

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

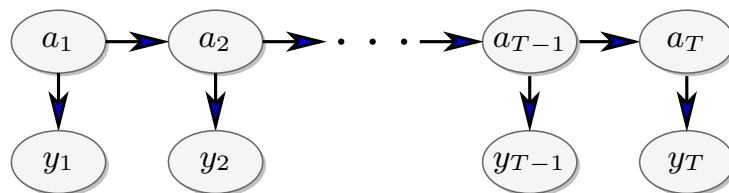


Рис. 1: Схема модели пространства состояний

*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).

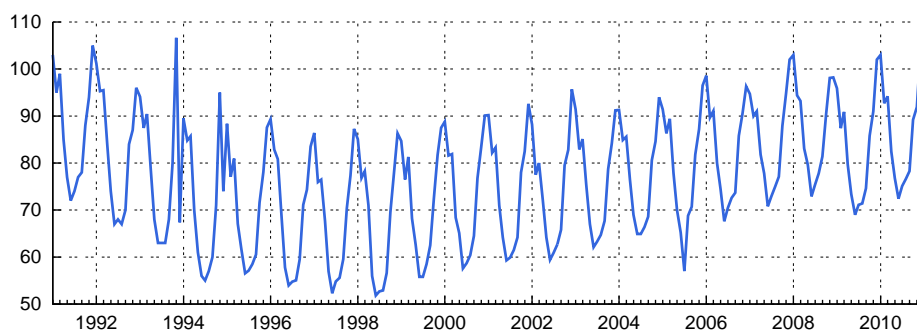


Рис. 2: Потребление электроэнергии в России, млрд кВт·ч.

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.

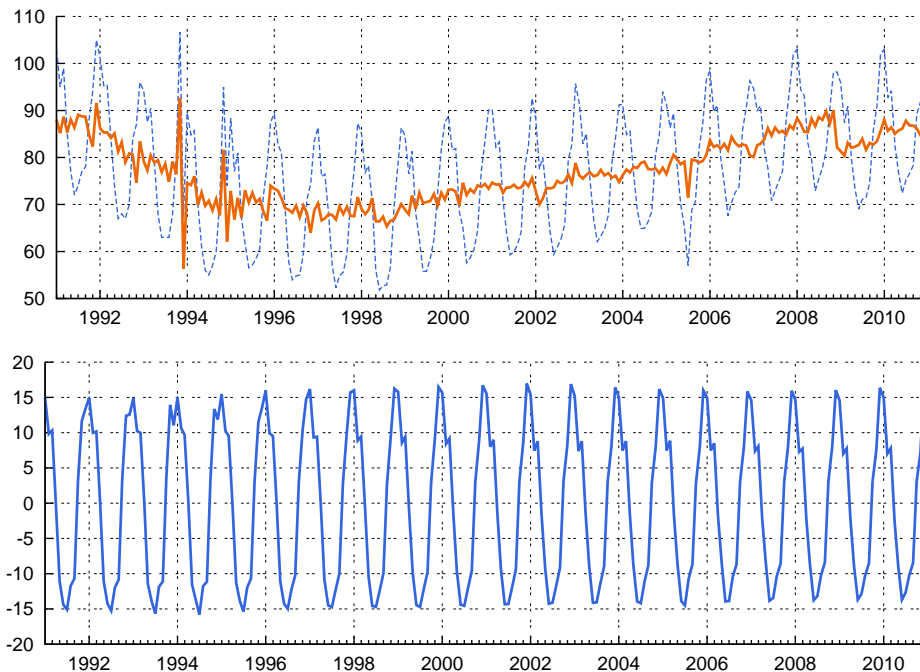


Рис. 3: Потребление электроэнергии в России, скорректированное на сезонность, и соответствующая сглаженная сезонная составляющая.

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.

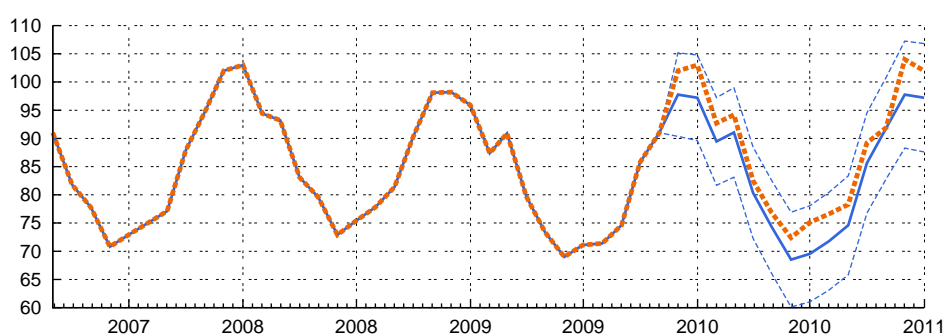


Рис. 4: Прогноз потребления электроэнергии в России.

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 ограничительными.

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

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

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 *двойной* (*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 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

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

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 мартингальное 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 мартингалное 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, мартингалное 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.