqrt{n} :

$$G_{ni}(y|\Omega_i, \theta_0) = \left(1 - \frac{\delta}{\sqrt{n}}\right) F_i(y|\Omega_i, \theta_0) + \frac{\delta}{\sqrt{n}} H_i(y|\Omega_i, \theta_0),$$

где $\delta > 0$, $1 < \delta/\sqrt{n}$, а H_i — условное распределение отличное от F_i , так что

$$k(u) = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H_i(F_i^{-1}(u|\Omega_i, \theta_0) | \Omega_i, \theta_0) - u \neq 0.$$

При локальной альтернативе новые СВ $U_i^* = G_{ni}(Y_i|\Omega_i, \hat{\theta})$ равномерно распределены и независимы, поэтому теперь нам понадобится разложение $\hat{V}_n(u)$ вокруг $V_n^*(u)$, построенному по U_i^*

$$V_n^*(u) = \sqrt{n} \left[F_n^{U_i^*}(u) - u \right], \quad u \in [0, 1],$$

и слабо сходящемуся к Броуновскому процессу. Bai (2003) установил аналог (9) для локальной альтернативы:

$$\hat{V}_n(u) = V_n^*(u) - g(u)\sqrt{n}(\hat{\theta} - \theta_0) + \delta k(u) + o_p(1)$$

равномерно по $u \in [0, 1]$. Если теперь применить мартингальное преобразование, то получим, что $\hat{W}_n(u)$ слабо сходится к Броуновскому процессу $W(u)$ плюс добавка:

$$\hat{W}_n(u) \Rightarrow W(u) + \delta k(u) - \delta \phi_g(k)(u),$$

где $\phi_g(k)(u) = \int_0^u (\dot{g}(s)')C^{-1}(s) \int_s^1 \dot{g}dk ds$. Таким образом, если удастся показать, что добавка ненулевая, то критерий обнаруживает локальные альтернативы.

Описанный тест имеет существенный недостаток. Легко видеть, что процесс (8) проверяет в явном виде равномерность, но не зависимость U_i . Поэтому такой тест плохо отлавливает динамику и зависимости в моделях. Например, если истинный процесс является авторегрессионным второго порядка, AR(2) с нормальными ошибками, а предполагаемый — AR(1), то такой тест имеет лишь номинальную мощность. Тест не работает и в общем случае при нарушении динамики и эллиптических ошибках (Corradi & Swanson, 2005). Более формально, (8) проверяет (11) в среднем, т. е. правильнее говорить о

$$H_0 : \exists \theta_0 \quad E(F_i(y|\Omega_i, \theta_0)) = F_i(y) \quad \forall i = 1, \dots, n.$$

Однако при такой нулевой гипотезе не гарантируется независимость U_i , меняются предельные распределения статистик, а это значит, что ни параметрический бутстрап, ни мартингальное преобразование использовать нельзя. Для проверки такой гипотезы Corradi & Swanson (2005) предложили приблизить предельное распределение статистики D_n , построенной по (8) с помощью блочного бутстрапа.

Для случая НОР Z_i и $V = 1$, Delgado & Stute (2008) предложили использовать двумерное преобразование Розенблатта (6). Если гипотеза (11) верна, преобразованные $U_i = F(X_i, \theta_0)$ и $V_i = F(Y_i|X_i, \theta_0)$ независимы и равномерно распределены на единичном квадрате $(U_i, V_i) \sim$

$U[0, 1]^2$. Это можно проверить сравнивая ФР, оцененную непараметрически и обозначенную как $F_n^{U,V}(u, v)$, и предполагаемую совместную ФР (U_i, V_i) :

$$V_{2n}(u, v) = \sqrt{n} [F_n^{U,V}(u, v) - uv], \quad u, v \in [0, 1]. \quad (15)$$

Этот процесс слабо сходится к двумерному Броуновскому мосту. Однако, при тестировании сложной гипотезы наблюдается эффект от оценивания параметра. Для борьбы с этим Delgado & Stute (2008) используют мартингальное преобразование к Броуновскому мосту.

Для тестирования в моделях временных рядов без экзогенных регрессоров, т. е. для проверки (11) при зависимых Y_i и $V = 0$, Kheifets (2010) предложил использовать процесс

$$V_{2n}(u_1, u_2) = \sqrt{n} [F_n^{U^{-1},U}(u_1, u_2) - u_1 u_2], \quad u_1, u_2 \in [0, 1], \quad (16)$$

где $F_n^{U^{-1},U}$ обозначает совместную эмпирическую ФР (U_i, U_{i-1}) , U_i получены в (13). Можно рассматривать также и другие лаги (U_i, U_{i-j}) , $j = 1, \dots, i-1$. В отличие от (15), (16) не является суммой независимых слагаемых, что не позволяет применять стандартные предельные теоремы. Kheifets (2010) изучил слабую сходимость процессов (16) и доказал законность использования параметрического бутстрапа для приближения критических значений суп-статистики

$$D_{2n} = \sup_{(u_1, u_2) \in [0, 1]^2} |V_{2n}(u_1, u_2)|$$

и ω^2 -статистики

$$\omega_{2n}^2 = \int_0^1 \int_0^1 V_{2n}^2(u_1, u_2) du_1 du_2.$$

4 Тестирование плотностей и диффузий

Для тестирования плотностей можно использовать подходы, заключающиеся в сравнении параметрических, зависящих от модели, и непараметрических оценок плотностей. В качестве непараметрических оценок плотностей берут, например, ядерные или серийные оценки, см. обзор Расин (2008). В отличие от оценок ФР, такие приближения сходятся медленнее, и свойства тестов зависят от настроек непараметрической оценки (например, выбор ядерной функции и ширины окна). Такие критерии не могут отличить \sqrt{n} -локальные альтернативы. В то же время, благодаря меньшей, чем \sqrt{n} скорости сходимости, нет эффекта от оценивания параметра при тестировании сложных гипотез.

Для оценки производных ценных бумаг (деривативов) используются модели непрерывного времени, в которых величины задаются как решения стохастических дифференциальных уравнений. Как проверять спецификацию таких моделей? Рассмотрим, как применять описанные методы для тестирования диффузионных процессов

$$dX_t = \mu_0(X_t, t)dt + \sigma_0(X_t, t)dW_t, \quad (17)$$

где W_t — стандартный Броуновский процесс. Тестирование параметрической диффузии заключается в проверке того, что дрейф $\mu_0(X_t, t)$ и диффузия $\sigma_0(X_t, t)$ принадлежат параметрическим семействам $\mu(X_t, t, \theta)$ и $\sigma(X_t, t, \theta)$, $\theta \in \Theta$ с вероятностью 1. Такая спецификация включает в себя большое количество моделей. Например, модель Vasicek (1977) задается как

$$\begin{aligned} \mu(x, t, \theta) &= \beta(\alpha - x), \\ \sigma(x, t, \theta) &= \gamma^2, \end{aligned}$$

и $\theta = (\alpha, \beta, \gamma)$, а модель Cox, Ingersoll & Ross (1985) задается как

$$\begin{aligned}\mu(x, t, \theta) &= \beta(\alpha - x), \\ \sigma(x, t, \theta) &= \gamma^2 x,\end{aligned}$$

и $\theta = (\alpha, \beta, \gamma)$.

Модель диффузии характеризует маргинальную плотность X_t , $\pi_0(x)$, и транзитивные плотности $p_0(x, t|y, s)$, $s < t$ (условная плотность $X_t = x$ при $X_s = y$), поэтому ее тестирование можно свести к тестированию плотностей. В качестве данных используют $\{X_{\tau\Delta}\}_{\tau=1}^n$, полученные дискретизацией временного периода $[0, T]$ интервалами длины Δ из непрерывной модели для X_t . Ait-Sahalia (1996) предложил тестировать диффузии, проверяя маргинальные плотности

$$H_0 : \exists \theta_0 \quad \pi(x, \theta_0) = \pi_0(x) \quad \text{почти всюду.}$$

Маргинальная плотность выражается в явном виде через дрейф и диффузию

$$\pi(x, \theta) = \frac{\xi(\theta)}{\sigma(x, \theta)} \exp \left[\int_{x_0}^x \frac{2\mu(u, \theta)}{\sigma(u, \theta)} du \right],$$

где выбор x_0 и $\xi(\theta)$ обеспечивает интегрирование плотности к 1. Непараметрическая ядерная оценка этой же плотности дает

$$\hat{\pi}_0(x) = \frac{1}{n} \sum_{\tau=1}^n \frac{1}{b_n} K \left(\frac{u - X_{\tau\Delta}}{b_n} \right),$$

где $K(\cdot)$ и b_n — ядерная функция и ширина окна. Ait-Sahalia (1996) установил асимптотическую нормальность статистики

$$M = nb_n \min_{\theta} \frac{1}{n} \sum_{\tau=1}^n (\pi(X_{\tau\Delta, \theta}) - \hat{\pi}_0(X_{\tau\Delta}))^2.$$

Как и критерий, предложенный в Bai (2003), такой тест плохо контролирует динамику.

Вместо маргинальных плотностей можно рассматривать переходные плотности, которые полностью определяют динамику. Рассмотрим

$$H_0 : \exists \theta_0 \quad p(x, t|y, s, \theta_0) = p_0(x, t|y, s) \quad \text{п.н.} \quad (18)$$

Транзитивные плотности плохо оцениваются непараметрически из-за большой размерности. Однако, при нулевой гипотезе (18)

$$U_{\tau} = \int_{-\infty}^{\tau\Delta} p(x, \tau\Delta | X_{(\tau-1)\Delta}, (\tau-1)\Delta, \theta_0) dx \sim \text{НОР } U[0, 1]. \quad (19)$$

Hong & Li (2005) использовали для проверки (19) разность ядерной оценки совместной плотности $(U_{\tau}, U_{\tau-1})$, $\hat{g}(u_1, u_2)$ и единицы — плотности равномерного распределения

$$Q = \int_0^1 \int_0^1 [\hat{g}(u_1, u_2) - 1]^2 du_1 du_2.$$

Можно рассматривать также и другие лаги $(U_{\tau}, U_{\tau-j})$, $j = 1, \dots, \tau - 1$. После должной нормировки предельное распределение оказывается стандартным нормальным, но скорость сходимости ниже, чем \sqrt{n} .

5 Заключение

В данном эссе мы рассмотрели различные методы тестирования предположений о функциональной форме ФР и плотностей. Часто такие критерии основываются на эмпирическом процессе, сравнивающем две оценки подходящей ФР: непараметрическую, сходящуюся к истинной ФР всегда, и параметрическую, сходящуюся к истинной ФР только если предположение верно.

Не всегда возможно затабулировать критические значения таких тестов. Распределение может зависеть от распределения и от факта оценивания параметров. В этом случае используют либо преобразование процесса (например, мартингальное преобразование к Броуновскому процессу), либо бутстрапируют критические значения. Мы видели примеры применения параметрического бутстрапа для обобщенного критерия Колмогорова и мартингального преобразования для табулирования критических значений теста в динамических моделях. В обоих случаях важную роль играет разложение оцененного процесса вокруг истинного. Это разложение также важно для вывода распределения процесса при локальной альтернативе.

В основном, мы рассматривали одномерные модели. Для тестирования многомерных динамических моделей можно использовать аналог преобразования Розенблатта, см., например, Diebold, Hahn & Tay (1999), Bai & Chen (2008), Kheifets (2011). Мы также предполагали непрерывность ФР. Velasco & Kheifets (2010) исследовали аналог (13) без предположения о непрерывности ФР и описали критерии для тестирования динамических моделей выбора. Анализ и сравнение других методов тестирования моделей выбора см. в Mora & Moro-Egido (2007).

Рассмотренные тесты должны работать против общих отклонений в противовес параметрическим тестам, заточенным на определенные альтернативы.

Список литературы

- Крил, М. (2008). Некоторые ловушки параметрической инференции. *Квантиль* 4, 1–6.
- Расин, Дж. (2008). Непараметрическая эконометрика: вводный курс. *Квантиль* 4, 7–56.
- Росси, Э. (2010). Одномерные GARCH-модели: обзор. *Квантиль* 8, 1–67.
- Цыплаков, А. (2010). Сделать тайное явным: искусство моделирования с помощью стохастической волатильности. *Квантиль* 8, 69–122.
- Ait-Sahalia, Y. (1996). Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies* 9, 385–426.
- Andrews, D.W.K. (1997). A conditional Kolmogorov test. *Econometrica* 65, 1097–1128.
- Bai, J. (2003). Testing Parametric Conditional Distributions of Dynamic Models. *The Review of Economics and Statistics* 85, 531–549.
- Bai, J. & Z. Chen (2008). Testing multivariate distributions in GARCH models. *Journal of Econometrics* 143, 19–36.
- Billingley, P. (1968) *Convergence of Probability Measures*. New York: John Wiley & Sons.
- Corradi, V. & R. Swanson (2006). Bootstrap conditional distribution test in the presence of dynamic misspecification. *Journal of Econometrics* 133, 779–806.
- Cox, J.C., Ingersoll, J.E. & S.A. Ross (1985). A theory of the term structure of interest rates. *Econometrica* 53, 385–407.
- Delgado, M. & W. Stute (2008). Distribution-free specification tests of conditional models. *Journal of Econometrics* 143, 37–55.
- Diebold, F.X., Gunther, T. & A.S. Tay (1998). Evaluating density forecasts with applications to finance and management. *International Economic Review* 39, 863–883.

- Diebold, F.X., Hahn, J. & A.S. Tay (1999). Multivariate density forecast evaluation and calibration in financial risk management: high frequency returns on foreign exchange. *Review of Economics and Statistics* 81, 661–673
- Durbin, J. (1973). Weak convergence of sample distribution functions when parameters are estimated. *Annals of Statistics* 1, 279–290.
- Hong, Y. & H. Li (2005). Nonparametric Specification Testing for continuous-Time Models with Applications to Term Structure of Interest Rates. *Review of Financial Studies* 72, 499–541.
- Kheifets, I. (2010). Specification tests for nonlinear time series models. Mimeo, Universidad Carlos III de Madrid.
- Kheifets, I. (2011). Specification tests for multivariate time series models. Mimeo, Universidad Carlos III de Madrid.
- Khmaladze, E.V. (1981). Martingale approach in the theory of goodness-of-tests. *Theory of Probability and its Applications* 26, 240–257.
- Kolmogorov, A.N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari* 4, 83.
- Mora, J. & A.I. Moro-Egido (2007). On specification testing of ordered discrete choice models. *Journal of Econometrics* 143, 191–205.
- Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics* 23, 470–72.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.
- Velasco, C. & I. Kheifets (2010) Model evaluation of the Fed monetary rules. Mimeo, Universidad Carlos III de Madrid.

Goodness-of-fit testing

Igor Kheifets

New Economic School, Moscow, Russia

We consider goodness-of-fit tests based on the empirical process theory. There are two main obstacles in obtaining critical values for such tests: the parameter estimation effect and distribution dependence. We discuss solutions to these problems: martingale transformation and bootstrap. As an illustration we show how to test GARCH and diffusion models.

Задачи и решения

Решения

Решение 8.1

Рассмотрим регрессию на стационарных временных рядах

$$y_t = \alpha + \beta x_t + e_t,$$

где ошибка e_t имеет нулевое среднее и не коррелирует с регрессором x_t . Нас интересует асимптотическая дисперсия V_β МНК-оценки β .

- (а) Постройте пример, в котором e_t серийно коррелирована, но тем не менее

$$V_\beta = \frac{\mathbb{V}[e_t]}{\mathbb{V}[x_t]},$$

то есть можно не поправлять стандартные ошибки на серийную корреляцию, хотя таковая имеется.

- (б) Постройте пример, в котором e_t серийно некоррелирована, но тем не менее

$$V_\beta \neq \frac{\mathbb{V}[x_t e_t]}{(\mathbb{V}[x_t])^2}.$$

то есть нужно поправлять стандартные ошибки на серийную корреляцию, хотя таковой в ошибках нет.

- (в) Какие последствия для предварительного тестирования на серийную корреляцию имеют результаты в (а) и (б)?

Данная задача была опубликована как Anatolyev (2001).

- (а) Пусть x_t – НОД-процесс с нулевым средним, независимый от e_t , тогда для $j \neq 0$, хотя и $\mathbb{C}(e_t, e_{t-j}) \neq 0$, выполнено

$$\mathbb{C}[x_t e_t, x_{t-j} e_{t-j}] = \mathbb{C}[x_t, x_{t-j}] \mathbb{C}[e_t, e_{t-j}] = 0$$

(где первое равенство следует из того, что x_t – НОД с нулевым средним), в то время как

$$\mathbb{V}[x_t e_t] = \mathbb{V}[x_t] \mathbb{V}[e_t],$$

так что

$$V_\beta = \frac{\mathbb{V}[e_t]}{\mathbb{V}[x_t]}.$$

- (б) Пусть x_t – процесс с нулевым средним и единичной дисперсией, следующий авторегрессии первого порядка, порожденной строгим белым шумом:

$$x_t = \rho x_{t-1} + \eta_t, \quad \eta_t \sim IID(0, 1 - \rho^2),$$

где $0 < \rho < 1$. Пусть стандартный двумерный нормальный белый шум $(v_t, w_t)'$ независим от процесса x_t . Построим e_t как

$$e_t = v_{t+1} + w_{t+1} x_{t+1} + \theta(v_t - w_t x_t),$$

где $0 < \theta < 1$. Тогда можно обнаружить, что

$$\mathbb{C}[x_t, e_t] = 0, \quad \mathbb{C}[e_t, e_{t-j}] = 0 \quad \forall j \neq 0$$

и

$$\mathbb{C}[x_t e_t, x_{t-1} e_{t-1}] \neq 0, \quad \mathbb{C}[x_t e_t, x_{t-j} e_{t-j}] = 0 \quad \forall j > 1.$$

Поэтому

$$V_{\beta} = \frac{\mathbb{V}[x_t e_t] + 2\mathbb{C}[x_t e_t, x_{t-1} e_{t-1}]}{(\mathbb{V}[x_t])^2} \neq \frac{\mathbb{V}[x_t e_t]}{(\mathbb{V}[x_t])^2}.$$

- (с) Следовательно, более разумно тестировать на серийную корреляцию процесс $x_t e_t$, а не процесс e_t . Аналогично, следует тестировать процесс $z_t e_t$, если рассматривается оценивание с инструментальными переменными с вектором инструментов z_t .

Решение 8.2

Рассмотрим стандартную панельную линейную регрессию с индивидуальными эффектами. Пусть для простоты константа отсутствует, а все переменные имеют нулевое среднее. Разработайте тест на случайные эффекты на основе разности внутри-оценки и между-оценки коэффициентов регрессии.

Пусть $P = I_n \otimes \frac{1}{T} \omega \omega'$, $Q = I_{nT} - P$, и вспомним, что $\hat{\beta}_W - \beta = (X'QX)^{-1} X'QU$ и $\hat{\beta}_B - \beta = (X'PX)^{-1} X'PU$, где U – вектор двухкомпонентной ошибки, и что $\mathbb{V}[U|X] = \alpha_Q Q + \alpha_P P$ для определенных весов α_Q и α_P , зависящих от дисперсий компонент. Тогда, условно на X ,

$$\begin{aligned} \mathbb{C}[\hat{\beta}_W, \hat{\beta}_B] &= (X'QX)^{-1} X'Q(\alpha_Q Q + \alpha_P P)PX(X'PX)^{-1} \\ &= \alpha_Q (X'QX)^{-1} X'QQPX(X'PX)^{-1} + \alpha_P (X'QX)^{-1} X'QPPX(X'PX)^{-1} \\ &= 0, \end{aligned}$$

поскольку $QP = PQ = 0$. Используя этот результат, имеем, условно на X ,

$$\begin{aligned} \mathbb{V}[\hat{\beta}_W - \hat{\beta}_B] &= \mathbb{V}[\hat{\beta}_W] + \mathbb{V}[\hat{\beta}_B] \\ &= \alpha_Q (X'QX)^{-1} + \alpha_P (X'PX)^{-1}. \end{aligned}$$

Известно, что при случайных эффектах $\hat{\beta}_W$ и $\hat{\beta}_B$ состоятельны и асимптотически нормальны, поэтому тестовая статистика

$$\mathcal{R} = (\hat{\beta}_W - \hat{\beta}_B)' \left(\hat{\alpha}_Q (X'QX)^{-1} + \hat{\alpha}_P (X'PX)^{-1} \right)^{-1} (\hat{\beta}_W - \hat{\beta}_B)$$

асимптотически распределена как хи-квадрат с $k = \dim(\beta)$ степенями свободы при случайных эффектах. Здесь $\hat{\alpha}_Q = RSS_W / (nT - n - k)$ и $\hat{\alpha}_P = RSS_B / (n - k)$. Когда случайные эффекты неприемлемы, оценка $\hat{\beta}_W$ по-прежнему состоятельна, но $\hat{\beta}_B$ несостоятельна, так что $\hat{\beta}_W - \hat{\beta}_B$ сходится к ненулевой предельной величине, в то время как $X'QX$ и $X'PX$ расходятся, из-за чего \mathcal{R} тоже расходится.

Решение 8.3

Выведите асимптотические свойства оценки Надарая–Уотсона регрессионной функции в регрессии с абсолютной подгонкой, т.е. когда регрессионной ошибки нет.

Используем обозначения, принятые в Анатолев (2009). Когда дисперсия ошибки равна нулю,

$$\hat{g}(x) - g(x) = \frac{(nh)^{-1} \sum_{i=1}^n (g(x_i) - g(x)) K(h^{-1}(x_i - x))}{(nh)^{-1} \sum_{i=1}^n K(h^{-1}(x_i - x))}.$$

Обычного источника дисперсии, регрессионной ошибки, в данном случае нет, так что дисперсия оценки порождается дисперсией регрессоров.

Знаменатель сходится к $f(x)$ при $n \rightarrow \infty, h \rightarrow 0$. Рассмотрим числитель, обозначив его как $\hat{q}(x)$. Он представляет собой среднее НОР случайных величин, скажем ζ_i . Следуя Анатолев (2009), получаем, что

$$\mathbb{E}[\hat{q}(x)] = h^2 B(x) f(x) + o(h^2), \quad \mathbb{V}[\hat{q}(x)] = o((nh)^{-1}).$$

Взглянем попристальней на дисперсию.

$$\begin{aligned} \mathbb{E}[\zeta_i^2] &= h^{-2} \int (g(x_i) - g(x))^2 K\left(\frac{x_i - x}{h}\right)^2 f(x_i) dx_i \\ &= h^{-1} \int (g(x + hu) - g(x))^2 K(u)^2 f(x + hu) du \\ &= h^{-1} \int (g'(x)hu + o(h))^2 K(u)^2 (f(x) + o(h)) du \\ &= hg'(x)^2 f(x) \int u^2 K(u)^2 du + o(h), \end{aligned}$$

так что

$$\mathbb{V}[\zeta_i] = \mathbb{E}[\zeta_i^2] - \mathbb{E}[\zeta_i]^2 = hg'(x)^2 f(x) \Psi_K^2 + o(h),$$

где $\Psi_K^2 = \int u^2 K(u)^2 du$. Поэтому, согласно ЦПТ,

$$\begin{aligned} &\sqrt{nh^{-1}} \left((nh)^{-1} \sum_{i=1}^n (g(x_i) - g(x)) K\left(\frac{x_i - x}{h}\right) - h^2 B(x) f(x) + o(h^2) \right) \\ &\xrightarrow{d} \mathcal{N}\left(0, g'(x)^2 f(x) \Psi_K^2\right). \end{aligned}$$

Пусть $\lambda = \lim_{n \rightarrow \infty, h \rightarrow 0} \sqrt{nh^3}$, и предположим, что $\lambda < \infty$. Тогда

$$\sqrt{nh^{-1}} (\hat{g}(x) - g(x)) \xrightarrow{d} \mathcal{N}\left(\lambda B(x), \frac{g'(x)^2}{f(x)} \Psi_K^2\right).$$

Список литературы

Анатолев, С. (2009). Непараметрическая регрессия. *Квантиль* 7, 37–52.

Anatolyev, S. (2001). Serial correlation and asymptotic variance. *Econometric Theory* 17, задача 01.5.3, стр. 1026.

Статьи: финансовая эконометрика

Моделирование многомерных параметрических плотностей финансовых доходностей*

Алексей Балаев[†]

Высшая школа экономики, Москва, Россия

В данной работе сравниваются некоторые двумерные параметризации условных функций плотности для доходностей фондовых индексов. Сравнение проводится в терминах качества внутривыборочной подгонки и предсказательной способности вне выборки при предсказании условной функции плотности в целом. Рассматриваются скошенное нормальное распределение, скошенное распределение Стьюдента, скошенное обобщенное распределение ошибки и распределение, основанное на разложении Грамма–Шарлье. Основное внимание уделяется способности данных функций плотности учитывать асимметрию и толщину так называемых «многомерных хвостов» распределения. Используя тест, основанный на информационном критерии Кульбака–Лейблера, мы проводим попарное сравнение оцененных моделей для условной функции плотности отдельно внутри и вне выборки. Затем модели упорядочиваются по качеству подгонки и предсказательной способности. В работе обсуждаются причины превосходства той или иной спецификации функции плотности над другой.

Ключевые слова: условная функция плотности, разложение Грамма–Шарлье, скошенное распределение, качество подгонки, предсказательная способность
Классификация JEL: C14, C16, C22, C32, C51, C53, C58

1 Введение

Типичная задача в теории выбора финансового портфеля – это максимизация ожидаемой доходности портфеля при наличии доступной к данному моменту времени информации и некоторых ограничений на уровень риска портфеля. В теории для построения оптимального портфеля инвестору необходимо знание совместного распределения доходностей имеющихся на рынке активов, условного на доступной к текущему моменту времени информации. В реальности, однако, такое знание отсутствует, и это условное распределение должно оцениваться. Как правило, интерес представляет оценка условной функции плотности для вектора доходностей доступных активов.

Для оценивания этой функции плотности можно исходить из различных параметризаций или использовать копула-функции (копулы). В настоящее время копульный подход приобретает все большую популярность на практике. Однако при изучении литературы возникает ощущение, что копулами начали пользоваться, не придя к консенсусу в параметрическом подходе. В рамках параметрического подхода остаются вопросы: какие параметризации функции плотности следует использовать, каково их соответствие данным внутри выборки оценивания, и насколько хорошо с их помощью предсказывается условное распределение доходностей вне выборки оценивания. В настоящей статье мы в рамках параметрического подхода частично отвечаем на эти вопросы.

*Цитировать как: Балаев, Алексей (2011) «Моделирование многомерных параметрических плотностей финансовых доходностей», Квантиль, №9, стр. 39–60. Citation: Balaev, Alexey (2011) “Modeling multivariate parametric densities of financial returns,” *Quantile*, No.9, pp. 39–60.

[†]Адрес: 109028, г. Москва, Покровский бульвар, д. 11. Электронная почта: a.balaev@gmail.com

В работе проведено сравнение некоторых двумерных параметризаций условных функций плотности для доходностей нескольких фондовых индексов. Сравнение проводится в терминах качества внутривыборочной подгонки и предсказательной способности вне выборки оценивания при предсказании условной функции плотности в целом. Рассматриваются скошенное нормальное распределение, скошенное распределение Стьюдента, скошенное распределение GED (обобщенное распределение ошибки) и распределение, основанное на разложении Грамма–Шарлье (для краткости, далее – распределение Грамма–Шарлье). Первые два из упомянутых распределений достаточно широко известны, чего нельзя сказать о двух последних. При выборе параметризаций основное внимание уделялось эффектам, порождаемым формой соответствующих функций плотности. Поэтому мы не используем сложные структуры для вектора условных средних и условной ковариационной матрицы.

Все модели оцениваются и тестируются на выборках дневных доходностей фондовых индексов различных стран. Используя тест, основанный на информационном критерии Кульбака–Лейблера (далее, KLIC-тест), мы проводим попарное сравнение оцененных моделей для условной функции плотности отдельно внутри и вне выборки оценивания. Затем модели упорядочиваются по качеству подгонки и предсказательной способности. Процедура приводит к следующему ранжированию распределений. Внутри выборки оценивания: 1 – скошенное распределение Стьюдента, 2 – скошенное распределение GED и распределение Грамма–Шарлье, 3 – скошенное нормальное распределение. Вне выборки оценивания: 1 – скошенное распределение Стьюдента и скошенное распределение GED, 2 – распределение Грамма–Шарлье, 3 – скошенное нормальное распределение. В работе обсуждаются причины превосходства той или иной спецификации функции плотности над другой внутри или вне выборки оценивания.

Работа построена следующим образом. В разделе 2 приведены спецификации для вектора условных средних и ковариационной матрицы, которые используются во всех моделях условной функции плотности, рассматриваемых в работе. В разделе 3 приведено описание параметризаций многомерных функций плотности, которые используются для учета асимметрии и так называемых «многомерных тяжелых хвостов» в распределениях векторов доходностей. Раздел 4 посвящен описанию KLIC-теста и его применению для сравнения невложенных моделей для функции плотности. В разделе 5 дано описание использованных данных, и проведен некоторый предварительный эмпирический анализ. В разделах 6 и 7 представлены результаты оценивания всех построенных моделей для двумерных условных функций плотности, а также результаты эмпирического сравнения этих моделей. Наконец, в разделе 8 приводятся выводы и возможные направления дальнейшей работы.

2 Условное среднее и ковариационная матрица

В этом разделе описаны простые спецификации для вектора условных средних и условной ковариационной матрицы, которые в дальнейшем используются со всеми параметризациями условной функции плотности, рассмотренными в разделе 3.

В одномерном случае довольно часто AR-компонента в динамике финансовых доходностей является незначимой. Поэтому, как правило, условное среднее в одномерном случае можно игнорировать (например, его можно зафиксировать на уровне 0) без значительных потерь точности любой модели для условной функции плотности распределения доходностей.

Однако в многомерном случае следует всегда использовать нетривиальную спецификацию вектора условных средних, поскольку доходности различных активов (или различных рынков) могут коррелировать. Например, в разделе 5 будет показано, что существуют довольно четкие динамические связи между доходностями фондовых индексов различных стран. Лагированная величина доходности на одном рынке может рассматриваться как информация, которая влияет на доходности на других рынках.

Пусть N – число рынков (или активов) в модели, а r_t – N -мерный вектор доходностей этих

рынков в момент времени t . Для того чтобы учесть в модели упомянутые динамические связи между рынками, мы используем простую векторную авторегрессию 1-го порядка с вектором-константой, то есть модель VAR(1):

$$r_t = c + Qr_{t-1} + \epsilon_t, \quad (1)$$

где, как обычно принято обозначать, ϵ_t – это N -мерный вектор шоков.

Это простейшая из доступных спецификация вектора условных средних. Безусловно, можно использовать и более сложные модели условного среднего. Например, в структуре условного среднего можно учитывать многомерные GARCH-эффекты, что реализовано в моделях MGARCH-in-mean (MGARCH-эффекты в условном среднем). Однако, как было упомянуто, нас главным образом интересует форма оцениваемых функций плотности. В такой постановке, на наш взгляд, простой VAR(1) спецификации для условного среднего вполне достаточно.

Далее, одномерный анализ финансовых доходностей практически всегда выявляет статистически значимую GARCH-структуру для их условной дисперсии. Поэтому в многомерном случае эффекты кластеризации волатильности также должны быть учтены. При этом не исключено, что если доходности на различных рынках связаны, то и их волатильности могут быть связаны. Для учета стандартных GARCH-эффектов и возможных связей между волатильностями мы используем многомерную GARCH-модель BEKK(1,1)¹ для условной ковариационной матрицы вектора ϵ_t . Ее спецификация является достаточно гибкой и гарантирует положительную определенность условной ковариационной матрицы в любой момент времени.² Модель BEKK(1,1) имеет следующую структуру:

$$\begin{aligned} \epsilon_t &= H_t^{1/2} \eta_t, \\ H_t &= \Omega\Omega' + A\epsilon_{t-1}\epsilon_{t-1}'A' + BH_{t-1}B', \end{aligned} \quad (2)$$

$$\eta_t \sim N(0, I_N), \quad (3)$$

где H_t – условная ковариационная матрица, а η_t – вектор так называемых стандартизованных ошибок.

Это одна из простейших спецификаций для условной ковариационной матрицы вектора доходностей. Существует масса других MGARCH-моделей, из которых можно выбирать. Например, для H_t можно использовать широко известную в литературе модель DCC-GARCH³ Но опять-таки, поскольку нашей основной задачей является сравнение форм функций плотности, мы не используем сложные спецификации для ковариационной матрицы. Более того, использование BEKK-модели выгодно тем, что это одна из наиболее гибких на сегодняшний день MGARCH-моделей.

3 Распределение стандартизованных ошибок

В рамках нашего анализа мы заменяем многомерное нормальное распределение в уравнении (3) более общим условным распределением с функцией плотности $p_{t-1}(\eta_t|\theta)$, где θ – конечный вектор параметров.

В этом разделе рассмотрены 4 параметризации для функции плотности $p_{t-1}(\eta_t|\theta)$, которые мы затем используем для моделирования распределения вектора стандартизованных ошибок η_t . Для каждой параметризации описано то, как в ней учитывается асимметрия и многомерные «тяжелые хвосты» распределений.

¹ Baba–Engle–Kraft–Kroner MGARCH-модель, предложенная в Engle & Kroner (1995).

² Обзор MGARCH-моделей и обсуждение возникающих при их построении проблем можно найти, например, в Silvennoinen & Terasvirta (2008).

³ Предложена в Engle (2002).

Параметры во всех рассматриваемых параметризациях функции $p_{t-1}(\eta_t|\theta)$ предполагаются постоянными. Это делает модели в некотором смысле равноправными и упрощает их сравнение. Если предположить изменяющиеся во времени параметры, то желательно для каждой параметризации найти наилучшую модель динамики параметров. На наш взгляд, это достаточно сложная задача, и в данной работе мы ограничиваемся предположением о постоянных параметрах.

3.1 Скошенное нормальное распределение

Функция плотности многомерного скошенного нормального распределения, являющаяся обобщением соответствующей одномерной функции плотности, была предложена в Azzalini & Dalla Valle (1996) и дополнительно обобщена в Azzalini & Capitanio (1999).

Пусть $\mu \in R^N$ и τ – параметры расположения и масштаба, где $\tau = \text{diag}(\tau_1, \dots, \tau_N)$, $\tau_i > 0$, $i = 1, \dots, N$. Тогда для вектора стандартизованных ошибок условная функция плотности скошенного нормального распределения запишется в виде

$$p_{t-1}(\eta_t|\mu, \tau, \Sigma, \alpha) = 2\phi_N(\eta_t - \mu, \Sigma)\Phi(\alpha'\tau^{-1}(\eta_t - \mu)), \quad (4)$$

где $\phi_N(\cdot, \Omega)$ – N -мерная функция плотности нормального распределения с нулевым вектором средних и ковариационной матрицей Ω , $\Phi(\cdot)$ – одномерная функция распределения $N(0, 1)$, α – N -мерный вектор, определяющий направления скошенности. Центр, вокруг которого располагается плотность, уже учитывается в используемой нами VAR(1)-модели для условного среднего, и поэтому мы полагаем $\mu = 0$. Вектор параметров α не может быть идентифицирован при произвольной матрице τ . Поэтому мы также полагаем $\tau = I_N$ перед оцениванием модели.

Вектор параметров асимметрии α напрямую контролирует направления скошенности распределения, и случай $\alpha = 0$ соответствует симметричному нормальному распределению. Основным недостатком данной параметризации состоит, очевидно, в том, что она не позволяет контролировать толщину многомерных хвостов распределения. Поэтому можно ожидать, что оцененная по данной параметризации функция плотности не будет иметь многомерных тяжелых хвостов и, как результат, одномерных тяжелых хвостов в соответствующих маргинальных распределениях, что является одной из основных широко известных характеристик распределения финансовых доходностей.

3.2 Скошенное распределение Стьюдента

В работе Bauwens & Laurent (2005) предложена общая процедура введения асимметрии в симметричные многомерные распределения. Применяя данную процедуру к многомерному распределению Стьюдента, Bauwens & Laurent (2005) строят так называемое многомерное скошенное распределение Стьюдента.

Пусть ν – скаляр, параметр степеней свободы и ξ – вектор направлений скошенности распределения. Тогда для вектора стандартизованных ошибок условная функция плотности скошенного распределения Стьюдента запишется в виде⁴

$$p_{t-1}(\eta_t|\xi, \nu) = \left(\frac{2}{\sqrt{\pi}}\right)^N \frac{\Gamma((\nu + N)/2)}{\Gamma(\nu/2)(\nu - 2)^{N/2}} \left(1 + \frac{\eta_t^* \eta_t^*}{\nu - 2}\right)^{-(N+\nu)/2} \prod_{i=1}^N \frac{\xi_i s_i}{1 + \xi_i^2}, \quad (5)$$

где использованы следующие обозначения:

$$\eta_t^* = (\eta_{t1}^*, \dots, \eta_{tN}^*)',$$

⁴ В разделе 6 при оценке соответствующей модели мы используем немного другую параметризацию для скошенного распределения Стьюдента из-за специфики процедуры языка GAUSS, которую мы применяем при оценивании.

$$\begin{aligned}\eta_{ti}^* &= (s_i \eta_{ti} + m_i) \xi_i^{I_i}, \\ m_i &= \frac{\Gamma((\nu - 1)/2) \sqrt{\nu - 2}}{\sqrt{\pi} \Gamma(\nu/2)} \left(\xi_i - \frac{1}{\xi_i} \right), \\ s_i^2 &= \left(\xi_i^2 + \frac{1}{\xi_i^2} - 1 \right) - m_i^2\end{aligned}$$

и

$$I_i = \begin{cases} -1, & \text{если } \eta_{ti} \geq -m_i/s_i, \\ 1, & \text{если } \eta_{ti} < -m_i/s_i. \end{cases}$$

Это довольно гибкая параметризация. Асимметрия распределения напрямую контролируется вектором ξ , а толщина многомерных хвостов контролируется параметром степеней свободы ν .

3.3 Скошенное распределение GED

Насколько известно автору, симметричное многомерное распределение GED⁵ было впервые предложено в работе Giller (2005). В рамках нашего анализа для вектора стандартизованных ошибок симметричная многомерная условная функция плотности распределения GED запишется в виде

$$p_{t-1}(\eta_t | \kappa) = \frac{1}{\sqrt{\pi}^N} \frac{\Gamma(1 + N/2)}{\Gamma(1 + n\kappa)} \left(\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right)^{N/2} \exp \left(- \left(\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \eta_t' \eta_t \right)^{1/2\kappa} \right), \quad (6)$$

где κ – скалярный параметр, контролирующий толщину многомерных хвостов распределения. Опять-таки, насколько известно автору, в литературе не рассматривалась скошенная версия данного распределения. Для построения такого распределения можно применить процедуру, предложенную в Bauwens & Laurent (2005), к симметричному распределению GED.

Данная процедура на основе симметричной N -мерной функции плотности $f(x)$ строит функцию плотности $f^s(x|\xi)$, допускающую скошенность с вектором направлений скошенности ξ . Построение основано на следующих преобразованиях:

$$f^s(x|\xi) = 2^N f(x^*) \prod_{i=1}^N \frac{\xi_i}{1 + \xi_i^2},$$

$$x^* = (x_1^*, \dots, x_N^*)',$$

$$x_i^* = x_i \xi_i^{I_i}, \quad i = 1, \dots, N,$$

$$I_i = \begin{cases} -1, & \text{если } x_i \geq 0, \\ 1, & \text{если } x_i < 0. \end{cases}$$

Применяя эти преобразования к функции (6), получаем условную функцию плотности скошенного распределения GED для вектора стандартизованных ошибок в виде

$$p_{t-1}(\eta_t | \xi, \kappa) = \frac{\Gamma(1 + N/2)}{\Gamma(1 + n\kappa)} \left(\frac{3\Gamma(3\kappa)}{\pi\Gamma(\kappa)} \right)^{N/2} \exp \left(- \left(\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \eta_t^{*'} \eta_t^* \right)^{1/2\kappa} \right) \prod_{i=1}^N \frac{\xi_i}{1 + \xi_i^2}, \quad (7)$$

$$\eta_t^* = (\eta_{t1}^*, \dots, \eta_{tN}^*)',$$

⁵ Generalized Error Distribution, или обобщенное распределение ошибки.

$$\eta_{ti}^* = \eta_{ti} \xi_i^{I_i}, \quad i = 1, \dots, N,$$

$$I_i = \begin{cases} -1, & \text{если } \eta_{ti} \geq 0, \\ 1, & \text{если } \eta_{ti} < 0. \end{cases}$$

Эта параметризация также является довольно гибкой. Как и для скошенного распределения Стьюдента, асимметрия напрямую контролируется вектором ξ , а толщина многомерных хвостов контролируется параметром κ .

3.4 Распределение Грамма–Шарлье

Функция плотности данного распределения получается на основе разложения истинной многомерной условной функции плотности вектора стандартизованных ошибок в ряд Грамма–Шарлье вокруг распределения $N(0, I_N)$. В полученном ряде все члены после третьего отбрасываются, а оставшаяся часть объявляется функцией плотности. Для вектора стандартизованных ошибок данная условная функция плотности имеет вид⁶

$$p_{t-1}(\eta_t | D_3, D_4) = \phi_N(\eta_t, I_N) \left(1 + \frac{1}{3!} \text{tr}(D_3 [(\eta_t \eta_t' - 3I_N) \otimes \eta_t]) + \frac{1}{4!} \text{tr}([D_4 - 3 \text{vec}(I_N) \text{vec}(I_N)'] \Omega(\eta_t)) \right), \quad (8)$$

где

$$\Omega(\eta_t) = \text{vec}(\eta_t \eta_t' - I_N) \text{vec}(\eta_t \eta_t' - I_N)' - 4I_N \otimes (\eta_t \eta_t') + 2I_{N^2},$$

и D_3, D_4 – матрицы соответственно третьих и четвертых моментов истинного условного распределения стандартизованных ошибок с истинной функцией плотности $p_{t-1}(\eta_t)$, которую мы хотим аппроксимировать:

$$D_3 = \int_{R^N} \eta \eta' \otimes \eta' p_{t-1}(\eta) d\eta,$$

$$D_4 = \int_{R^N} \eta \eta' \otimes \eta \eta' p_{t-1}(\eta) d\eta.$$

Заметим, что поскольку мы рассматриваем только первые три члена ряда Грамма–Шарлье, функция (8) не является функцией плотности распределения в соответствии со строгим определением. К счастью, эта функция интегрируется в единицу для любых матриц D_3 и D_4 . Однако для некоторых D_3 и D_4 существуют области вектора η_t , в которых данная функция принимает отрицательные значения.⁷ Как мы увидим в разделе 6, это приводит к значительным трудностям при оценивании соответствующей модели.

Чтобы гарантировать положительность функции плотности Грамма–Шарлье на всем R^N , можно возвести в квадрат некоторые из ее членов. Однако данное преобразование приводит к излишне нелинейной функции, которая к тому же, как впрочем и исходная функция, содержит довольно много параметров. В частности, для случая $N = 2$, который мы рассматриваем в разделе 6, модель VAR(1)–BEKK(1,1) с условной функцией плотности Грамма–Шарлье для стандартизованных ошибок предполагает оценивание 26 параметров. При такой высокой размерности после возведения в квадрат каких-либо членов в функции плотности Грамма–Шарлье сходимость процедур оптимизации при оценивании модели значительно ухудшается. Из-за этих сложностей мы рассматриваем только исходную нескорректированную версию функции плотности Грамма–Шарлье (8).

⁶ Подробное описание построения данной функции можно найти, например, в Del Brio, Niguez & Perote (2008).

⁷ На Рис. 4 в Приложении показаны области для параметров индивидуальной скошенности и куртозиса, в которых функция плотности Грамма–Шарлье как функция аргумента η_t всюду положительна и унимодальна при условии что компоненты вектора η_t некоррелированы.

4 Процедура сравнения моделей

В этом разделе описывается метод сравнения двух конкурирующих моделей условной функции плотности как внутри, так и вне выборки оценивания. Поскольку рассматриваемые модели функции плотности являются невлуженными, для такого сравнения удобно использовать информационный критерий Кульбака–Лейблера (Vuong, 1989). Сравнение двух функций плотности проводится в форме теста на основе этого информационного критерия (далее, КЛИС-теста). В данном разделе КЛИС-тест описан в контексте моделирования условной функции плотности для вектора стандартизованных ошибок $\eta_t = (\eta_{t1}, \dots, \eta_{tN})'$.

Пусть имеются две конкурирующие условные функции плотности для вектора стандартизованных ошибок $f(\eta_t|I_{t-1}, \theta)$ и $g(\eta_t|I_{t-1}, \phi)$. Чтобы выбрать из них лучшую, с точки зрения качества подгонки внутри выборки или предсказательной способности вне ее, проводится следующий тест.

Нулевая гипотеза состоит в эквивалентности моделей:

$$H_0 : f(\eta_t|I_{t-1}, \theta) \sim g(\eta_t|I_{t-1}, \phi).$$

Это означает, что расстояние Кульбака–Лейблера между ними равно нулю:

$$E \left[\log \frac{f(\eta_t|I_{t-1}, \theta_*)}{g(\eta_t|I_{t-1}, \phi_*)} \right] = 0,$$

где θ_* и ϕ_* – это псевдоистинные значения параметров для двух моделей функции плотности.

Имеются две альтернативные гипотезы, соответствующие случаям, когда одна модель предпочтительнее другой:

$$H_f : f(\eta_t|I_{t-1}, \theta) \succ g(\eta_t|I_{t-1}, \phi),$$

$$H_g : g(\eta_t|I_{t-1}, \phi) \succ f(\eta_t|I_{t-1}, \theta).$$

Обозначим через ω_* дисперсию логарифмической разности рассматриваемых функций плотности:

$$\omega_* = \mathbb{V} \left(\log \frac{f(\eta_t|I_{t-1}, \theta_*)}{g(\eta_t|I_{t-1}, \phi_*)} \right).$$

Эту дисперсию можно состоятельно оценить следующим образом:

$$\widehat{\omega}_* = \frac{1}{T} \sum_{t=2}^T \left(\log \frac{f(\eta_t|I_{t-1}, \widehat{\theta})}{g(\eta_t|I_{t-1}, \widehat{\phi})} \right)^2 - \left(\frac{1}{T} \sum_{t=2}^T \log \frac{f(\eta_t|I_{t-1}, \widehat{\theta})}{g(\eta_t|I_{t-1}, \widehat{\phi})} \right)^2,$$

где $\widehat{\theta}$ и $\widehat{\phi}$ – это оценки параметров двух моделей методом максимального квазиправдоподобия (QML-оценки).

Далее, обозначим через LR сумму выборочных логарифмических разностей рассматриваемых функций плотности:

$$LR = \sum_{t=2}^T \log \frac{f(\eta_t|I_{t-1}, \widehat{\theta})}{g(\eta_t|I_{t-1}, \widehat{\phi})}.$$

КЛИС-тест основан на следующей асимптотике LR в рамках каждой из гипотез:

$$\frac{LR}{\sqrt{T\widehat{\omega}_*}} \xrightarrow{H_0} N(0, 1), \quad \xrightarrow{H_f} +\infty, \quad \xrightarrow{H_g} -\infty.$$

КЛИС-тест может быть проведен как внутри выборки оценивания (для сравнения качества подгонки моделей) так и вне ее (для сравнения предсказательной способности моделей).

Потенциальная проблема данного теста – возможность ситуации, когда $\omega_* = 0$. Для проверки этого равенства необходимо проводить предварительное тестирование. Однако, существующий на данный момент тест для проверки гипотезы $\omega_* = 0$ использует статистику с нестандартным и непивотальным распределением, и чтобы не усложнять расчеты, мы не проводим предварительное тестирование в данной работе.

5 Данные и предварительный анализ

В этом разделе описаны используемые в расчетах данные и предварительно отмечены некоторые их свойства, которые предполагается учесть в оцениваемых двумерных моделях условной функции плотности.

5.1 Данные

Мы используем дневные цены закрытия фондовых индексов разных стран: S&P500 (США), FTSE100 (Великобритания), CAC40 (Франция), DAX (Германия), Hang-Seng (Китай), NIKKEI225 (Япония).⁸ Стартовая дата – 26 ноября 1990 г. – первый день, когда начали вычислять немецкий индекс DAX. Последняя дата в выборке – 29 августа 2008 г. Финансовый кризис вызвал ультравысокую волатильность в сентябре 2008 г., и поэтому мы не включаем в наши данные динамику индексов начиная с этого месяца, поскольку в этом случае оценки параметров моделей слишком чувствительны к малым изменениям в данных. Мы работаем с дневными логарифмическими доходностями на индексы, то есть с величинами $r_t = 100 \ln(S_t/S_{t-1})$, где S_t – значение индекса в момент времени t .

Чтобы иметь возможность оценивать многомерные условные функции плотности, мы синхронизировали данные. То есть мы исходим из данных о ценах только в такие дни, когда торговались все 6 индексов. Чтобы оставить в выборке только такие дни, пришлось удалить некоторое количество наблюдений для каждого из индексов. Поэтому, строго говоря, построенные логарифмические доходности соответствуют временным промежуткам различной длины. В таблице 1 приведено распределение длин временных интервалов, соответствующих доходностям в нашей выборке.

Таблица 1. Временные интервалы доходностей

Дней между торгами	0	1	2	3	4	5	6	7
Доля в выборке, %	72,9	3,49	15,18	5,41	1,64	0,81	0,51	0,05
Накопленная доля, %	72,9	76,39	91,57	96,99	98,63	99,44	99,95	100

Двухдневные промежутки (суббота и воскресенье) неизбежны при моделировании дневных доходностей. Потенциально, проблемными могут оказаться доходности за период в 3 и более дней. Однако, как видно из таблицы, для наших данных такие доходности составляют около 8% наблюдений, что на наш взгляд является приемлемым. По этой причине мы не учитываем специфику доходностей, соответствующих промежуткам в 3 и более дней и рассматриваем все доходности как однородные.

Итоговая выбока состоит из 3952 наблюдений для каждого из 6 рядов доходностей. Построенную выборку разделяем на две части: первые 2635 наблюдений используем для оценивания моделей («внутривыборочные» наблюдения), а последние 1317 используем для оценки предсказательной способности моделей («вневыборочные» наблюдения). Таким образом, соотношение числа наблюдений внутри и вне выборки составляет 2:1. Причина такого выбора (внутри выборки наблюдений гораздо больше, чем вне ее) в том, что при меньшем числе наблюдений внутри выборки в некоторых из наших функций правдоподобия возникают

⁸ Используется база данных Yahoo Finance <http://finance.yahoo.com>.

большие осцилляции, и это значительно ухудшает сходимость оптимизационных процедур при оценивании.

В таблице 2 представлены описательные статистики рассматриваемых рядов доходностей. Распределения всех доходностей имеют тяжелые хвосты (лептокуртозис) и отрицательную скошенность. Это два широко известных стилизованных факта о маргинальном распределении финансовых доходностей. Доходность индекса Hang-Seng имеет наибольшее стандартное отклонение и куртозис – этот индекс наиболее волатилен среди рассматриваемых. Вероятно, по этой причине мы столкнулись с особенными трудностями со сходимостью оптимизационных процедур для моделей, в которые входит Hang-Seng.⁹

Таблица 2. Описательные статистики для рядов логарифмических доходностей

Статистика	S&P500	FTSE100	CAC40	DAX	HANGSENG	NIKKEI225
Среднее	0,035	0,024	0,026	0,038	0,050	-0,015
Медиана	0,056	0,038	0,030	0,088	0,058	-0,011
Максимум	5,574	5,904	7,002	8,754	17,247	7,826
Минимум	-7,113	-8,565	-8,775	-10,358	-14,735	-9,754
Ст. откл.	1,059	1,096	1,388	1,473	1,715	1,499
Скошенность	-0,059	-0,099	-0,061	-0,217	-0,180	-0,007
Куртозис	6,328	6,636	6,081	7,725	12,629	5,643
Наблюдений	3952	3952	3952	3952	3952	3952

5.2 Корреляция

В рассматриваемых рядах логарифмических доходностей нет значимой связности. Гипотеза о постоянном нулевом условном среднем в простой AR(1) модели с константой не отвергается на 5% уровне значимости тестом Вальда для всех 6 индексов, как показано в таблице 3.

Таблица 3. Тест Вальда на постоянное нулевое условное среднее в моделях AR(1)

	S&P500	FTSE100	CAC40	DAX	HANGSENG	NIKKEI225
Статистика Вальда	4,689	2,971	1,845	2,695	3,356	1,079
(P-значение)	0,096	0,226	0,397	0,260	0,187	0,583

Таким образом, в доходностях нет значимой автокорреляции. Однако существует значимая корреляция между доходностями различных индексов. В таблице 4 приведены одновременные и лагированные корреляции доходностей рассматриваемых индексов. Корреляции в этой таблице можно объяснить географией рынков (временными зонами). Для рассматриваемых индексов время торговли определено (приблизительно) следующим образом. По московскому времени рынки в Японии и Гонконге открываются в 3:00 и закрываются в 11:00. Европейские рынки открываются в 9:00 и закрываются в 17:00. Американский рынок открывается в 16:00 и закрывается в 00:00. Динамика на рынках, на которых торгуют раньше, является информацией для тех рынков, на которых торгуют позже. Отсюда существенная положительная корреляция между лагированной доходностью S&P500 и текущей доходностью NIKKEI225 и Hang-Seng. По той же причине лагированная доходность S&P500 положительно коррелирует с текущей доходностью европейских индексов FTSE100, CAC40 и DAX. Корреляции в этом случае ниже, чем для NIKKEI225 и Hang-Seng, поскольку с момента закрытия американского рынка до момента открытия европейских рынков проходит больше времени и поступает больше информации. Аналогичные связи можно наблюдать для временных пар Восточная Азия–Европа и Европа–США.

⁹ Эта проблема была решена, когда мы стали брать оценки параметров простых вложенных моделей в качестве стартовых значений при оценке более сложных моделей.

Таблица 4. Одновременные и лагированные корреляции доходностей

SP&500(-1)	1	-0,008	0,467	0,267	0,482	0,256	0,510	0,224	0,157	0,359	0,148	0,316
SP&500	1	0,001	0,468	0,468	0,013	0,487	-0,001	0,511	-0,017	0,156	-0,028	0,148
FTSE100(-1)	1	-0,016	0,795	0,001	0,795	-0,006	0,713	0,016	0,349	0,209	0,305	0,1866
FTSE100	1	0,001	0,001	0,001	0,001	0,795	0,011	0,713	-0,025	0,349	-0,056	0,305
CAC40(-1)	1	0,011	0,011	0,011	0,011	0,011	0,803	0,035	0,330	0,182	0,306	0,202
CAC40	1	0,026	0,026	0,026	0,026	0,026	0,803	0,803	-0,023	0,330	-0,046	0,306
DAX(-1)	1	-0,002	0,344	0,344	0,344	0,344	0,162	-0,002	0,344	0,162	0,285	0,204
DAX	1	0,008	0,008	0,008	0,008	0,008	0,344	1	-0,008	0,344	-0,035	0,285
HANGSENG(-1)	1	0,003	0,420	0,420	0,420	0,420	-0,003	1	1	-0,003	0,420	0,010
HANGSENG	1	-0,031	-0,031	-0,031	-0,031	-0,031	1	1	1	1	-0,031	0,420
NIKKEI225(-1)	1	0,013	0,013	0,013	0,013	0,013	1	1	1	1	1	-0,013
NIKKEI225	1	0,013	0,013	0,013	0,013	0,013	1	1	1	1	1	1

(-1) означает лагированную доходность

Среди шести рассматриваемых нами индексов S&P500 приносит на рынок наиболее важную информацию. В то же время доходность на этот индекс не зависит существенным образом от информации о реализовавшихся доходностях на других рынках. Это продемонстрировано в таблице 5, в которой приведены оценки VAR(1)-модели для вектора доходностей всех шести рассматриваемых индексов.

Таблица 5. Оценки векторной авторегрессии для всех шести индексов

	S&P500	FTSE100	CAC40	DAX	HANGSENG	NIKKEI225
S&P500(-1)	-0,014 (0,019)	0,387 (0,018)	0,454 (0,024)	0,424 (0,025)	0,570 (0,028)	0,391 (0,025)
FTSE100(-1)	-0,010 (0,026)	-0,109 (0,026)	-0,133 (0,033)	-0,088 (0,035)	0,213 (0,039)	0,031 (0,035)
CAC40(-1)	0,043 (0,024)	-0,016 (0,024)	-0,022 (0,030)	0,096 (0,033)	0,017 (0,036)	0,064 (0,032)
DAX(-1)	-0,015 (0,020)	-0,054 (0,020)	-0,043 (0,025)	-0,177 (0,027)	-0,099 (0,030)	0,035 (0,027)
HANGSENG(-1)	-0,007 (0,011)	0,006 (0,011)	0,001 (0,014)	0,014 (0,015)	-0,045 (0,017)	-0,038 (0,015)
NIKKEI225(-1)	-0,021 (0,012)	-0,040 (0,012)	-0,043 (0,016)	-0,044 (0,017)	-0,098 (0,019)	-0,071 (0,017)
c	0,036 (0,017)	0,015 (0,017)	0,014 (0,021)	0,028 (0,023)	0,028 (0,025)	-0,032 (0,023)

В скобках приведены QML-стандартные ошибки. Жирным показана значимость на 5% уровне.

Согласно оцененной векторной авторегрессии, лагированная доходность S&P500 имеет большое положительное влияние на доходности европейских и азиатских индексов (коэффициенты в первой строке таблицы), но в обратную сторону такого влияния нет (коэффициенты в первом столбце). Таким образом, лагированная доходность на S&P500 – особенная переменная, которая может рассматриваться как предиктор текущих доходностей на других рынках. Предполагается, что отмеченные динамические связи между доходностями индексов будут учтены моделью VAR(1) для условного среднего.

5.3 Волатильность

На Рис. 1 изображена динамика логарифмических доходностей рассматриваемых шести индексов.

Очевидна кластеризация волатильности, которая, как правило, наблюдается в рядах финансовых доходностей. Это также очевидно и из стандартных GARCH(1,1) моделей, результаты оценивания которых представлены в таблице 6.

Таблица 6. Оценки моделей GARCH(1,1)

Параметр	S&P500	FTSE100	CAC40
β_0	0,005 (0,003)	0,013 (0,004)	0,032 (0,012)
β_1	0,042 (0,009)	0,079 (0,011)	0,068 (0,013)
β_2	0,954 (0,011)	0,912 (0,012)	0,915 (0,016)
Параметр	DAX	HANGSENG	NIKKEI225
β_0	0,034 (0,013)	0,037 (0,019)	0,056 (0,020)
β_1	0,075 (0,013)	0,075 (0,018)	0,083 (0,015)
β_2	0,908 (0,015)	0,913 (0,021)	0,894 (0,019)

В скобках приведены QML-стандартные ошибки.

Жирным показана значимость на 5% уровне.

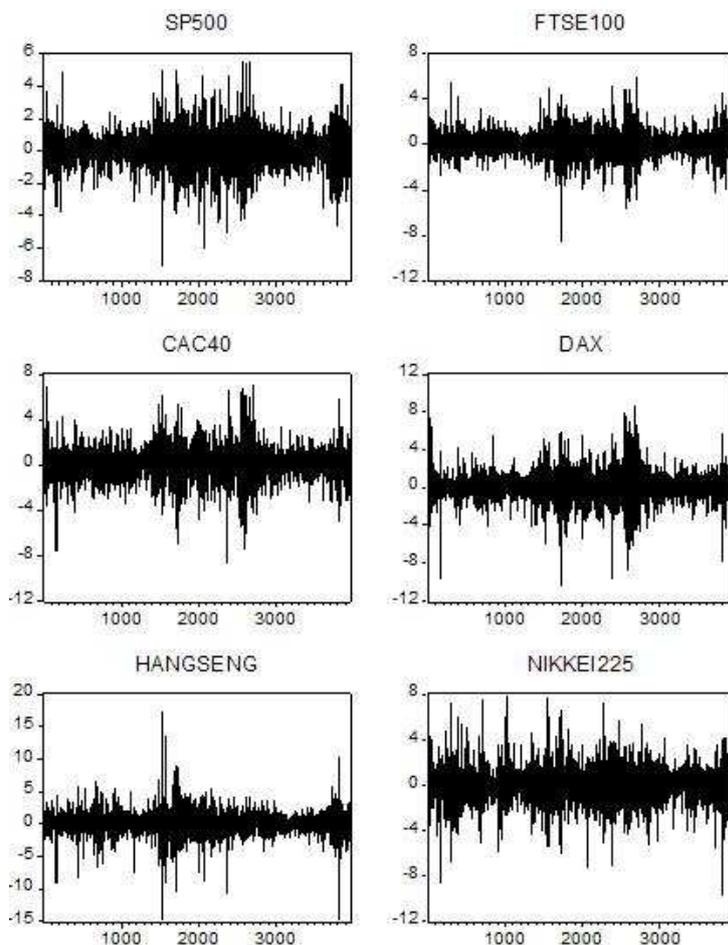


Рис. 1: Динамика логарифмических доходностей фондовых индексов

Из Рис. 1 можно также заметить, что кластеры волатильности на одном рынке, как правило, соответствуют кластерам волатильности на других рынках (например, все индексы имеют кластеры высокой волатильности в середине выборки). Таким образом, волатильности доходностей различных рынков могут быть связаны. Предполагается, что MGARCH-модель для условной ковариационной матрицы будет учитывать как индивидуальные GARCH - эффекты, так и возможную связь между волатильностями на различных рынках.

5.4 Условная скошенность и куртозис

Источником отрицательной безусловной скошенности и лептокуртозиса (тяжелых хвостов) могут быть их условные аналоги.¹⁰ Чтобы предварительно оценить эти величины мы проводим непараметрическую оценку одномерных условных функций плотности для рассматриваемых рядов логарифмических доходностей и по полученным оценкам функций плотности рассчитываем условную скошенность и куртозис.

Предположим, что ряд доходностей r_t является строго стационарным, так что неизвестные совместные функции плотности $f(r_t, r_{t-1})$ и маргинальные функции плотности $f(r_t)$ не зависят от времени. Тогда условная функция плотности $f(r_t|r_{t-1})$ не зависит от времени и может быть оценена непараметрически (для простоты включаем в доступную информацию I_{t-1} только r_{t-1}).

¹⁰ Источником лептокуртозиса является также переменная волатильность.

Мы используем одношаговую непараметрическую оценку условной функции плотности, основанную на стандартных ядерных оценках. Пусть h_1 и h_2 – это ширина окон для r_t и r_{t-1} соответственно, и $K_h(u) = h^{-1}K(u/h)$ для некоторой функции ядра $K(x)$. Стандартные ядерные оценки Надарая–Уотсона для совместной и маргинальной функций плотности и получаемая из них одношаговая оценка условной функции плотности имеют вид

$$\widehat{f}(y|x) = \frac{\widehat{f}(y, x)}{\widehat{f}(x)},$$

где

$$\widehat{f}(y, x) = \frac{1}{n} \sum_{t=2}^T K_{h_1}(y - r_t) K_{h_2}(x - r_{t-1}),$$

$$\widehat{f}(x) = \frac{1}{n} \sum_{t=2}^T K_{h_2}(x - r_{t-1}).$$

Существует также и двухшаговая непараметрическая оценка условной функции плотности, основанная на оценке функции условного среднего. Эта оценка имеет меньшее асимптотическое смещение, чем одношаговая оценка, и в общем случае она предпочтительнее.¹¹ Но мы строим непараметрические оценки условных функций плотности в целях получения предварительной оценки условной скошенности и куртозиса, и на наш взгляд, одношаговой оценки для этих целей вполне достаточно.

Итак, с помощью одношаговой процедуры мы строим оценки условных функций плотности логарифмических доходностей.¹² Ширина окна вычисляется по правилу Сильвермана $h = 1.06\sigma_r T^{-1/5}$. После оценивания каждой из условных функций плотности мы приводим ее к нулевому среднему (которое само по себе близко к нулю) и единичной дисперсии (используя оцененное среднее и дисперсию). Полученные оценки функций плотности стандартизованных ошибок при условии $\eta_{t-1} = 0$ представлены на Рис. 2 вместе с функцией плотности стандартного нормального распределения.

Для оцененных условных функций плотности характерен лептокуртозис, но относительно отрицательной скошенности однозначных выводов сделать нельзя. В таблице 7 показаны условные скошенности и куртозисы, вычисленные по непараметрическим оценкам функций плотности. Заметим, что оценки функций плотности для FTSE100 и NIKKEI225 предполагают небольшую положительную скошенность, и этот результат устойчив к изменению ширины окна при оценивании. Для индекса Hang-Seng характерен относительно небольшой условный куртозис по сравнению с безусловным.

Таблица 7. Оценки условной скошенности и куртозиса

Статистика	S&P500	FTSE100	CAC40	DAX	HANGSENG	NIKKEI225
Скошенность	-0,510	0,014	-0,253	-0,254	-0,052	0,092
Куртозис	5,301	4,278	3,859	4,266	3,786	3,720

Предполагается, что условная скошенность и куртозис в распределении стандартизованных ошибок будут учтены с помощью рассмотренных в разделе 3 гибких параметризаций, которые предоставляют возможность управлять асимметрией и толщиной хвостов распределения.

¹¹ Описание данной оценки можно найти, например, в Hansen (2004).

¹² Используется пакет pr в языке программирования R.

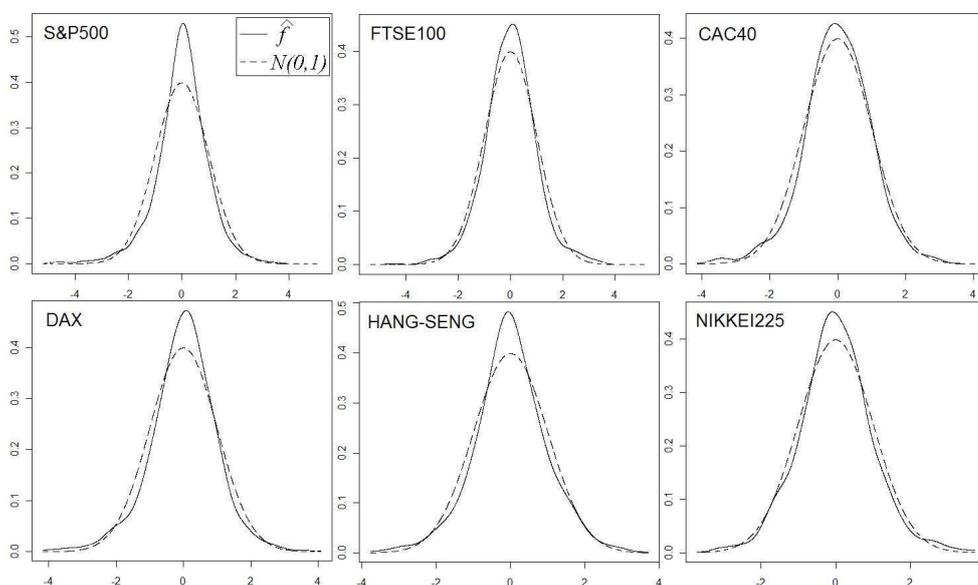


Рис. 2: Нормализованные непараметрические оценки условных функций плотности и функция плотности распределения $N(0, 1)$

6 Результаты оценивания

В предыдущем разделе было отмечено, что доходности на различных рынках динамически связаны, их волатильности также динамически связаны, а условное распределение доходностей асимметрично и имеет тяжелые хвосты. Чтобы учесть все эти свойства данных, мы строим двумерные модели VAR(1)–BEKK(1,1) с функциями плотности распределения стандартизованных ошибок, рассмотренными в разделе 3, которые допускают скошенность и лептокуртозис. Каждая из этих моделей описывается уравнениями (1)–(2) и одной из функций плотности (4), (5), (7) или (8). Имеется 15 пар фондовых индексов и 4 параметризации функции плотности распределения стандартизованных ошибок – в общей сложности оценено 60 моделей.

Прежде чем анализировать полученные оценки параметров, сделаем два замечания относительно технической процедуры оценивания. Во-первых, как уже было упомянуто, функция плотности Грамма–Шарлье не является функцией плотности в соответствии со строгим определением. Для некоторых значений элементов матриц D_3 и D_4 она, как функция аргумента η_t , может иметь области отрицательных значений. Это значительно усложняет максимизацию функции правдоподобия для параметризации Грамма–Шарлье по сравнению с другими параметризациями. Вероятная причина этого, на наш взгляд, в том, что граница области положительной определенности двумерной плотности Грамма–Шарлье такова, что градиент соответствующей функции правдоподобия очень нестабилен вблизи нее. Поэтому при оценивании параметров функции плотности Грамма–Шарлье следует уделять особое внимание выбору вектора начальных значений параметров, чтобы избежать раннего прерывания оптимизационной процедуры.¹³

Во-вторых, поскольку количество параметров в рассматриваемых моделях различно, можно оценить, как это количество влияет на время, необходимое для сходимости процедуры максимизации функции правдоподобия. Этот вопрос имеет практическое значение, поскольку в практических расчетах число активов может быть довольно большим, что приводит к большому числу параметров. В нашем случае время, необходимое для максимизации функ-

¹³ Мы используем процедуру оптимизации из библиотеки CML для языка GAUSS с алгоритмом спуска ВННН и методом поиска на линии STEPBT.

ции правдоподобия с параметризацией Грамма–Шарлье, имеющей 26 параметров, оказывается примерно в 4 раза большим, чем соответствующее время для скошенного нормального распределения, скошенного распределения Стьюдента и скошенного распределения GED (19, 20 и 20 параметров соответственно). Таким образом, для практического применения функция плотности Грамма–Шарлье, вероятно, имеет слишком много параметров. Однако это лишь предварительный вывод. Чтобы полноценно сравнить время оценивания моделей, необходимо рассматривать случай большого числа активов, а не только простой случай $N = 2$.

В таблицах 8–11 для пар S&P500–NIKKEI225 и S&P500–HANGSENG представлены результаты оценивания методом максимального квазиправдоподобия моделей VAR(1)–BEKK(1,1) с 4-мя различными параметризациями функции плотности стандартизованных ошибок. Оценивание для остальных пар индексов дает схожие результаты. Все оцененные модели имеют похожие оценки параметров в части условного среднего и ковариационной матрицы, вне зависимости от того, какую параметризацию для функции плотности распределения стандартизованных ошибок они используют. Например, для моделей, где присутствует индекс S&P500, важный коэффициент Q_{21} , отражающий влияние лагированной доходности S&P500 на текущую доходность другого индекса, всегда положителен, значим на 5% уровне значимости и имеет приблизительно одинаковые значения для различных параметризаций плотности. В уравнении динамики условной ковариационной матрицы коэффициенты A_{11} , A_{22} , B_{11} и B_{22} практически не зависят от используемой параметризации функции плотности. Другие параметры в уравнениях условного среднего и ковариационной матрицы более чувствительны к выбору параметризации: их P-значения, а иногда и знак существенно зависят от выбора параметризации.

Таблица 8. Оценки VAR(1)–BEKK(1,1) модели со скошенным нормальным распределением для стандартизованных ошибок

S&P500–NIKKEI225				S&P500–HANGSENG			
c		Q		c		Q	
0,146 (0,077)	0,049 (0,021)	–0,002 (0,011)		0,135 (0,081)	0,041 (0,021)	–0,009 (0,011)	
0,252 (0,103)	0,406 (0,027)	–0,046 (0,021)		0,299 (0,161)	0,522 (0,033)	–0,002 (0,020)	
Ω				Ω			
0,041 (0,030)				0,033 (0,048)			
–0,033 (0,127)	0,242 (0,079)			0,140 (0,305)	0,315 (0,087)		
A		B		A		B	
0,160 (0,046)	–0,004 (0,015)	0,986 (0,008)	0,001 (0,004)	0,159 (0,053)	0,018 (0,019)	0,988 (0,008)	–0,009 (0,010)
–0,070 (0,044)	0,270 (0,042)	0,017 (0,009)	0,949 (0,016)	–0,097 (0,067)	0,340 (0,050)	0,029 (0,012)	0,917 (0,027)
α				α			
–0,111 (0,121)				–0,103 (0,119)			
–0,280 (0,110)				–0,230 (0,155)			

В скобках приведены QML-стандартные ошибки. Жирным выделены коэффициенты, значимо отличные от 0 или релевантного значения на 5% уровне значимости.

Таблица 9. Оценки VAR(1)–BEKK(1,1) модели со скошенным распределением Стьюдента для стандартизованных ошибок

S&P500–NIKKEI225				S&P500–HANGSENG			
c		Q		c		Q	
0,117 (0,043)		0,032 (0,019)	–0,006 (0,010)	0,118 (0,044)		0,024 (0,019)	–0,004 (0,009)
0,076 (0,070)		0,405 (0,025)	–0,048 (0,018)	0,156 (0,068)		0,501 (0,029)	0,002 (0,019)
Ω				Ω			
0,031 (0,021)				0,034 (0,020)			
–0,062 (0,101)	0,180 (0,058)			–0,032 (0,087)	0,207 (0,052)		
A		B		A		B	
0,136 (0,030)	0,001 (0,010)	0,986 (0,006)	–0,001 (0,004)	0,142 (0,026)	0,005 (0,009)	0,985 (0,005)	–0,002 (0,004)
–0,029 (0,044)	0,227 (0,023)	0,013 (0,010)	0,951 (0,010)	–0,029 (0,046)	0,265 (0,040)	0,014 (0,011)	0,938 (0,019)
ξ				ξ			
–0,092 (0,078)				–0,099 (0,077)			
–0,135 (0,080)				–0,100 (0,074)			
ν				ν			
6,671 (0,585)				6,659 (0,605)			

Таблица 10. Оценки VAR(1)–BEKK(1,1) модели со скошенным распределением GED для стандартизованных ошибок

S&P500–NIKKEI225				S&P500–HANGSENG			
c		Q		c		Q	
0,099 (0,032)		0,023 (0,020)	–0,007 (0,010)	0,096 (0,032)		0,012 (0,020)	–0,003 (0,010)
–0,017 (0,043)		0,410 (0,026)	–0,052 (0,018)	0,106 (0,046)		0,505 (0,030)	–0,001 (0,019)
Ω				Ω			
0,036 (0,020)				0,034 (0,022)			
–0,046 (0,097)	0,206 (0,056)			–0,011 (0,120)	0,250 (0,058)		
A		B		A		B	
0,141 (0,030)	–0,001 (0,010)	0,986 (0,006)	–0,000 (0,003)	0,146 (0,026)	0,008 (0,010)	0,986 (0,005)	–0,004 (0,004)
–0,025 (0,042)	0,235 (0,026)	0,012 (0,009)	0,950 (0,011)	–0,044 (0,048)	0,284 (0,042)	0,019 (0,011)	0,931 (0,021)
ξ				ξ			
0,956 (0,024)				0,957 (0,024)			
0,982 (0,022)				0,967 (0,023)			
κ				κ			
0,805 (0,030)				0,814 (0,034)			

В скобках приведены QML-стандартные ошибки. Жирным выделены коэффициенты, значимо отличные от 0 или релевантного значения на 5% уровне значимости.

Таблица 11. Оценки VAR(1)–BEKK(1,1) модели с распределением Грамма–Шарлье для стандартизованных ошибок

S&P500–NIKKEI225				S&P500–HANGSENG			
c		Q		c		Q	
0,052 (0,017)		0,031 (0,019)	–0,004 (0,010)	0,050 (0,017)		0,027 (0,020)	–0,008 (0,009)
–0,039 (0,026)		0,408 (0,025)	–0,052 (0,019)	0,036 (0,029)		0,514 (0,031)	–0,007 (0,020)
Ω				Ω			
0,034 (0,026)				0,025 (0,035)			
–0,134 (0,140)	0,207 (0,119)			0,134 (0,327)	0,313 (0,089)		
A		B		A		B	
0,156 (0,039)	–0,006 (0,011)	0,988 (0,006)	0,002 (0,003)	0,153 (0,041)	0,011 (0,015)	0,990 (0,005)	–0,006 (0,007)
–0,059 (0,040)	0,250 (0,032)	0,016 (0,008)	0,954 (0,012)	–0,043 (0,070)	0,325 (0,061)	0,020 (0,011)	0,925 (0,030)
D_3				D_3			
–0,078 (0,075)	–0,012 (0,034)	–0,012 (0,034)	0,014 (0,033)	–0,084 (0,074)	–0,071 (0,034)	–0,071 (0,034)	0,062 (0,033)
–0,012 (0,034)	0,014 (0,033)	0,014 (0,033)	0,056 (0,067)	–0,071 (0,034)	0,062 (0,033)	0,062 (0,033)	–0,082 (0,073)
D_4				D_4			
4,208 (0,176)	–0,025 (0,076)	–0,025 (0,076)	1,183 (0,057)	4,193 (0,168)	–0,031 (0,073)	–0,031 (0,073)	1,179 (0,063)
–0,025 (0,076)	1,183 (0,057)	1,183 (0,057)	0,028 (0,078)	–0,031 (0,073)	1,179 (0,063)	1,179 (0,063)	0,132 (0,084)
–0,025 (0,076)	1,183 (0,057)	1,183 (0,057)	0,028 (0,078)	–0,031 (0,073)	1,179 (0,063)	1,179 (0,063)	0,132 (0,084)
1,183 (0,057)	0,028 (0,078)	0,028 (0,078)	4,031 (0,163)	1,179 (0,063)	0,132 (0,084)	0,132 (0,084)	4,082 (0,171)

В скобках приведены QML-стандартные ошибки. Жирным выделены коэффициенты, значимо отличные от 0 или релевантного значения на 5% уровне значимости.

Таблицы 8–11 показывают, что все оцененные условные функции плотности имеют отрицательную скошенность вдоль обеих осей. Для параметризации со скошенным нормальным распределением (таблица 8) оценки обеих компонент вектора скошенности α отрицательны, так же как и для вектора ξ в параметризации со скошенным распределением Стьюдента (таблица 9). Модель со скошенным распределением GED (таблица 10) также характеризуется отрицательной скошенностью, поскольку обе компоненты вектора ξ в ней меньше единицы. Наконец, параметризация с распределением Грамма–Шарлье (таблица 11) имеет отрицательные оценки для некоторых третьих моментов в матрице D_3 , что также означает наличие отрицательной скошенности.

Модель со скошенным нормальным распределением не может учесть толщину многомерных хвостов распределения, поскольку в ней отсутствуют соответствующие параметры, но, как видно из таблиц 9–11, остальные модели указывают на наличие многомерных тяжелых хвостов. Оценка параметра степеней свободы ν в модели со скошенным распределением Стьюдента невелика, что означает наличие тяжелых многомерных хвостов. Модель со скошенным распределением GED дает оценку параметра κ , меньшую единицы, что также указывает на тяжелые многомерные хвосты. Наконец, параметризация Грамма–Шарлье дает такую оценку матрицы D_4 , которая также подразумевает тяжелые многомерные хвосты: элементы этой матрицы, равные 1 и 3 в случае нормальности, статистически значимо превышают соответственно 1 и 3 в оцененной модели с распределением Грамма–Шарлье.

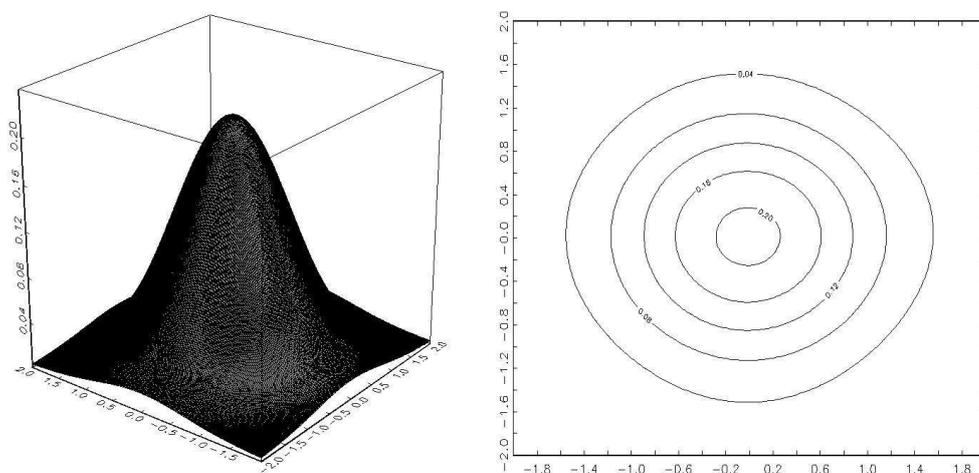


Рис. 3: Поверхность и линии уровня оцененной функции плотности распределения Грамма–Шарлье для стандартизованных ошибок (пара S&P500–NIKKEI225)

Все четыре оцененные функции плотности визуально похожи. На Рис. 3 для пары S&P500–NIKKEI225 изображены поверхность и линии уровня оцененной условной функции плотности распределения вектора стандартизованных ошибок с параметризацией Грамма–Шарлье, при условии $\eta_{t-1} = 0$. Существенной асимметрии функции плотности на Рис. 3 не наблюдается.

7 Результаты сравнения

В этом разделе приведены результаты KLIC-теста для всех описанных параметризаций и пар фондовых индексов. Тестирование проведено отдельно внутри и вне выборки. В таблице 12 приведены результаты KLIC-теста внутри выборки оценивания, а в таблице 13 – результаты теста вне ее.

Внутривыборочный KLIC-тест сопоставляет качество подгонки двух моделей. Вневыборочный KLIC-тест сопоставляет качество прогнозов условной функции плотности, полученных на основе двух моделей. Для теста ‘M1 vs M2’ («Модель 1 против Модели 2») KLIC-статистика вычисляется как разница значений функций правдоподобия Модели 1 и Модели 2. Поэтому на 5% уровне значимости выполнение условия $KLIC > 1,96$ означает, что Модель 1 предпочтительнее Модели 2, а условие $KLIC < -1,96$ означает, что предпочтительнее Модель 2. Если же $-1,96 < KLIC < 1,96$, то Модели 1 и 2 признаются эквивалентными на 5% уровне значимости.

Таблицы 12 и 13 показывают, что сравнение оцененных нами моделей в большинстве случаев однозначно и почти не зависит от рассматриваемой пары фондовых индексов.

Модель со скошенным нормальным распределением для стандартизованных ошибок доминируется всеми остальными моделями как внутри, так и вне выборки оценивания. Этот результат ожидаем, поскольку данная параметризация не учитывает важную характеристику распределения доходностей – тяжелые хвосты, а остальные параметризации ее учитывают.

Модель со скошенным распределением Стьюдента доминирует модель со скошенным распределением GED внутри выборки оценивания, а вне ее эти модели, в целом, эквивалентны. Соответствующие функции плотности имеют одинаковое число параметров и строятся по одному и тому же принципу: в обеих есть вектор направлений скошенности и параметр, контролирующий величину вероятностной массы в центре распределения. Однако данные функции плотности отличаются тем, каким образом вектор направлений скошенности и параметр тяжести хвостов в них входят. Судя по результатам сравнений, параметризация со скошен-

ным распределением Стьюдента является более гибкой, чем параметризация со скошенным распределением GED, то есть позволяет модели лучше соответствовать данным внутри выборки оценивания. Однако при этом данная гибкость не дает существенного выигрыша в точности предсказаний условной функции плотности вне выборки оценивания, и поэтому по предсказательной способности модели со скошенными распределениями Стьюдента и GED эквивалентны.

Таблица 12. Результаты KLIC-теста внутри выборки оценивания (2635 наблюдений)

Пара индексов	SN-SS	SN-SG	SN-GC	SS-SG	SS-GC	SG-GC
S&P500-FTSE100	-4,893	-4,586	-5,107	3,118	1,127	-0,745
S&P500-CAC40	-4,219	-4,287	-4,458	2,065	2,212	1,806
S&P500-DAX	-3,672	-3,727	-5,083	2,808	1,946	1,343
S&P500-HANGSENG	-4,204	-4,285	-5,381	2,780	2,186	1,702
S&P500-NIKKEI225	-5,199	-4,996	-5,272	3,663	3,166	2,046
FTSE100-CAC40	-3,692	-3,545	-3,866	2,954	2,690	1,953
FTSE100-DAX	-3,040	-2,995	-3,822	2,737	1,981	1,508
FTSE100-HANGSENG	-4,131	-3,992	-4,478	3,074	2,345	1,462
FTSE100-NIKKEI225	-4,424	-4,157	-3,232	2,983	3,450	2,840
CAC40-DAX	-2,719	-2,650	-3,606	2,589	1,740	1,165
CAC40-HANGSENG	-3,263	-3,363	-4,030	2,154	1,800	1,515
CAC40-NIKKEI225	-3,479	-3,416	-3,589	2,531	2,100	1,584
DAX-HANGSENG	-3,658	-3,790	-2,963	2,453	2,837	2,815
DAX-NIKKEI225	-3,453	-3,556	-4,868	2,513	1,944	1,632
HANGSENG-NIKKEI225	-4,100	-3,831	-4,127	3,744	2,915	2,215

SN – скош. нормальное, SS – скош. Стьюдента, SG – скош. GED, GC – Грамма-Шарлье

‘M1 vs M2’: $KLIC > 1,96 \Rightarrow M1$ лучше, $KLIC < -1,96 \Rightarrow M2$ лучше (5% уровень значимости)

Таблица 13. Результаты KLIC-теста вне выборки оценивания (1317 наблюдений)

Пара индексов	SN-SS	SN-SG	SN-GC	SS-SG	SS-GC	SG-GC
S&P500-FTSE100	-3.142	-3.533	-3.062	0.609	2.083	2.445
S&P500-CAC40	-4.303	-5.021	-3.141	-0.722	3.702	4.861
S&P500-DAX	-3.963	-4.924	-3.175	-1.191	3.505	5.091
S&P500-HANGSENG	-4.199	-4.539	-3.133	0.509	4.029	4.638
S&P500-NIKKEI225	-3.151	-3.082	-1.832	1.706	3.655	3.636
FTSE100-CAC40	-14.483	-11.224	-7.004	4.316	18.521	10.588
FTSE100-DAX	-8.103	-8.432	-8.028	1.400	4.684	4.966
FTSE100-HANGSENG	-5.123	-4.970	-4.257	2.186	3.676	3.397
FTSE100-NIKKEI225	-3.771	-3.451	-2.210	2.599	3.789	3.204
CAC40-DAX	-4.853	-6.130	-5.774	-2.736	1.721	3.320
CAC40-HANGSENG	-4.033	-4.133	-3.469	1.738	3.128	3.235
CAC40-NIKKEI225	-4.049	-3.293	-2.306	3.421	4.108	2.945
DAX-HANGSENG	-3.612	-4.098	-2.106	0.381	3.554	4.286
DAX-NIKKEI225	-2.963	-2.993	-1.569	1.170	3.072	3.271
HANGSENG-NIKKEI225	-5.712	-5.826	-3.300	0.645	5.773	6.511

SN – скош. нормальное, SS – скош. Стьюдента, SG – скош. GED, GC – Грамма-Шарлье

‘M1 vs M2’: $KLIC > 1,96 \Rightarrow M1$ лучше, $KLIC < -1,96 \Rightarrow M2$ лучше (5% уровень значимости)

Примечательно, что модель с распределением Грамма-Шарлье доминируется вне выборки оценивания моделями со скошенными распределениями Стьюдента и GED. Этот результат неожидан, поскольку функция плотности Грамма-Шарлье имеет существенно больше параметров, чем функции плотности скошенных распределений Стьюдента и GED. Больше число параметров дает большую гибкость распределению Грамма-Шарлье, и тем не

менее даже внутри выборки оно оказывается лишь эквивалентным скошенному распределению GED, а скошенному распределению Стьюдента даже проигрывает. На наш взгляд, можно предложить следующее объяснение этим результатам. Вспомним, что для некоторых значений параметров функция плотности Грамма–Шарлье теряет положительную определенность. Далее вспомним, что параметры функции плотности Грамма–Шарлье в наших моделях соответствуют нормализованным условным третьим и четвертым моментам двумерного распределения доходностей. Вероятно, истинные условные моменты распределения доходностей таковы, что функция плотности Грамма–Шарлье не является положительно определенной, если в качестве значений ее параметров подставить эти моменты (в предположении, что они постоянны). Если это так, то процедура максимизации функции правдоподобия выберет матрицы D_3 и D_4 , заведомо отличающиеся от истинных матриц третьих и четвертых моментов.¹⁴ Это и может быть причиной низкого качества подгонки и предсказательной способности модели с распределением Грамма–Шарлье. К сожалению, мы не можем явным образом проверить данное предположение, поскольку истинные третьи и четвертые условные моменты распределения доходностей ненаблюдаемы.

Таким образом, на основе проведенного с помощью KLIC-теста попарного сравнения рассмотренных моделей можно заключить, что скошенное распределение Стьюдента дает наилучшее качество подгонки, и так же как и скошенное распределение GED – наилучшую предсказательную способность. Распределение на основе разложения Грамма–Шарлье, что неожиданно, показывает более низкое качество подгонки и предсказательной способности, особенно в сравнении со скошенным распределением Стьюдента. Наконец, скошенное нормальное распределение, что вполне ожидаемо, дает самое низкое качество подгонки и предсказательную способность среди рассмотренных распределений.

В заключение отметим, что глядя на данные результаты, необходимо помнить, что KLIC-тест может быть некорректным, если дисперсия ω^* логарифмической разности сравниваемых функций плотности близка к нулю, о чем упоминалось в разделе 4. В данной работе мы не проводим предварительного тестирования гипотезы $\omega^* = 0$, и наши результаты надежны настолько, насколько данная гипотеза не соответствует действительности.

8 Выводы

В данной работе произведена оценка двумерных моделей VAR(1)–BEKK(1,1) для логарифмических доходностей индексов крупнейших мировых фондовых рынков. Для моделирования условной функции плотности стандартизованных ошибок были использованы 4 спецификации: скошенное нормальное распределение, скошенное распределение Стьюдента, скошенное распределение GED и распределение Грамма–Шарлье.

Для простоты процедуры сравнения моделей предполагалось постоянство параметров распределения стандартизованных ошибок. Используя тест, основанный на информационном критерии Кульбака–Лейблера, мы провели попарное сравнение оцененных моделей. В результате сравнений получен следующий ранжированный список распределений стандартизованных ошибок. По качеству подгонки модели внутри выборки оценивания: 1 – скошенное распределение Стьюдента, 2 – скошенное распределение GED и распределение Грамма–Шарлье, 3 – скошенное нормальное распределение. По предсказательной способности модели вне выборки оценивания: 1 – скошенное распределение Стьюдента и скошенное распределение GED, 2 – распределение Грамма–Шарлье, 3 – скошенное нормальное распределение.

Предположение о постоянных параметрах распределения стандартизованных ошибок упрощает сравнение функций плотности. Однако не исключено, что при этом оцененные нами модели могут иметь достаточно низкое качество подгонки и предсказательной способности

¹⁴ Процедурой будут рассматриваться только такие матрицы D_3 и D_4 , для которых функция плотности Грамма–Шарлье является положительно определенной.

в некотором абсолютном смысле, например, с точки зрения распределения их обобщенных остатков.¹⁵ Поэтому важное направление дальнейшей работы – оценка качества рассмотренных моделей именно в абсолютном смысле.

Другое возможное продолжение данной работы – отказаться от постоянства параметров распределения стандартизованных ошибок и рассмотреть различные модели изменения этих параметров, а затем для каждой из них сравнить соответствующие функции плотности распределения доходностей. Учет изменения параметров во времени может улучшить качество данных моделей в абсолютном смысле.

Еще одним из направлений дальнейших исследований может быть поиск возможностей улучшения качества моделей за счет смешения различных функций плотности. Например, интерес может представлять нахождение функции плотности, которая при смешении с функцией плотности скошенного распределения Стьюдента улучшает качество подгонки и предсказательную способность соответствующей модели.

9 Благодарности

Автор выражает благодарность своему научному руководителю, профессору РЭШ Станиславу Анатольеву за ценные замечания, в особенности, за идею расширить работу с многомерным распределением Грамма–Шарлье и рассмотреть другие конкурирующие параметризации для условной функции плотности распределения доходностей. Автор также благодарен своим коллегам по исследованиям в РЭШ – Юрию Бедному, Максиму Спиряеву и Родиону Ломиворотову – за плодотворные обсуждения. Наконец, автор благодарит своего сокурсника Александра Степанова за полезные советы о библиотеках языка программирования R. Автор несет ответственность за все возможные ошибки и неточности.

Приложение

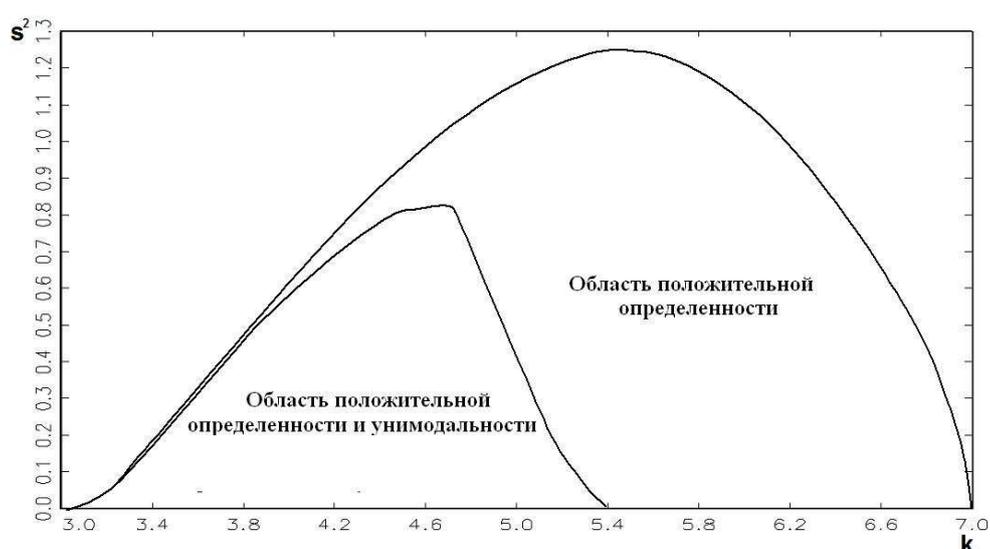


Рис. 4: Область положительной определенности для одномерной функции плотности распределения Грамма–Шарлье

¹⁵ Один из первых методов оценки качества модели на основе анализа распределения обобщенных остатков предложен в Diebold, Gunther & Tay (1998).

Список литературы

- Azzalini, A. & A. Capitanio (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society B* 61, 579–602.
- Azzalini, A. & A. Dalla Valle (1996). The multivariate skew-normal distribution. *Biometrika* 83, 715–726.
- Bauwens, L. & S. Laurent (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics* 23, 346–354.
- Del Brio, E., T. Niguez & J. Perote (2008). Multivariate Gram–Charlier densities. *Working paper*. Доступно на <http://mpira.ub.uni-muenchen.de/29073>.
- Diebold, F., T. Gunther & A. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863–883.
- Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics* 20, 339–350.
- Engle, R., & K. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122–150.
- Fernandez, C., J. Osiewalski & M. Steel (1995). Modeling and inference with v -spherical distributions. *Journal of the American Statistical Association* 90, 1331–1340.
- Giller, G. (2005). A generalized error distribution. *Giller Investments Research Note* 20031222/1. Доступно на <http://www.gillerinvestments.com/Downloader/Files/A%20Generalized%20Error%20Distribution.pdf>.
- Hansen, B. (2004). Nonparametric conditional density estimation. *Working paper*, University of Wisconsin. Доступно на <http://www.ssc.wisc.edu/~bhansen/papers/ncde.pdf>.
- Hong, Y., H. Li & F. Zhao (2007). Can the random walk model be beaten in out-of-sample density forecasts? Evidence from intraday foreign exchange rates. *Journal of Econometrics* 141, 736–776.
- Leon, A., G. Rubio & G. Serna (2005). Autoregressive conditional volatility, skewness and kurtosis. *Quarterly Review of Economics and Finance* 45, 599–618.
- Silvennoinen, A. & T. Terasvirta (2008). Multivariate GARCH models. *SSE/EFI Working Paper Series in Economics and Finance* 669. Доступно на <http://swopec.hhs.se/hastef/papers/hastef0669.pdf>.
- Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57, 307–333.

Modeling multivariate parametric densities of financial returns

Alexey Balaev

Higher School of Economics, Moscow, Russia

This paper compares several bivariate conditional density parameterizations for stock market returns in terms of in-sample fit and out-of-sample predictive ability for the whole conditional density. We consider Skew-Normal, Skew-Student, Skew-GED and Gram–Charlier densities. We focus on the ability of these density specifications to capture asymmetry and so called ‘multivariate tails’. Using a test based on Kullback–Leibler information criterion we conduct pairwise comparisons of estimated conditional density models in sample and out of sample. The models are ranked according to their quality of fit and predictive ability. We discuss the causes behind superiority of this or that density specification.

Keywords: conditional density, Gram–Charlier expansion, skewed distribution, quality of fit, predictive ability

JEL Classification: C14, C16, C22, C32, C51, C53, C58

Хеджирование фьючерсами: многомерные GARCH с динамическими условными корреляциями*

Алексей Колоколов[†]

*Российский экономический университет им. Г.В. Плеханова, Москва, Россия
Университет Тор Вергата, Рим, Италия*

В настоящей статье исследуются способы моделирования взаимосвязей между фьючерсными и спот-ценами финансовых индексов, а также проверяется практическая ценность эконометрических моделей для хеджирования фьючерсами на российских и зарубежных данных. Динамика фьючерсных и спот-цен описывается векторной моделью исправления ошибок, а волатильности и корреляции – различными многомерными GARCH-моделями из класса моделей с динамическими условными корреляциями разной степени детализации. Проведенное в работе эмпирическое исследование позволяет сделать выводы об эффективности применения стратегий хеджирования, основанных на многомерных GARCH-моделях, о сходствах и различиях взаимозависимостей между фьючерсами и базовыми активами на российском и иностранных финансовых рынках и о практически оправданной степени детализации многомерных GARCH-моделей.

Ключевые слова: фьючерсы, хеджирование, многомерные GARCH-модели, динамические условные корреляции

Классификация JEL: C32, C51, C53, G11, G15

1 Введение

Хеджирование фьючерсами заключается в создании короткой или длинной позиции по фьючерсным контрактам для ослабления эффектов неблагоприятных изменений цен базовых инструментов. Ключевая проблема хеджирования – выбор оптимального, в том или ином смысле, отношения хеджирования, которое определяется как отношение позиции по фьючерсам к позиции по базовому активу.

Долгое время при оценивании оптимальных отношений хеджирования превалировал статистический подход, предложенный и разработанный в Johnson (1960) и Ederington (1979), в рамках которого оптимальное отношение хеджирования определяется как коэффициент наклона в регрессии изменения спот-цены на изменение цены фьючерсного контракта и может быть оценено методом наименьших квадратов (МНК). Однако данный метод подвергся интенсивной критике: такая оценка базируется на оценках безусловных дисперсии и ковариации, а условная информация опускается (Myers & Thompson, 1989), полученные МНК-оценки неэффективны из-за наличия автокорреляционных связей в рядах цен, и игнорируются характерные для финансовых данных эффекты, такие как гетероскедастичность (Park & Bera, 1987).

В настоящее время развитие теории временных рядов и финансовой эконометрики позволяет получить оценки условных (относительно всей имеющейся к данному моменту времени информации) отношений хеджирования. Такие динамические отношения хеджирования рассчитываются как отношения условных ковариаций между ценами фьючерсов и спот-ценами активов к условным дисперсиям цен фьючерсов и минимизируют дисперсию суммарной позиции инвестора по хеджевым и хеджируемым активам.

*Цитировать как: Колоколов, Алексей (2011) «Хеджирование фьючерсами: многомерные GARCH с динамическими условными корреляциями», Квантиль, №9, стр. 61–75. Citation: Kolokolov, Alexei (2011) “Futures hedging: Multivariate GARCH with dynamic conditional correlations,” Quantile, No.9, pp. 61–75.

[†]Адрес: 117997, г. Москва, Стремянный переулок, д. 36. Электронная почта: alexeiuo@gmail.com

Для получения оценок динамических отношений хеджирования необходимо оценить условное математическое ожидание и условную ковариационную матрицу двумерного случайного процесса цен фьючерсных контрактов и спот-цен финансовых индексов. Интуитивная догадка о коинтегрируемости рядов значений индексов и фьючерсов зачастую подтверждается свойствами наблюдаемых данных. Более того, эмпирически установлено, что игнорирование коинтеграции ведет к недооценке отношений хеджирования (Ghosh, 1993).

Большую роль при оценивании оптимальных динамических отношений хеджирования играет выбор модели для условной ковариационной матрицы. Моделирование ковариационных матриц началось с модели VEC – прямого обобщения одномерных GARCH-моделей на векторный случай (Bollerslev, Engle & Wooldridge, 1988), после видоизмененной в более компактную модель – BEKK (Engle & Kroner, 1995). Данные модели оказались, однако, неудобными из-за сложности ограничений, гарантирующих положительную определенность ковариационной матрицы, большого количества параметров, подлежащих оцениванию, и неясности их интерпретации. Поэтому на смену им пришли новые модели, отдельно представляющие динамику корреляций и волатильности, – сначала модель постоянных корреляций (Bollerslev, 1990), затем – модели динамических корреляций (Tse & Tsui, 2002 и Engle, 2002). Признание и широкое распространение за ее простоту и эффективность заслужила модель динамических условных корреляций Энгла (Engle, 2002). Позже она неоднократно модифицировалась и дорабатывалась, что позволило исследователям и практикам учитывать при моделировании разнообразные эффекты динамики корреляций, в частности, эффект асимметрии (Cappiello, Engle & Sheppard, 2006).

Большое количество литературы посвящено изучению практической ценности различных моделей с меняющимися условными корреляциями, например, Bystrom (2003), Lee & Yoder (2007), Skintzi & Xanthopolous-Sisinis (2007), Yang & Allen (2004), и неоднократно констатировалась их высокая эффективность. В то же время известны случаи, когда эффективность применения моделей более высокого уровня может быть существенно снижена из-за ошибок, возникающих по причине сложности оценивания их параметров и увеличения транзакционных издержек (Tse & Tsui, 2002).

Данная работа посвящена оцениванию оптимальных отношений хеджирования с помощью трех многомерных GARCH-моделей класса изменяющихся условных корреляций разной степени детализации и эмпирическому исследованию эффективности предложенных подходов на российском и зарубежных финансовых рынках.

Во втором разделе приводятся теоретические аспекты хеджирования фьючерсными контрактами. В третьем и четвертом разделах кратко описаны модели, использованные в работе для прогнозирования условных ковариаций, и способы оценивания их параметров. В пятом разделе представлены данные, использованные в работе, приведены оценки параметров моделей и результаты проверки связанных с ними статистических гипотез. Исследование эффективности применения моделей для оценки оптимальных отношений хеджирования представлено в шестом разделе.

2 Стратегии хеджирования фьючерсными контрактами

При вычислении количества фьючерсных контрактов, необходимого для сокращения риска конкретной позиции по базовым хеджируемыми активам, исследователи и практики апеллируют к понятию отношения хеджирования. По определению Халла (Hull, 2006) отношение хеджирования – это отношение объема инвестиций в хеджевый актив к объему инвестиций в актив, подверженный риску.

Предположим, что инвестор занимает длинную позицию по одному базовому активу, и обозначим через s_t и f_t логарифмы цен хеджируемого (базового) и хеджевого (фьючерса)

активов, соответственно, в момент времени t . Тогда доходность инвестора в момент t составит

$$R_t^u = s_t - s_{t-1}$$

в случае, если инвестор не предпринимает действий по хеджированию инвестиций, а при хеджировании фьючерсами она может быть принята примерно равной

$$R_t^h = (s_t - s_{t-1}) - h_t(f_t - f_{t-1}),$$

где $s_t - s_{t-1}$ и $f_t - f_{t-1}$ – доходности инвестора по базовому активу и по фьючерсным контрактам соответственно, а h_t – отношение хеджирования в момент времени t . Нетрудно проверить, что условная, относительно информации до момента времени t , дисперсия доходности инвестора в момент t представима в виде суммы:

$$\mathbb{V}_{t-1}(R_t^h) = \mathbb{V}_{t-1}(s_t) + h_t^2 \mathbb{V}_{t-1}(f_t) - 2h_t \mathbb{C}_{t-1}(s_t, f_t),$$

где $\mathbb{V}_{t-1}(s_t)$ и $\mathbb{V}_{t-1}(f_t)$ – условные дисперсии логарифмов цен базового актива и фьючерса, а $\mathbb{C}_{t-1}(s_t, f_t)$ – условная ковариация между ними.

Исходя из принципа полного ослабления риска, оптимальное в момент t отношение хеджирования можно определить как такое отношение хеджирования, при котором условная, относительно информации до момента времени t , дисперсия доходности в момент t достигает минимума (см. Ederington, 1979 и Hull, 2006):

$$h_t^* = \frac{\mathbb{C}_{t-1}(s_t, f_t)}{\mathbb{V}_{t-1}(f_t)}. \quad (1)$$

На практике при расчете оптимального отношения хеджирования в момент времени t , располагая данными до момента t , условные ковариации заменяются прогнозными значениями, построенными на основе доступной статистики. Таким образом, точность прогноза оптимального отношения хеджирования напрямую зависит от точности эконометрических моделей, применяющихся для прогнозирования условных ковариаций.

Эффективность хеджирования измеряется как выраженное в относительных единицах сокращение безусловной дисперсии дохода инвестора при хеджировании (Ederington, 1979):

$$u(\hat{h}_t^*) = \frac{\mathbb{V}_{t-1}(R_t^u)}{\mathbb{V}_{t-1}(R_t^h)}. \quad (2)$$

Величина данного показателя зависит от точности прогнозирования оптимального отношения хеджирования, а, следовательно, от выбора эконометрической модели для прогноза условных ковариаций. Таким образом, показатель эффективности хеджирования может использоваться для сравнения прогностических способностей этих моделей.

Справедливо заметить, однако, что стратегия минимизации дисперсии дохода плохо описывает поведение реального инвестора, зависящее и от ожидаемого дохода. Учитывая этот факт, можно предложить более реалистичное определение оптимального отношения хеджирования, как такого отношения хеджирования, при котором функция полезности инвестора достигает максимума.

Согласно Brooks, Henry & Persaud (2002), для инвестора, характеризуемого функцией полезности вида

$$U(\mathbb{E}_{t-1}(R_t^h), \mathbb{V}_{t-1}(R_t^h)) = \mathbb{E}_{t-1}(R_t^h) - \gamma \mathbb{V}_{t-1}(R_t^h),$$

где γ – отношение инвестора к риску (боязнь риска), а $\mathbb{E}_{t-1}(R_t^h)$ – условное математическое ожидание доходности, оптимальное отношение хеджирования имеет вид:

$$h_t^* = \frac{\mathbb{C}_{t-1}(s_t, f_t)}{\mathbb{V}_{t-1}(f_t)} - \frac{\mathbb{E}_{t-1}(f_t - f_{t-1})}{\gamma \mathbb{V}_{t-1}(f_t)}. \quad (3)$$

Второе слагаемое, появляющееся в правой части, принято интерпретировать как спекулятивный спрос на фьючерсы. Два приведенных определения оптимального отношения хеджирования эквивалентны, если выполнено одно из дополнительных предположений: либо коэффициент γ стремится к бесконечности (инвестор не приемлет никакого риска вообще, вне зависимости от ожидаемого дохода), либо ожидаемый доход по фьючерсам равен нулю. Конечно, данные дополнительные предположения нереалистичны, однако определение, основанное на минимизации дисперсии, является важным и зачастую более удобным для теоретических исследований.

Основной целью данной работы является изучение возможностей применения многомерных GARCH-моделей для оценивания и прогнозирования условных ковариаций на различных финансовых рынках, а не разработка «реальной стратегии», поэтому в дальнейшем используется оптимальное отношение хеджирования, определяемое формулой (1). Такой подход позволяет проще получить представление о том, насколько хорошо та или иная эконометрическая модель описывает эволюцию условных ковариаций между хеджируемыми активами и фьючерсами, исходя из цепочки простых рассуждений, не являющихся, впрочем, строгим математическим доказательством: большее сокращение дисперсии дохода при хеджировании достигается при более точной оценке оптимального отношения хеджирования, а точность оценок оптимального отношения хеджирования повышается с увеличением точности оценок условных ковариаций. Следовательно, эконометрические модели, обеспечивающие большую эффективность хеджирования (в смысле сокращения дисперсии), лучше описывают реальный процесс, генерирующий наблюдаемые значения, и должны обеспечивать и большую точность оценок оптимального отношения хеджирования, определенного при помощи функции полезности. При построении реальной стратегии хеджирования инвестор сможет полагаться на теоретические результаты данного исследования для оценивания ковариационной матрицы и определять оптимальное отношение хеджирования в зависимости от своего отношения к риску по формуле (3).

3 Модели постоянных и динамических условных корреляций

Пусть каждому моменту времени t соответствует случайный двумерный вектор y_t , компонентами которого являются доходности фьючерсного контракта и финансового индекса. Предполагается, что данный векторный случайный процесс $\{y_t\}_{t \in \mathbb{Z}}$ имеет вид

$$y_t = \mathbb{E}(y_t | F_{t-1}) + \epsilon_t, \quad \epsilon_t = \Sigma_t^{1/2} z_t, \quad (4)$$

где F_{t-1} – вся информация, доступная к моменту времени $t - 1$, Σ_t – положительно определенная матрица, $\Sigma_t^{1/2}$ – разложение Холецкого для Σ_t , z_t – независимые одинаково распределенные случайные векторы с $\mathbb{E}(z_t) = 0$ и $\mathbb{V}(z_t) = I$. Тогда

$$\mathbb{V}(y_t | F_{t-1}) = \Sigma_t^{1/2} \mathbb{V}(z_t) (\Sigma_t^{1/2})' = \Sigma_t.$$

Условное математическое ожидание и ковариационная матрица являются функциями от неизвестных параметров и наблюдаемых значений. Дальнейшее моделирование заключается в придании конкретных параметрических форм $\mathbb{E}(y_t | F_{t-1})$ и Σ_t . В данной работе для $\mathbb{E}(y_t | F_{t-1})$ используется классическая векторная модель исправления ошибок, VECM (см. приложение), а для условной ковариационной матрицы – три многомерных GARCH-модели разной степени детализации: модель с постоянными корреляциями и модели симметричной и асимметричной динамики условных корреляций.

Общим для моделей ковариационной матрицы является ее представление в виде произведения:

$$\Sigma_t = D_t R_t D_t,$$

где $R_t = (\rho_{ijt})$ – матрица условных корреляций, а D_t – диагональная матрица с элементами σ_{iit} (корень из условной дисперсии компоненты i , где $i = 1, 2$) на главной диагонали. Положительная определенность матрицы Σ_t обеспечивается положительной определенностью матрицы R_t и положительностью σ_{iit} .

Элементы σ_{iit} являются функциями от параметров и соответствующих компонент вектора ошибок ϵ_t . В общем случае формы процессов σ_{iit} могут быть различны для каждого номера i (для каждого одномерного ряда), но в данной работе все они моделируются как одномерные GARCH.

Простейшей моделью для матрицы R_t является модель постоянных условных корреляций (*Constant Conditional Correlation, CCC*), предложенная Тимом Болерслевым (Bollerslev, 1990), в которой полагается, что матрица R_t постоянна во времени:

$$R_t = R = (\rho_{ij}),$$

Данная модель имеет очевидную интерпретацию и легко оценивается в два шага: сначала находят оценки параметров одномерных GARCH, после чего рассчитывают выборочные ковариации между стандартизированными остатками. Тем не менее, априорное предположение о неизменности условных корреляций часто является необоснованным и может привести к недопустимым неточностям.

Естественным обобщением CCC-модели, допускающим изменение условных корреляций во времени, является модель динамических условных корреляций (*Dynamic Conditional Correlation, DCC*) Роберта Энгла (Engle, 2002). Для обеспечения особой формы корреляционной матрицы (симметричности, единиц на главной диагонали и меньших единицы по модулю недиагональных элементов) R_t представляется в виде

$$R_t = (\text{diag}(Q_t))^{-\frac{1}{2}} Q_t (\text{diag}(Q_t))^{-\frac{1}{2}},$$

где Q_t – положительно определенная симметричная матрица, эволюционирующая в соответствии с процессом

$$Q_t = (1 - \theta_1 - \theta_2) \bar{Q} + \theta_1 u_{t-1} u'_{t-1} + \theta_2 Q_{t-1},$$

где $u_t = [u_{1t}, u_{2t}]'$, $u_{it} = \epsilon_{it} / \sigma_{iit}$ – стандартизированные остатки, $i = 1, 2$, \bar{Q} – безусловная ковариационная матрица u_t .

Для обеспечения положительной определенности матриц R_t и, следовательно, Σ_t , параметры θ_1 и θ_2 должны быть положительны и их сумма не должна превышать единицу (Engle, Sheppard, 2001).

Динамика условных корреляций в DCC-моделях объясняется их зависимостью от предыдущих значений шоков (ошибок), поправленных на волатильность: одинаково направленные остатки ведут к увеличению условных корреляций, разнонаправленные – к уменьшению. Однако не склонные к риску инвесторы скорее реагируют на негативную информацию о рынке, поэтому для более точного описания динамики условных корреляций разумно допустить, что изменение условных корреляций может различаться для положительных и отрицательных значений остатков. Для учета такого эффекта асимметрии динамики используется модификация DCC модели – асимметричная модель динамических условных корреляций (*Asymmetric Dynamic Conditional Correlation, ADCC*).

Различие между DCC- и ADCC-моделями заключается в параметрической форме процесса для матриц Q_t . Для ADCC-матрицы Q_t изменяются в соответствии с уравнением:

$$Q_t = (1 - \theta_1 - \theta_2) \bar{Q} - \theta_3 \bar{N} + \theta_1 u_{t-1} u'_{t-1} + \theta_2 Q_{t-1} + \theta_3 \eta_{t-1} \eta'_{t-1},$$

где $\eta_t = \mathbb{I}\{u_t < 0\} \odot u_t$, $\bar{N} = \mathbb{E}[\theta_t \theta'_t]$, \odot обозначает поэлементное умножение. В данном случае положительная определенность матрицы R_t может быть гарантирована следующими условиями: параметры θ_1 , θ_2 и θ_3 неотрицательны, и их сумма меньше единицы. Второе условие

может быть заменено менее строгим (Cappiello, Engle & Sheppard, 2006) $\theta_1 + \theta_2 + \delta\theta_3 < 1$, где δ – наибольшее собственное значение матрицы $\bar{Q}^{-\frac{1}{2}}\bar{N}\bar{Q}^{-\frac{1}{2}}$.

Приведенные многомерные GARCH-модели относятся к классу моделей с изменяющимися условными корреляциями. Заметим, что CCC-модель может рассматриваться как DCC при $\theta_1 = \theta_2 = 0$, которая в свою очередь является ADCC-моделью при $\theta_3 = 0$.

4 Способы оценивания параметров

Векторная модель исправления ошибок (условное математическое ожидание) оценивается методом наименьших квадратов (см. Lütkepohl, 2005). Оценивание параметров многомерных GARCH-моделей осуществляется на основе остатков $\hat{\epsilon}_t$, полученных после оценивания условного математического ожидания $E[y_t|F_{t-1}]$.

Предположив некоторый закон распределения для z_t , можно оценить параметры многомерных GARCH-моделей методом максимального правдоподобия. В частности, если $z_t \sim N(0, I_m)$, то логарифм функции правдоподобия (без константы) имеет вид:

$$\ln L(\theta) = -\frac{1}{2} \sum_{t=1}^n (\ln \det \Sigma_t + \hat{z}_t' \Sigma_t^{-1} \hat{z}_t)$$

Вектор θ здесь обозначает вектор всех параметров модели. Точка максимума $\ln L(\theta)$, $\hat{\theta}$, является состоятельной, асимптотически нормальной и эффективной оценкой θ . Если же предположение $z_t \sim N(0, I_m)$ не выполняется, $\hat{\theta}$ следует понимать как оценку методом максимального псевдо- (квази-) правдоподобия, также являющуюся состоятельной и асимптотически нормальной, однако не эффективной.

Большим достоинством многомерных GARCH-моделей с динамическими условными корреляциями является возможность использования для их оценивания двухшаговый метод (Engle, 2002): на первом шаге оцениваются параметры одномерных GARCH, на втором – параметры уравнения динамики условных корреляций. Такой подход облегчает работу с одномерными волатильностями и позволяет избежать численных трудностей, возникающих при оптимизации. Формально он может быть изложен следующим образом: функция правдоподобия

$$\ln L(\theta) = -\frac{1}{2} \sum_{t=1}^n (\ln \det D_t R_t D_t + \hat{u}_t' R_t^{-1} \hat{u}_t), \quad u_t = D_t^{-1} \epsilon_t,$$

может быть представлена в виде суммы:

$$-\frac{1}{2} \sum_{t=1}^T (2 \ln \det D_t + u_t' u_t) - \frac{1}{2} \sum_{t=1}^T (\ln \det R_t + u_t' R_t^{-1} u_t - u_t' u_t) = \ln L^1(\theta_1) + \ln L^2(\theta_2|\theta_1).$$

Здесь θ_1 и θ_2 обозначают векторы параметров соответственно условных дисперсий (параметры в D_t) и условных корреляций (параметры в R_t). Заметим, что $\ln L^1(\theta_1)$ является функцией правдоподобия совместно для всех одномерных GARCH. На первом шаге оцениваются параметры θ_1 :

$$\hat{\theta}_1 = \arg \min_{\theta_1} \ln L^1(\theta_1),$$

а на втором – параметры θ_2 , используя оценки, полученные на первом шаге:

$$\hat{\theta}_2 = \arg \min_{\theta_2} \ln L^2(\theta_2|\hat{\theta}_1),$$

Оценки двухшагового метода $\hat{\theta}_1$ и $\hat{\theta}_2$ можно рассматривать как оценки обобщенного метода моментов (см. Engle & Sheppard, 2001 и Newey & McFadden, 1994). Они состоятельны и

асимптотически нормальны, но не эффективны. Их асимптотическое распределение имеет вид:

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, A_0^{-1} B_0 A_0^{-1}),$$

где

$$A_0 = \begin{pmatrix} \partial^2 \ln L^1(\theta_1) / \partial \theta_1 \partial \theta_1' & 0 \\ \partial^2 \ln L^2(\theta_2) / \partial \theta_1 \partial \theta_2' & \partial^2 \ln L^2(\theta_2) / \partial \theta_2 \partial \theta_2' \end{pmatrix},$$

$$B_0 = \mathbb{V} \left[T^{-\frac{1}{2}} \sum_{t=1}^T \left(\frac{\partial \ln L^1(\theta_1)}{\partial \theta_1'}, \frac{\partial \ln L^1(\theta_1, \theta_2)}{\partial \theta_2'} \right)' \right]$$

5 Описание данных и оценки параметров

Для проведения эмпирического исследования используются исторические дневные значения российского индекса RTS и мировых индексов: DAX, S&P500 (далее S&P) и NASDAQ COMPOSITE (далее NASDAQ) в период с 1 января 2008 г. по 22 июля 2010 г., а также котировки соответствующих фьючерсных контрактов. Обработка данных и процедуры оценки параметров моделей осуществлялись с использованием программы MATLAB.

Индексы RTS, DAX и S&P являются основными показателями, характеризующими соответственно российский, немецкий и американский фондовые рынки. Финансовый индекс NASDAQ является индикатором одной из основных фондовых бирж США, NASDAQ, специализирующейся на торговле акциями высокотехнологичных компаний.

Общее число наблюдений составляет 635 и 668 торговых дней для индексов RTS и DAX, S&P500 и NASDAQ соответственно, причем различие в длине рядов обусловлено различным количеством праздничных нерабочих дней. Для каждого ряда последние 60 пар наблюдений (около трех торговых месяцев) используются для проведения вневыборочного анализа. Описательные статистики рядов данных представлены в таблице 1.

Таблица 1: Описательные статистики

	RTC		DAX		NASDAQ		S&P	
	фьючерс	индекс	фьючерс	индекс	фьючерс	индекс	фьючерс	индекс
mean	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
med	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
std	0,04	0,03	0,02	0,02	0,02	0,02	0,02	0,02
skewness	-0,40	-0,23	0,31	0,29	0,11	-0,09	0,12	-0,14
kurtosis	10,87	9,91	9,08	8,19	10,17	7,02	10,30	8,17
max	0,23	0,20	0,12	0,11	0,13	0,11	0,13	0,11
min	-0,31	-0,21	-0,08	-0,07	-0,11	-0,10	-0,10	-0,09

В таблице представлены описательные статистики для рядов доходностей индексов (индекс) и соответствующих фьючерсных контрактов (фьючерс). По строкам расположены статистики: mean – среднее значение, med – оценка медианы, std – оценка стандартного отклонения, skewness – коэффициент асимметрии, kurtosis – эксцесс, max – максимальное значение, min – минимальное значение.

Стоит отметить, что коэффициенты асимметрии доходности всех индексов отличны от нуля. Для индексов DAX, NASDAQ и S&P они равны соответственно 0,31, 0,11 и 0,12, а коэффициент асимметрии доходности индекса RTC составляет -0,23. Таким образом, у доходности индекса RTS более вероятно появление значений из левого хвоста распределения значений, а для DAX, NASDAQ и S&P – из правого.

Для всех рассматриваемых пар ряды значений индексов и фьючерсных контрактов демонстрируют признаки коинтегрируемости. Для каждой пары построена двумерная модель

исправления ошибок. Оценки их параметров и среднеквадратические отклонения оценок представлены в таблице 2.

Количество лагов в VECM выбрано согласно показаниям информационных критериев (Шварца, Акаике, Ханнана–Куина). Для рядов с индексами RTS и DAX оно равно трем, для NASDAQ – двум, и для S&P – одному; таким образом, процессы с индексами RTS и DAX обладают более длинной памятью, чем с индексами NASDAQ и S&P. Оценки коинтегрирующего вектора близки к -1 для всех пар индекс–фьючерс, за исключением пары с индексом NASDAQ. Кроме того, для нее оценки поправочных коэффициентов положительны, в то время как оценки этих параметров для всех остальных пар отрицательны.

Таблица 2: Оценки параметров двумерных VECM

RTS								
alpha	beta'	A1		A2		A3		const
-0,430*	1,000	0,004	0,148	-0,175*	0,112	-0,115*	0,109	-0,027*
(0,120)		(0,116)	(0,131)	(0,102)	(0,110)	(0,007)	(0,074)	(0,008)
0,067	-1,008*	0,470*	-0,390*	0,306*	-0,235*	0,073	0,014	0,000
(0,083)	(0,005)	(0,080)	(0,091)	(0,071)	(0,076)	(0,053)	(0,051)	(0,006)
DAX								
alpha	beta'	A1		A2		A3		const
-0,760*	1,000	-0,330	0,277	-0,338	0,272	-0,350	0,308	-0,043
(0,220)		(0,318)	(0,322)	(0,313)	(0,318)	(0,283)	(0,288)	(0,120)
-0,641*	-1,007*	0,044	-0,112	-0,058	-0,015	-0,062	0,010	-0,037
(0,220)	(0,004)	(0,318)	(0,322)	(0,313)	(0,318)	(0,283)	(0,288)	(0,120)
NASDAQ								
alpha	beta'	A1		A2		A3		const
0,131*	1,000	-0,173	0,051	-0,412*	0,301*	–	–	0,950*
(0,035)		(0,146)	(0,143)	(0,145)	(0,141)			(0,025)
0,141*	-0,066*	0,219*	-0,330*	-0,253*	0,150	–	–	0,026
(0,036)	(0,031)	(0,148)	(0,146)	(0,148)	(0,144)			(0,026)
S&P								
alpha	beta'	A1		A2		A3		const
-1,869*	1,00	0,654*	-0,812*	–	–	–	–	-0,107
(0,287)		(0,251)	(0,256)					(0,160)
-1,349*	-1,008*	0,813*	-0,986*	–	–	–	–	-0,177
(0,289)	(0,001)	(0,253)	(0,251)					(0,160)

В таблице представлены оценки параметров двумерных моделей исправления ошибок и их стандартные отклонения (в скобках) для четырех пар индекс–фьючерс. Параметры записаны в виде матриц: alpha – матрица поправочных коэффициентов, beta – стандартизованная матрица коинтегрирующих векторов, A1, A2, A3 – матрицы параметров векторной авторегрессии, const – константы. Значимые на 5% уровне доверия параметры помечены как *.

Для каждой пары индекс–фьючерс построены модели CCC, DCC и ADCC. Таблица 3 содержит оценки параметров двумерных GARCH-моделей и их стандартные ошибки. Альтернативные модели корреляций протестированы одна против другой с помощью критерия отношения правдоподобия. Результаты проверки гипотез представлены в таблице 4.

Оценка постоянного условного коэффициента корреляции CCC-модели для индекса RTS равна 0,862 и существенно меньше аналогичных оценок для других индексов, близких к 1, что свидетельствует о существовании в российской финансовой системе существенного объема информации, по-разному влияющей на фондовый рынок и рынок фьючерсов. Оценки параметров асимметрии ADCC-модели для всех изучаемых пар незначимы на 5% уровне значимости.

Согласно критериям отношения правдоподобия нулевая гипотеза о постоянстве условных корреляций может быть отвергнута при уровне значимости 0,05 для всех изучаемых рядов данных. Нулевая гипотеза о симметричной динамике условных корреляций не может быть отвергнута в пользу альтернативной гипотезы об асимметричной динамике для всех рядов

Таблица 3: Оценки параметров многомерных GARCH-моделей

	RTS			DAX		
	CCC	DCC	ADCC	CCC	DCC	ADCC
const1	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)
delta1	0,127* (0,001)	0,127* (0,001)	0,127* (0,001)	0,127* (0,001)	0,127* (0,001)	0,127* (0,001)
gamma1	0,872* (0,001)	0,872* (0,001)	0,872* (0,001)	0,872* (0,001)	0,872* (0,001)	0,872* (0,001)
const2	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)
delta2	0,136* (0,001)	0,136* (0,001)	0,136* (0,001)	0,136* (0,001)	0,136* (0,001)	0,136* (0,001)
gamma2	0,854* (0,001)	0,854* (0,001)	0,854* (0,001)	0,854* (0,001)	0,854* (0,001)	0,854* (0,001)
corr	0,862* (0,000)	—	—	0,995* (0,000)	—	—
theta1	—	0,029* (0,000)	0,029* (0,001)	—	0,185* (0,005)	0,184* (0,002)
theta2	—	0,958* (0,000)	0,950* (0,002)	—	0,487* (0,088)	0,487* (0,080)
theta3	—	—	0,011 (9,260)	—	—	0,328 (4,800)
	NASDAQ			S&P		
	CCC	DCC	ADCC	CCC	DCC	ADCC
const1	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)
delta1	0,127* (0,001)	0,127* (0,001)	0,127* (0,001)	0,127* (0,001)	0,127* (0,001)	0,127* (0,001)
gamma1	0,872* (0,001)	0,872* (0,001)	0,872* (0,001)	0,872* (0,001)	0,872* (0,001)	0,872* (0,001)
const2	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)
delta2	0,136* (0,001)	0,136* (0,001)	0,136* (0,001)	0,136* (0,001)	0,136* (0,001)	0,136* (0,001)
gamma2	0,854* (0,001)	0,854* (0,001)	0,854* (0,001)	0,854* (0,001)	0,854* (0,001)	0,854* (0,001)
corr	0,972* (0,000)	—	—	0,987* (0,000)	—	—
theta1	—	0,069* (0,001)	0,069* (0,001)	—	0,063* (0,002)	0,103* (0,024)
theta2	—	0,786* (0,013)	0,786* (0,017)	—	0,782* (0,043)	0,000 (12,025)
theta3	—	—	0,143* (0,083)	—	—	0,895* (0,544)

В таблице представлены оценки параметров CCC-, DCC- и ADCC-моделей и их стандартные отклонения (в скобках) для четырех пар индекс-фьючерс. Обозначения параметров: const, delta, gamma – константы, ARCH и GARCH параметры в одномерных GARCH-моделях соответственно, corr – коэффициент корреляции в CCC-модели, theta1, theta2 и theta3 – параметры DCC- и ADCC-моделей. Значимые на 5% уровне параметры помечены как *.

Таблица 4: Тесты отношения правдоподобия

RTS				
тест	статистика	критические значения		
H_0/H_1	LR	0,10	0,05	0,01
CCC/DCC	7,60	4,61	5,99	9,21
DCC/ADCC	2,69	2,71	3,84	6,63
DAX				
тест	статистика	критические значения		
H_0/H_1	LR	0,10	0,05	0,01
CCC/DCC	82,53	4,61	5,99	9,21
DCC/ADCC	-0,65	2,71	3,84	6,63
NASDAQ				
тест	статистика	критические значения		
H_0/H_1	LR	0,10	0,05	0,01
CCC/DCC	15,08	4,61	5,99	9,21
DCC/ADCC	1,34	2,71	3,84	6,63
S&P				
тест	статистика	критические значения		
H_0/H_1	LR	0,10	0,05	0,01
CCC/DCC	6,84	4,61	5,99	9,21
DCC/ADCC	0,74	2,71	3,84	6,63

В таблице представлены результаты проверки гипотез CCC против DCC (CCC/DCC) и DCC против ADCC (DCC/ADCC) с помощью критериев отношения правдоподобия, статистики тестов (LR) и критические значения.

при уровне значимости 0,10.

К сожалению, мощность тестов отношения правдоподобия может быть существенно снижена из-за потенциально неверной спецификации одномерных GARCH-моделей или неверной параметрической формы уравнения динамики условных корреляций. Поэтому для тестирования нулевой гипотезы о постоянстве условных корреляций против динамических условных корреляций используется также тест, предложенный в Engle & Sheppard (2001), основанный лишь на состоятельных оценках CCC-модели: если нулевая гипотеза верна, стандартизированные остатки $\hat{u}_t = \hat{R}^{-\frac{1}{2}} D_t^{-1} \hat{\epsilon}_t$ представляют собой последовательность независимых одинаково распределенных величин, и, следовательно, параметры в авторегрессии векторного произведения стандартизированных остатков должны быть равны нулю. Результаты тестирования приводятся в таблице 5.

Тестирование показало, что для разного количества лагов вспомогательной регрессии вероятность отвергнуть нулевую гипотезу при условии, что она верна, значительно различается. С большой степенью уверенности можно отклонить гипотезу о постоянстве условных корреляций для ряда доходностей индексов и фьючерсов DAX, и принять ее для ряда индексов и фьючерсов RTS и NASDAQ.

6 Результаты хеджирования

Анализ показывает, что применение для расчета отношений хеджирования оценок ковариаций, основанных на более детализированных моделях, DCC и ADCC, приводит к незначительному увеличению качества хеджей внутри исходной выборки и уменьшению при прогнозировании для индексов DAX, NASDAQ и S&P, а для индекса RTS – напротив, эффективность DCC-хеджа растет при прогнозировании. Отношение выборочной оценки диспер-

Таблица 5: Тесты на постоянство условных корреляций

	nlags	sta	p-value
RTS	2	1,92	0,59
	5	2,65	0,85
	10	4,70	0,94
DAX	2	41,84	0,00
	5	63,45	0,00
	10	64,84	0,00
NASDAQ	2	0,21	0,98
	5	3,53	0,74
	10	6,10	0,87
SP	2	4,01	0,26
	5	5,98	0,42
	10	19,22	0,06

В таблице приводятся результаты тестов на постоянство условных корреляций. По столбцам справа налево расположены: наименование ряда, количество лагов вспомогательной регрессии, используемой в тесте (nlags), статистика теста (sta) и достигаемый уровень значимости (p-value).

сии доходности нехеджированной позиции к оценке дисперсии доходности хеджированной варьируется от 44,997 (при хеджировании индекса DAX с помощью DCC) до 1,637 (при хеджировании индекса RTS с помощью ADCC) внутри выборки и от 75,139 (при хеджировании индекса DAX с помощью CCC) до 1,881 (при хеджировании индекса RTS с помощью ADCC) вне выборки.

Лучшими для хеджирования вне исходной выборки индексов NASDAQ, DAX и S&P оказались модели постоянных условных корреляций, а для хеджирования индекса RTS – модель динамических условных корреляций. Относительное изменение дисперсии дохода хеджей при переходе от одной модели к другой является незначительным для индекса RTS: отношения оценок дисперсий составляют 1,012 (отношение оценки дисперсии DCC-хеджа к оценке дисперсии CCC-хеджа) и 0,851 (отношение оценки дисперсии ADCC-хеджа к оценке дисперсии CCC-хеджа). А для индексов NASDAQ, DAX и S&P наблюдается существенное снижение качества хеджей при переходе на DCC и ADCC модели.

Таким образом, существенного улучшения качества хеджей при детализации эконометрических моделей не наблюдается. Подобные эффекты присущи громоздким моделям, требующим оценивания слишком многих параметров, и объясняются накапливающейся неточностью оценок. Способность тестов выявлять наиболее адекватную для описания данных модель ослабляется по аналогичным причинам.

Сокращение дисперсии позиции при хеджировании как внутри, так и вне исходной выборки заметно выше для американского и немецкого рынков. Данное наблюдение объясняется более тесной связью между индексами и фьючерсными контрактами на этих развитых рынках, обусловленной высокой ликвидностью, быстротой реакции фьючерсного рынка на поведение спот-рынка, меньшим количеством спекулятивных сделок.

Для иллюстрации вневыборочных хеджей для всех пар индекс–фьючерс на Рис. 1 показаны оцененные методом ядерного сглаживания плотности распределения доходностей нехеджированных и хеджированных инвестиций, причем оценки плотностей распределений хеджей рассчитываются на основе лучших (в смысле показателя отношения дисперсий) моделей.

Таблица 6: Относительные изменения оценок дисперсий доходностей при хеджировании

Внутривыборочный анализ					
RTS			DAX		
u(CCC)	u(DCC)	u(ADCC)	u(CCC)	u(DCC)	u(ADCC)
1,646	1,639	1,637	44,668	44,997	44,973
CCC/CCC	DCC/CCC	ADCC/CCC	CCC/CCC	DCC/CCC	ADCC/CCC
1,000	0,996	0,995	1,000	1,007	1,007
NASDAQ			S&P		
u(CCC)	u(DCC)	u(ADCC)	u(CCC)	u(DCC)	u(ADCC)
13,360	13,438	13,435	30,084	30,097	29,980
CCC/CCC	DCC/CCC	ADCC/CCC	CCC/CCC	DCC/CCC	ADCC/CCC
1,000	1,006	1,006	1,000	1,000	0,997
Вневыборочный анализ					
RTS			DAX		
u(CCC)	u(DCC)	u(ADCC)	u(CCC)	u(DCC)	u(ADCC)
2,211	2,238	1,881	75,139	9,300	14,840
CCC/CCC	DCC/CCC	ADCC/CCC	CCC/CCC	DCC/CCC	ADCC/CCC
1,000	1,012	0,851	1,000	0,124	0,197
NASDAQ			S&P		
u(CCC)	u(DCC)	u(ADCC)	u(CCC)	u(DCC)	u(ADCC)
28,487	16,473	9,249	48,586	27,776	43,881
CCC/CCC	DCC/CCC	ADCC/CCC	CCC/CCC	DCC/CCC	ADCC/CCC
1,000	0,578	0,325	1,000	0,572	0,903

В таблице представлены отношения оценок дисперсий доходностей нехеджированных инвестиций к оценкам дисперсий хеджированных, построенных с помощью CCC (u(CCC)), DCC (u(DCC)) и ADCC (u(ADCC)) моделей, и отношения оценок дисперсий всех хеджей к оценке дисперсии CCC-хеджа (CCC/CCC, CCC/DCC, CCC/ADCC).

Для всех индексов распределения доходностей хеджей сконцентрированы вокруг нуля, и разброс их значений меньше разброса значений доходов нехеджированных инвестиций. Доход хеджа индекса RTS имеет наибольшую дисперсию среди всех рассматриваемых индексов.

В таблице 7 представлены статистики, характеризующие ряды оценок оптимальных отношений хеджирования, полученных для каждой пары индекс-фьючерс на основе моделей CCC, DCC и ADCC. Характерные отличия оценок отношений хеджирования, полученных на основе разных моделей, одинаковы для всех изучаемых индексов. Ряды оценок оптимальных отношений хеджирования демонстрируются на Рис. 2.

Оценки оптимальных отношений хеджирования, основанные на моделях ADCC и DCC, всегда меньше оценок, основанных на модели с постоянными условными корреляциями. Учитывая, что CCC-хедж является лучшим для почти всех рассматриваемых индексов, можно утверждать, что хеджи с динамическими корреляциями недооценивают истинные оптимальные отношения хеджирования.

7 Заключение

Приведенное исследование показывает, что многомерные GARCH-модели с динамическими условными корреляциями, DCC и ADCC, не способны ощутимо улучшить оценки оптимальных динамических отношений хеджирования по сравнению с простыми моделями с неизменными условными корреляциями, CCC. Учитывая показания тестов на постоянство условных корреляций, можно утверждать, что CCC-модели являются лучшей аппроксимацией для процесса условных корреляций, чем страдающие из-за неправильной спецификации

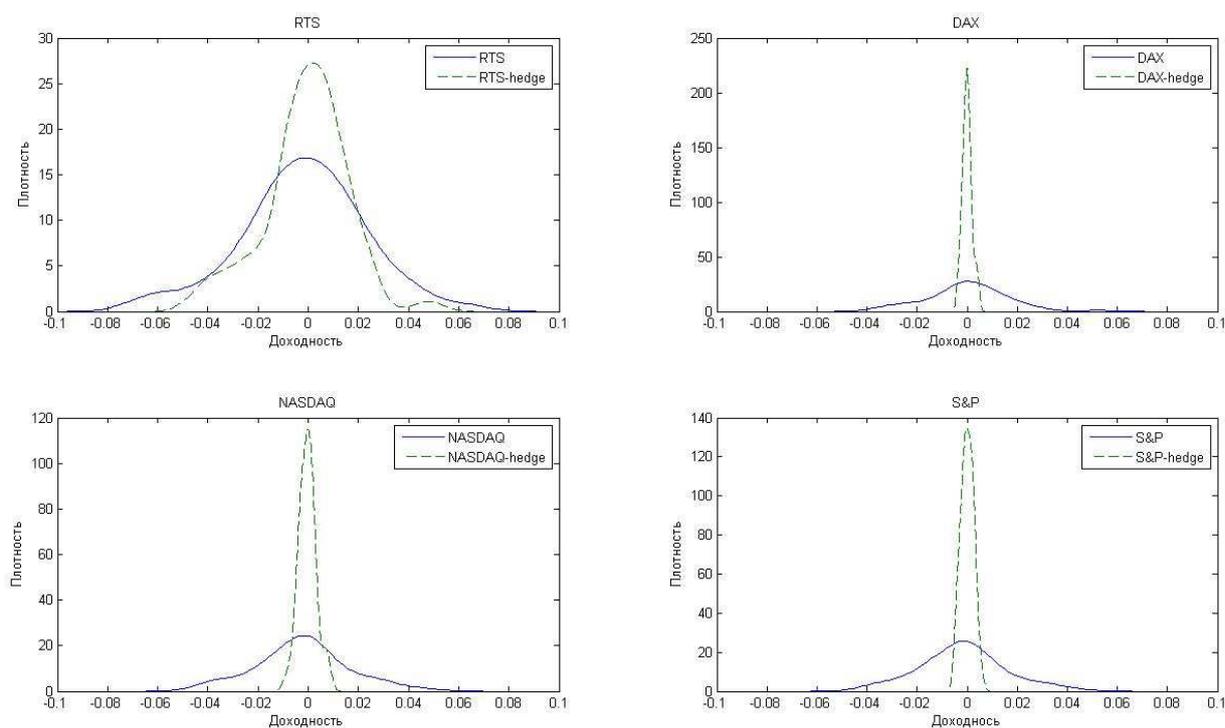


Рис. 1: Распределения доходностей хеджированных и нехеджированных инвестиций

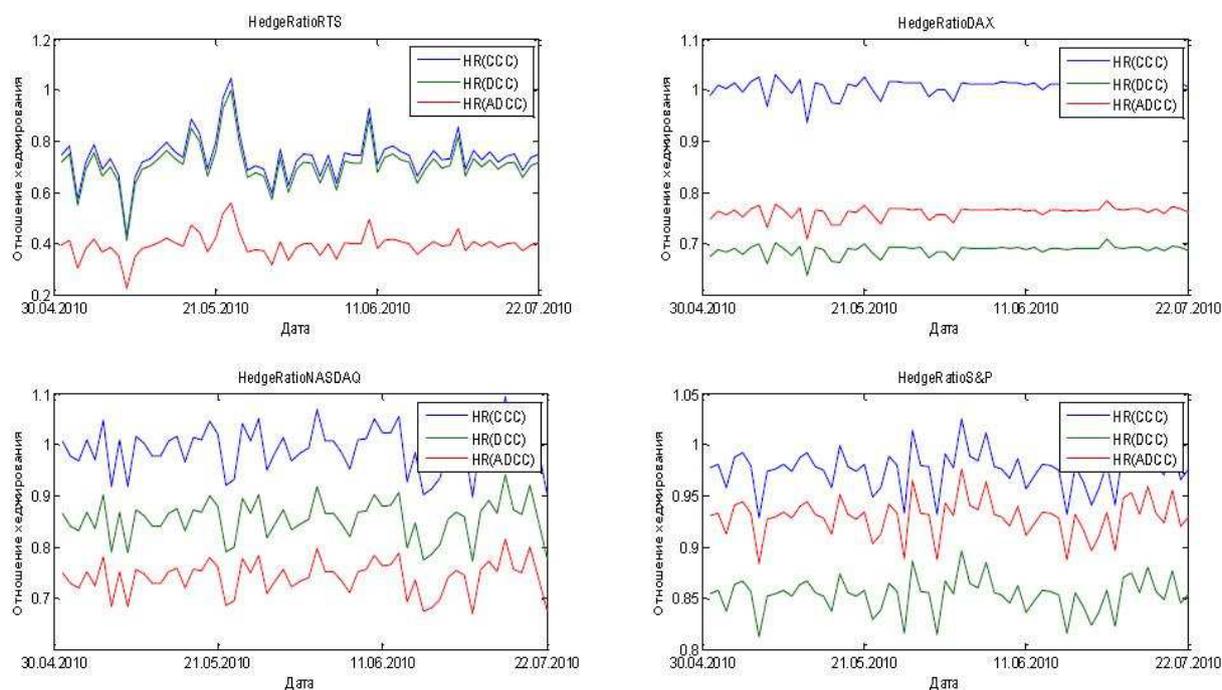


Рис. 2: Динамика отношений хеджирования

и накопления ошибок при оценивании параметров модели DCC и ADCC, недооценивающие истинные оптимальные отношения хеджирования.

Тем не менее, хеджи, основанные на CCC-моделях, позволяют существенно сократить дисперсии финансовых позиций по индексам, по крайней мере, на развитых финансовых рынках – немецком и американском.

Для российского рынка и специализированного индекса NASDAQ сконструированные хеджи относительно менее эффективны. Модели CCC, DCC и ADCC, управляющие динамикой условных корреляций лишь несколькими параметрами, оказываются недостаточно гибкими для исчерпывающего описания процесса, а использование более громоздких моделей рискованно в связи с уже отмеченной проблемой накопления неточностей и не может гарантировать значительное сокращение дисперсии хеджированной позиции.

Приложение. Коинтеграция и векторные модели исправления ошибок

Одномерные ряды значений финансовых индексов и соответствующих фьючерсных контрактов почти всегда оказываются нестационарными, а их флуктуации часто демонстрируют наличие устойчивых долгосрочных соотношений между рядами. Для формального описания таких рядов используются понятие о коинтеграции и векторные модели исправления ошибок (*Vector Error Correction Models, VECM*).

Двумерный нестационарный случайный процесс y_t называется коинтегрируемым, если существует вектор $\beta \in \mathbb{R}^2$ такой, что процесс $\beta' y_t$ стационарен, т.е. существует линейная комбинация компонент вектора y_t , являющаяся стационарным процессом. Данная линейная комбинация интерпретируется как долгосрочное равновесие (*long-run equilibrium*), устойчивое отношение между компонентами ряда, которое в каждый конкретный момент времени может и не выполняться точно, но равняться случайным величинам, последовательность которых стационарна. Вектор β называют коинтегрирующим вектором (*cointegrating vector*).

Основываясь на понятии о коинтеграции, можно предложить следующую модель временного ряда:

$$\Delta y_t = \alpha \beta' y_{t-1} + \mu + \Gamma_1 \Delta y_{t-2} + \dots + \Gamma_p \Delta y_{t-p} + u_t,$$

где u_t – последовательность мартингал-разностей относительно естественной фильтрации, $\Gamma_i \in \mathbb{R}^{2 \times 2}$, $i = 1, \dots, p$, $\mu \in \mathbb{R}^2$ – матрицы параметров, $\alpha \in \mathbb{R}^2$ – вектор параметров, и β – коинтегрирующий вектор. Величина $\beta' y_t$ выражает отклонение от долгосрочного равновесия в момент t , параметр α показывает скорость возвращения системы к состоянию равновесия. Сумма $\mu + \Gamma_1 \Delta y_{t-2} + \dots + \Gamma_p \Delta y_{t-p}$ интерпретируется как краткосрочные колебания.

Состоятельные оценки параметров VECM можно получить методом наименьших квадратов, но они имеют нестандартное асимптотическое распределение. На втором шаге можно получить суперсостоятельную оценку коинтегрирующего вектора обобщенным методом наименьших квадратов.

Более общее определение коинтеграции и подробное обсуждение VECM, оценивания их параметров и диагностики можно найти в книге Lütkepohl (2005).

Список литературы

- Bystrom, H.N.E. (2003). The hedging performance of electricity futures on the Nordic power exchange. *Applied Economics* 35, 1–11.
- Bollerslev, T., R.F. Engle & J.M. Wooldridge (1988). A capital asset pricing model with time varying covariances. *Journal of Political Economy* 96, 116–131.
- Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics* 52, 5–59.
- Brooks, C., O.T. Henry & G. Persaud (2002). The effect of asymmetries on optimal hedge ratios. *Journal of Business* 75, 333–352.
- Cappiello, L., R.F. Engle & K. Sheppard (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics* 4, 537–572.
- Ederington, L.H. (1979). The hedging performances of the new futures markets. *Journal of Finance* 34, 157–170.

- Engle, R.F. & K.F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122–150.
- Engle, R.F. & K. Sheppard (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. *NBER Working Paper 8554*.
- Engle, R.F. (2002). Dynamic conditional correlation. A simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* 20, 339–350.
- Ghosh, A. (1993). Hedging with stock index futures: Estimation and forecasting with error correction model. *Journal of Futures Markets* 13, 743–752.
- Hull, J. (2006). *Options, Futures, and Other Derivatives*. Шестое издание. Prentice–Hall: Englewood Cliffs, NJ.
- Johnson, L.L. (1960). The theory of hedging and speculation in commodity futures. *Review of Economic Studies* 27, 139–151.
- Lee, H.T. & J.K. Yoder (2007). Optimal hedging with a regime-switching time-varying correlation GARCH model. *The Journal of Futures Markets* 27, 495–516.
- Lien, D., Y.K. Tse & A.K.C. Tsui (2002). Evaluating the hedging performance of the constant-correlation GARCH model. *Applied Financial Economics* 12, 791–798.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer–Verlag: Berlin.
- Myers, R.J. & S.R. Thompson (1989). Generalized optimal hedge ratio estimation. *American Journal of Agricultural Economics* 71, 858–867.
- Newey, W.K. & D.L. McFadden (1994). Large sample estimation and hypothesis testing. Глава 36 в *Handbook of Econometrics* (под редакцией R. Engle & D. McFadden), том 4, 2111–2245. Elsevier Science.
- Park, H. & A. Bera (1987). Interest rate volatility, basis, and heteroscedasticity in hedging mortgages. *American Real Estate and Urban Economics Association* 15, 79–97.
- Skintzi, V.D. & S. Xanthopoulos–Sisinis (2007). Evaluation of correlation forecasting models for risk management. *Journal of forecasting* 26, 497–526.
- Tse, Y. & A. Tsui (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business and Economic Statistics* 20, 351–362.
- Yang, W. & D. Allen (2004). Multivariate GARCH hedge ratios and hedging effectiveness in Australian futures markets. *Accounting and Finance* 45, 301–321.

Futures hedging: Multivariate GARCH with dynamic conditional correlations

Alexei Kolokolov

*Plekhanov Russian University of Economics, Moscow, Russia
University of Rome “Tor Vergata”, Rome, Italy*

This article studies modeling dependence between futures and spot prices of financial indices and verifies a practical value of econometric models for futures hedging using Russian and foreign data. The dynamics of futures and spot prices is described by an error correction model, while volatilities and correlations are modeled by various multivariate GARCH models with dynamic conditional correlations of different degree of detail. The empirical investigation carried out in the article can answer questions on effectiveness of hedging strategies based on multivariate GARCH models, on similarities and differences of dependencies between futures and basic assets in Russian and foreign financial markets, and on a reasonable degree of detail in multivariate GARCH modeling.

Keywords: futures, hedging, multivariate GARCH models, dynamic conditional correlations

JEL Classification: C32, C51, C53, G11, G15

Quantile

No. 9, July 2011

English page in the world wide web: <http://quantile.ru/eng>

Electronic mail address: quantile@quantile.ru

Access to the journal is free and unlimited

EDITOR

Stanislav Anatolyev
New Economic School (Moscow, Russia)

EDITORIAL COUNSEL

Victoria Zinde-Walsh
McGill University (Montréal, Canada)

Rustam Ibragimov
Harvard University (Cambridge, USA)

Anna Mikusheva
Massachusetts Institute of Technology (Cambridge, USA)

Alexey Onatsky
Cambridge University (Cambridge, UK)

Konstantin Tyurin
Indiana University (Bloomington, USA)

Alexander Tsyplakov
Novosibirsk State University (Novosibirsk, Russia)

Victor Chernozhukov
Massachusetts Institute of Technology (Cambridge, USA)

GUIDE TO AUTHORS

Manuscripts for publication in the “Articles” section should be submitted by electronic mail to the address submit@quantile.ru. Submitted work may be drawn from any applied field within the economics profession. The main requirement is correct usage of adequate econometric methodology. The manuscript should be written in Russian (for Russian-speaking persons) or in English (for all others) in the *Microsoft Word* or (preferably) *LaTeX* formats, and not exceed 30 double-spaced A4 pages. All submissions are subject to quality control by the editorial counsel and independent referees. A promising manuscript may be returned to the author(s) for polishing or rewriting. The editor also invites econometrics experts worldwide to contribute to the methodological sections of the journal.

Articles and methodological material published in “Quantile” do not transfer original copyright, neither in full, nor in part.

Solutions to the problems from the “Problems and Solutions” section and new problems can be sent to the address ps@quantile.ru.

Quantile

*international econometric journal
in Russian language*

**No. 9
July 2011**

IN THIS ISSUE

Econometric literacy: time series

Tsyplakov, Alexander. An introduction to state space modeling	1
Kheifets, Igor. Goodness-of-fit testing	25

Problems and solutions

Solutions 8.1, 8.2, 8.3	35
-------------------------	----

Articles: financial econometrics

Balaev, Alexey. Modeling multivariate parametric densities of financial returns	39
Kolokolov, Alexei. Futures hedging: Multivariate GARCH with dynamic conditional correlations	61