

КОВАРИАЦИОННОЕ ПРИБЛИЖЕНИЕ НЕЛИНЕЙНОЙ РЕГРЕССИИ

© 2016 г. Л. В. Лабунец^{1,2}, Е. Л. Лабунец³, Н. Л. Лебедева⁴¹Московский государственный технический университет им. Н.Э. Баумана,
Российская Федерация, 105005 Москва, 2-я Бауманская ул., 5²НОУ ВПО “Российский новый университет”
Российская Федерация, 105005 Москва, ул. Радио, 22³АКБ “РосЕвроБанк” (ОАО)
Российская Федерация, 119991 Москва, ул. Вавилова, 24⁴ОАО “Банк ВТБ”, Российская Федерация, 119121 Москва, ул. Плющиха, 37

E-mail: labunets@bmsu.ru

Поступила в редакцию 20.05.2015 г.

На основе ковариационного приближения многомерного вероятностного распределения получена модель нелинейной регрессии. Модель представлена разложением по базисным функциям в виде частных производных от логарифма совместной плотности распределения факторов. Весовыми коэффициентами разложения являются ковариации результирующей и объясняющих переменных. На конкретных примерах продемонстрирована эффективность байесовского приближения предложенной регрессионной модели в рамках описания распределения факторов конечной смесью эллипсоидально симметричных плотностей.

DOI: 10.7868/S003384941607007X

ВВЕДЕНИЕ

В настоящее время ядерные методы оценивания широко применяют в алгоритмах интеллектуального анализа данных (ИАД) [1, 2]. К перечню подобного рода задач ИАД следует отнести обнаружение скрытых закономерностей с помощью классической ядерной регрессии Надарая–Ватсона [3–6], локально взвешенной полиномиальной регрессии [7–10], моделирования и прогнозирования временных рядов [9, 11], кластеризации и распознавания образов средствами радиальных базисных функций [12, 13]. Ядерная интерполяция является эффективным инструментом эмпирической модовой декомпозиции изображений [14].

Важный раздел ядерных моделей ИАД составляют задачи непараметрического и полупараметрического оценивания одномерных и многомерных плотностей, среди которых следует отметить гистограмму сглаженную сдвигом (Average Shifted Histogram, ASH) [8], адаптивные ядерные оценки распределения [15–17] и ассоциированные с ними модели конечной смеси [1, 18, 19] стандартных распределений.

Характерной особенностью регрессии Надарая–Ватсона является необходимость выбора, в определенной мере субъективного, параметров ядерной интерполяции данных или применения

достаточно сложных по критерию вычислительных затрат методов оптимизации этих параметров [13].

Меру субъективности ядерной модели регрессии возможно снизить, на наш взгляд, применяя альтернативный подход, основанный на кумулянтном описании N -мерного вероятностного распределения для объясняющих $\vec{x} = (x_1, \dots, x_{N-1})^T$ и результирующей x_N переменных.

На практике основной, а иногда и единственной информацией о системе случайных величин (СВ) X_1, \dots, X_N , которую удастся надежно оценить по экспериментальным данным, является их ковариационная матрица $\mathbf{B}_N = \{b_{n,m}\}_{n=1, (N-1)}^{m=(n+1), N}$ и одномерные плотности распределения вероятностей (ПРВ) $\varphi_1^{X_1}(x_1), \dots, \varphi_1^{X_N}(x_N)$. В рамках указанных статистик рациональным для формирования нелинейной регрессионной модели

$$\begin{aligned} \tilde{x}_N(\vec{x}) &= \frac{\varepsilon_{N-1}(\vec{x})}{\varphi_{N-1}(\vec{x})}, \\ \varepsilon_{N-1}(\vec{x}) &= \int_{-\infty}^{+\infty} x_N \varphi_N(\vec{x}, x_N) dx_N \end{aligned} \quad (1)$$

является ковариационное приближение много-местной ПРВ [20]

$$\varphi_N(\bar{x}, x_N) = \sum_{k_{1,2}} \dots \sum_{k_{(N-1),N}} \frac{b_{1,2}^{k_{1,2}} \dots b_{(N-1),N}^{k_{(N-1),N}}}{k_{1,2}! \dots k_{(N-1),N}!} \times \prod_{m=1}^N \frac{d^{k_m}}{dx_m^{k_m}} \{\varphi_1^{X_m}(x_m)\}. \quad (2)$$

Здесь $\varphi_{N-1}(\bar{x})$ – маргинальное распределение факторов. Индексы суммирования степенного ряда (2) – это элементы верхней треугольной матрицы $\{k_{n,m}\}_{n=1,(N-1)}^{m=(n+1),N}$, принимающие целые неотрицательные значения из области

$$0 \leq \sum_{n=1}^{N-1} \sum_{m=n+1}^N k_{n,m} \leq K; k_m = \sum_{n=1}^{m-1} k_{n,m} + \sum_{n=m+1}^N k_{m,n}.$$

1. СТРУКТУРА КОВАРИАЦИОННОГО ПРИБЛИЖЕНИЯ НЕЛИНЕЙНОЙ РЕГРЕССИИ

Аддитивно-мультипликативная форма ряда (2) позволяет получить в явной форме зависимость условного математического ожидания результирующей переменной от статистических характеристик факторов. Будем исходить из предположения, что ПРВ целевой переменной x_N имеет K -й порядок соприкосновения с осью абсцисс. В этом случае соответствующие интегралы приобретают вид

$$\int_{-\infty}^{+\infty} x_N \frac{d^{k_N}}{dx_N^{k_N}} \{\varphi_1^{X_N}(x_N)\} dx_N = \begin{cases} a_N, & k_N = 0; \\ -1, & k_N = 1; \\ 0, & k_N \geq 2. \end{cases} \quad (3)$$

Здесь a_N – безусловное математическое ожидание СВ X_N . Последующее вычисление интеграла (1) с учетом (2) и (3) дает

$$\varepsilon_{N-1}(\bar{x}) = a_N \varphi_{N-1}(\bar{x}) - \sum_{n=1}^{N-1} b_{n,N} \theta_n(\bar{x} | \mathbf{B}_{N-1}), \quad (4)$$

где \mathbf{B}_{N-1} – ковариационная матрица факторов;

$$\theta_n(\bar{x} | \mathbf{B}_{N-1}) = \sum_{k_{1,2}} \dots \sum_{k_{(N-2),(N-1)}} \frac{b_{1,2}^{k_{1,2}} \dots b_{(N-2),(N-1)}^{k_{(N-2),(N-1)}}}{k_{1,2}! \dots k_{(N-2),(N-1)}!} \times \prod_{m=1}^{n-1} \frac{d^{k_m}}{dx_m^{k_m}} \{\varphi_1^{X_m}(x_m)\} \times \frac{d^{k_{n+1}}}{dx_{n+1}^{k_{n+1}}} \{\varphi_1^{X_m}(x_m)\} \prod_{m=n+1}^{N-1} \frac{d^{k_m}}{dx_m^{k_m}} \{\varphi_1^{X_m}(x_m)\}.$$

Несложно заметить, что последнее равенство представляет собой частную производную сов-

местной ПРВ объясняющих переменных по фактору x_n , т.е.

$$\theta_n(\bar{x} | \mathbf{B}_{N-1}) = \frac{\partial}{\partial x_n} \{\varphi_{N-1}(\bar{x})\}.$$

В результате агрегирования (1) и (4) получают ковариационное приближение модели нелинейной регрессии

$$\tilde{x}_N(\bar{x}) = a_N + \sum_{n=1}^{N-1} b_{n,N} \frac{\partial}{\partial x_n} (-\ln \{\varphi_{N-1}(\bar{x})\}) = a_N + \bar{b}_N^T \nabla_{\bar{x}} (-\ln \{\varphi_{N-1}(\bar{x})\}), \quad (5)$$

где $\bar{b}_N = (b_{1,N}, \dots, b_{(N-1),N})^T$ – N -й вектор-столбец ковариаций факторов \bar{x} с целевой переменной x_N в матрице \mathbf{B}_N .

Закономерно, что полученная регрессионная модель имеет явно выраженную сетевую архитектуру, подобную нейронной сети радиальных базисных функций [13]. На входы сети поступают стандартный сигнал смещения (+1) и факторные переменные x_1, \dots, x_{N-1} . Роль “радиальных” нейронов скрытого слоя сети, обеспечивающих нелинейное преобразование и отображение на пространство информативных признаков, выполняют частные производные логарифма совместной ПРВ объясняющих переменных. Личнейный нейрон выходного слоя сети суммирует реакции скрытых нейронов, веса которых a_N и $b_{1,N}, \dots, b_{(N-1),N}$.

2. БАЙЕСОВСКАЯ МОДЕЛЬ НЕЛИНЕЙНОЙ РЕГРЕССИИ

Дальнейшая конкретизация модели нелинейной регрессии (5) возможна на основе аппроксимации совместной ПРВ факторов $\varphi_{N-1}(\bar{x})$ конечной смесью

$$\Phi_{N-1}(\bar{x}) = \sum_{k=1}^K p_k \Phi_{(N-1),k}(\bar{x} | \bar{A}_k, \mathbf{C}_k), \quad (6)$$

$$\sum_{k=1}^K p_k = 1$$

эллипсоидально симметричных парциальных распределений

$$\Phi_{(N-1),k}(\bar{x} | \bar{A}_k, \mathbf{C}_k) = \frac{f\{r^2(\bar{x} | \bar{A}_k, \mathbf{C}_k)\}}{S_1 \nu_{N-2} \sqrt{\det(\mathbf{C}_k)}}. \quad (7)$$

Здесь $r^2(\bar{x} | \bar{A}_k, \mathbf{C}_k) = (\bar{x} - \bar{A}_k)^T \mathbf{C}_k^{-1} (\bar{x} - \bar{A}_k)$ – квадрат расстояния Махаланобиса; p_k , $\bar{A}_k = (a_{k,1}, \dots, a_{k,N-1})^T$ и $\mathbf{C}_k = \{c_{k,n,m}\}_{n=1,(N-1)}^{m=1,(N-1)}$ – априорная

вероятность, вектор математического ожидания и ковариационная матрица k -го парциального распределения; $S_1 = 2\sqrt{\pi^{N-1}}/\Gamma\{(N-1)/2\}$ – площадь поверхности единичной сферы в $(N-1)$ -мерном пространстве факторов; $f(r^2)$ – одномерное, монотонно убывающее по мере $r \rightarrow \infty$ радиальное распределение с конечным моментом $v_{N-2} = \int_0^\infty r^{N-2} f(r^2) dr$.

На практике для моделирования эллипсоидально симметричных распределений (7) широко применяют три радиальные базисные функции [12, 19, 21, 22]:

Гаусса –

$$f(r^2) = \exp\left(-\frac{r^2}{2}\right), v_{N-2} = \frac{1}{2}\sqrt{2^{N-1}}\Gamma\left(\frac{N-1}{2}\right);$$

Пирсона с параметром $\gamma > 0$ –

$$f(r^2) = \begin{cases} \left(1 - \frac{r^2}{2\gamma + N - 1}\right)^{\gamma-1}, & r^2 \leq 2\gamma + N - 1, \\ 0, & r^2 > 2\gamma + N - 1, \end{cases}$$

$$v_{N-2} = \frac{1}{2}\sqrt{(2\gamma + N - 1)^{N-1}}B\left(\gamma, \frac{N-1}{2}\right);$$

Стьюдента с параметром $\gamma > N - 2$ –

$$f(r^2) = \left(1 + \frac{r^2}{\gamma + 2 - N}\right)^{-\frac{\gamma+1}{2}},$$

$$v_{N-2} = \frac{1}{2}\sqrt{(\gamma + 2 - N)^{N-1}}B\left(\frac{\gamma + 2 - N}{2}, \frac{N-1}{2}\right),$$

где $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$ – бета-функция.

Анализ частных производных квадрата расстояния Махаланобиса k -го парциального распределения факторов удобно выполнять в терминах матрицы точности $C_k^{-1} = \{c_k^{(n,m)}\}_{n=1, (N-1)}^{m=1, (N-1)}$, т.е. обратной к ковариационной матрице C_k , а именно

$$r^2(\bar{x}|\bar{A}_k, C_k) = \sum_{n=1}^{N-1} c_k^{(n,n)}(x_n - a_{k,n})^2 + 2 \sum_{n=1}^{N-2} \sum_{m=n+1}^{N-1} c_k^{(n,m)}(x_n - a_{k,n})(x_m - a_{k,m}).$$

(Далее, если нет необходимости, не приводим зависимость метрики Махаланобиса $r(\bar{x}|\bar{A}_k, C_k)$ от ста-

тистик положения \bar{A}_k и масштаба C_k .) Соответствующие частные производные приобретают вид

$$\frac{\partial}{\partial x_n} \{r^2(\bar{x})\} = \sum_{m=1}^{N-1} c_k^{(m,n)}(x_m - a_{k,m}), \quad n = 1, 2, \dots, (N-1). \quad (8)$$

Дифференцирование логарифма конечной смеси (6) с учетом представленных выше радиальных базисных функций и формулы (8) дает байесовское приближение для ковариационной модели (5) нелинейной регрессии:

$$\tilde{x}_N(\bar{x}) = a_N + \sum_{k=1}^K \Pr(k|\bar{x}) R_{k,N}(\bar{x}), \quad (9)$$

где $\Pr(k|\bar{x}) = p_k \Phi_{(N-1),k}(\bar{x}|\bar{A}_k, C_k) / \Phi_{N-1}(\bar{x})$ – апостериорная вероятность ассоциации текущих значений факторов с k -м парциальным ядром $R_{k,N}(\bar{x})$. Тип ядра определяется выбором радиальных базисных функций.

В случае модели Гаусса ядро имеет вид

$$R_{k,N}(\bar{x}) = S_{k,N}(\bar{x}); \quad (10)$$

в случае модели Пирсона –

$$R_{k,N}(\bar{x}) = \begin{cases} \frac{2(\gamma-1)S_{k,N}(\bar{x})}{2\gamma + N - 1} \times \\ \times \left\{1 - \frac{r^2(\bar{x})}{2\gamma + N - 1}\right\}^{-1}, & r^2(\bar{x}) \leq 2\gamma + N - 1, \\ 0, & r^2(\bar{x}) > 2\gamma + N - 1; \end{cases}$$

в случае модели Стьюдента –

$$R_{k,N}(\bar{x}) = \frac{(\gamma+1)S_{k,N}(\bar{x})}{\gamma + 2 - N} \left\{1 + \frac{r^2(\bar{x})}{\gamma + 2 - N}\right\}^{-1}.$$

Здесь локальная аппроксимирующая гиперплоскость

$$S_{k,N}(\bar{x}) = \sum_{m=1}^{N-1} W_{k,N}^{(m)}(x_m - a_{k,m}), \\ W_{k,N}^{(m)} = \sum_{n=1}^{N-1} c_k^{(m,n)} b_{n,N}$$

ассоциирована с k -м парциальным ядром. Выражение для этой плоскости удобно записать в векторной форме

$$S_k(\bar{x}|\bar{A}_k, C_k) = \bar{b}_N^T C_k^{-1} (\bar{x} - \bar{A}_k). \quad (11)$$

Факторизация матрицы точности

$$C_k^{-1} = U_k \Lambda_k^{-1/2} \Lambda_k^{-1/2} U_k^T = (\Lambda_k^{-1/2} U_k^T) (U_k \Lambda_k^{-1/2})$$

позволяет представить квадратичную форму в равенстве (11) в терминах преобразования, аналогичного декоррелирующему [23]

$$\begin{aligned} S_k(\bar{x}|\bar{A}_k, \mathbf{C}_k) &= \\ &= (\mathbf{U}_k \Lambda_k^{-1/2} \bar{b}_N)^T \{ \mathbf{U}_k \Lambda_k^{-1/2} (\bar{x} - \bar{A}_k) \} = \\ &= (\Lambda_k^{-1/2} \bar{b}_N)^T \mathbf{U}_k^T \mathbf{U}_k \{ \Lambda_k^{-1/2} (\bar{x} - \bar{A}_k) \} = \bar{w}_{k,N}^T \bar{y}_k. \end{aligned} \quad (12)$$

Здесь $\Lambda_k^{1/2}$ и \mathbf{U}_k – матрицы сингулярных чисел и ортонормированных собственных векторов ковариационной матрицы k -го парциального распределения; $\bar{w}_{k,N} = \Lambda_k^{-1/2} \bar{b}_N$ – нормированный вектор-столбец ковариаций целевой переменной x_N с факторами \bar{x} ; $\bar{y}_k = \Lambda_k^{-1/2} (\bar{x} - \bar{A}_k)$ – вектор-столбец центрированных и нормированных факторов.

Далее подробнее проанализируем ковариационное приближение нелинейной регрессии для популярного в практических приложениях случая полигауссовского описания совместной ПРВ факторов.

3. ПОЛИГАУССОВСКАЯ КОВАРИАЦИОННАЯ МОДЕЛЬ НЕЛИНЕЙНОЙ РЕГРЕССИИ

В случае полигауссовской аппроксимации ПРВ факторов байесовское приближение модели нелинейной регрессии получим в результате последовательной подстановки формул (12) и (10) в выражение (9)

$$\tilde{x}_N(\bar{x}) = a_N + \sum_{k=1}^K \Pr(k|\bar{x}) \bar{w}_{k,N}^T \bar{y}_k. \quad (13)$$

Геометрический смысл этой модели удобно анализировать в N -мерном информативном пространстве факторов и результирующей переменной. В текущей точке (\bar{x}, x_N) регрессионная поверхность $\tilde{x}_N(\bar{x})$ представляет собой локальную гиперплоскость. Смещение гиперплоскости, т.е. расстояние от начала координат информативного пространства до плоскости по нормали к ней составляет величину

$$d_N(\bar{x}) = \frac{a_N - \bar{b}_N^T \sum_{k=1}^K \Pr(k|\bar{x}) \Lambda_k^{-1} \bar{A}_k}{\left\| \bar{b}_N^T \sum_{k=1}^K \Pr(k|\bar{x}) \Lambda_k^{-1} \right\|}.$$

Поворот гиперплоскости определяется направлением ее нормали

$$\bar{u}_N(\bar{x}) = \frac{\bar{b}_N^T \sum_{k=1}^K \Pr(k|\bar{x}) \Lambda_k^{-1}}{\left\| \bar{b}_N^T \sum_{k=1}^K \Pr(k|\bar{x}) \Lambda_k^{-1} \right\|}.$$

Иными словами, средневзвешенные положение и ориентация локальной аппроксимирующей гиперплоскости определяются уровнями надежности $\Pr(k|\bar{x})$, с которыми текущие значения факторов \bar{x} ассоциируют с k -м парциальным распределением. Таким образом, байесовское взвешивание обеспечивает адаптацию параметров линейной регрессии к локальным областям подпространства факторов.

Важно отметить, что полигауссовское приближение (13) ковариационной модели нелинейной регрессии согласуется с методологией нелинейного метода главных компонент (Nonlinear Principal Component Analysis, NLPKA) [24, 25]. Наряду с иерархической NLPKA-моделью анализа данных, представленной в [25, разд. 2.3], рациональной нейросетевой реализацией байесовской модели (13) является, на наш взгляд, динамическая структура ассоциативной машины в виде архитектуры сети смещения мнений экспертов (Mixture of Experts, ME) [13] (рис. 1). Ассоциативная машина содержит K сетей экспертов и агрегирующий модуль в виде сети шлюза. Каждый эксперт предлагает модель линейной регрессии $\bar{w}_{k,N}^T \bar{y}_k$, $k = 1, 2, \dots, K$ наилучшую в k -й локальной области пространства факторов \bar{x} . Сеть шлюза содержит K нелинейных нейронов, оценивающих степени достоверности $\Pr(k|\bar{x})$ мнений экспертов. Выходной линейный нейрон машины агрегирует мнения экспертов, формируя средневзвешенную оценку (13) регрессионной поверхности. Ассоциативная машина, как известно [13], является универсальным аппроксиматором.

В последующих вычислительных экспериментах проанализируем результаты применения регрессионной модели (13) для конкретных данных.

4. ОДНОМЕРНАЯ РЕГРЕССИЯ

Для двумерных данных ($N = 2$) полигауссовское приближение (13) ковариационной модели нелинейной регрессии приобретает вид

$$\tilde{x}_2(x_1) = a_2 + \sum_{k=1}^K \Pr(k|x_1) \frac{b_{1,2}}{\sigma_{k,1}} \frac{x_1 - a_{k,1}}{\sigma_{k,1}}, \quad (14)$$

где $\sigma_{k,1} = \sqrt{\lambda_{k,1}}$ – среднее квадратичное отклонение k -й парциальной гауссовской ПРВ фак-

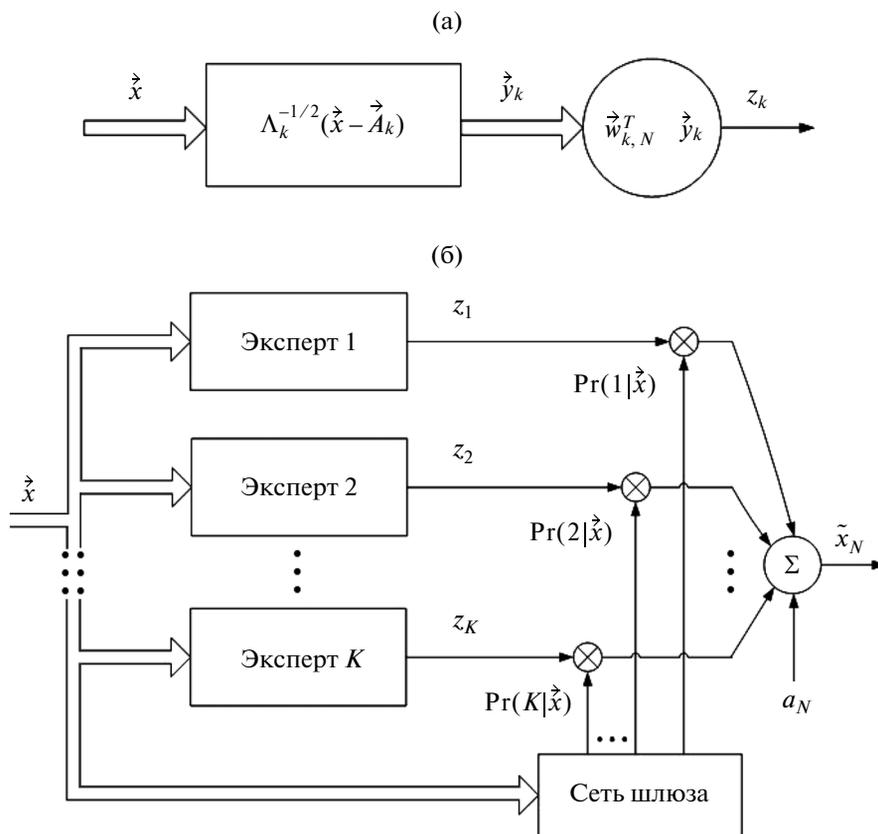


Рис. 1. Сеть смещения мнений экспертов: а) сеть k -го эксперта; б) МЕ-сеть.

тора x_1 . Очевидно, что при $K = 1$ равенство (14) эквивалентно уравнению одномерной линейной регрессии.

Нелинейная регрессия (14) тестировалась на примере задачи моделирования и прогнозирования временного ряда (ВР) $v(m)$, $m = 1, 2, \dots, 144$ объемов ежемесячных авиаперевозок пассажиров [26] в тысячах человек в течение 12 лет. Приближение (14) применялось для формирования нелинейной регрессионной зависимости $u(m) = \theta\{u(m-12)\} + e(m)$. Здесь $e(m)$ – ошибка моделирования данных;

$$u(m) = \{v(m) - a(m)\} / s(m)$$

– центрированный и нормированный ВР, полученный с помощью полиномиальных оценок тренда

Таблица 1. Параметры нелинейной регрессии авиаперевозок

k	p_k	$a_{k,1}$	$\sigma_{k,1}$	$\sigma_k^{(опт)}$
1	0.65	-0.1432	0.7625	1.01
2	0.35	0.3563	0.9156	0.97

$a(m) = 116.97 + 1.268m + 0.0134m^2$ и меры волатильности $s(m) = 11.411 + 0.2226m + 0.0017m^2$. Иными словами, фактор x_1 и целевую переменную x_2 ассоциировали с лаговым $u(m-12)$ и текущим $u(m)$ отсчетами ВР соответственно. Ниже представлены выборочные оценки безусловных статистик этих переменных, необходимые для реализации модели (14):

Статистика	a_2	$b_{1,2}$
Оценка	-0.003751	0.909439

Выборочная оценка ПРВ фактора x_1 в виде гистограммы сглаженной сдвигом гауссовского ядра представлена на рис. 2 (кривая 1). Гистограмма была получена по пяти разрядным интервалам и пяти суженным подынтервалам. В дальнейшем АШН-оценку аппроксимировали смесью двух гауссоид (кривая 2) с помощью модифицированного ЕМ-алгоритма [19, 27, 28]. Параметры данной полигауссовской смеси (6) представлены в табл. 1. Этим параметрам соответствует завышенное значение $\sigma = 0.051$ для среднего квадратич-

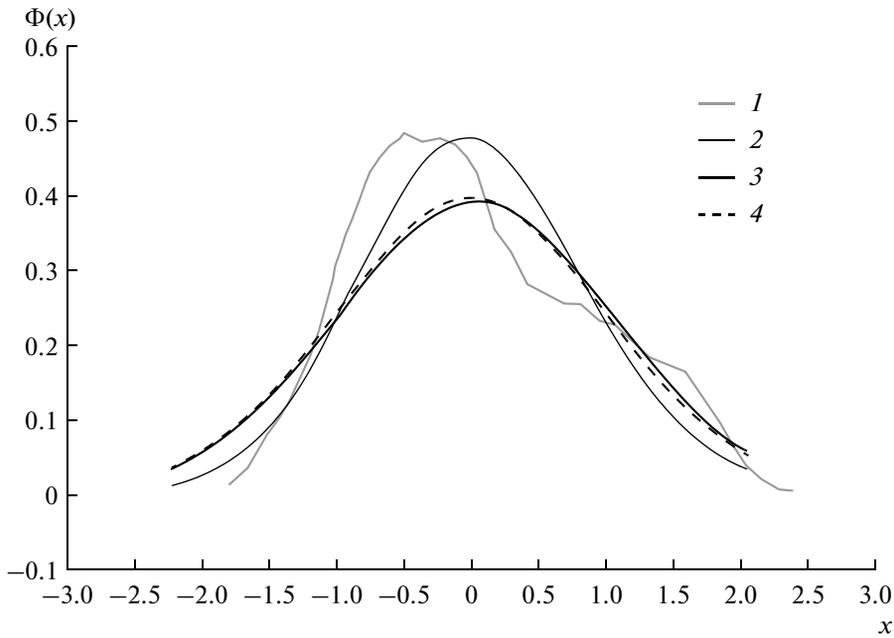


Рис. 2. Оценки ПРВ фактора: 1 – ASH-оценка распределения; 2 – полигауссовская аппроксимация; 3 – оптимальная полигауссовская модель; 4 – гауссоида соответствующая линейной регрессии.

ного отклонения (СКО) ошибки $e(m)$ моделирования данных.

Последующая настройка масштабов $\sigma_{1,1}$ и $\sigma_{2,1}$ регрессионной модели (14) обеспечила наименьшую величину СКО $\sigma = 0.0375$. Оптимальные значения $\sigma_k^{(опт)}$ этих параметров также приведены в табл. 1. Двухкомпонентная полигауссовская модель ПРВ фактора x_1 , отвечающая оптимальным значениям параметров регрессионной зависимости (14), представлена на рис. 2 (кривая 3). Эта ПРВ идеально согласуется с гауссоидой $N(-0.002; 1.008)$ (см. рис. 2 кривая 4), соответствующей модели линейной регрессии. Зависимость апостериорных вероятностей $\Pr(k|x_1)$ классов $k = 1, 2$ от значений фактора x_1 для оптимальной двухкомпонентной полигауссовской смеси (6) дана на рис. 3. Диаграмма рассеяния факторной $u(m-12)$ и целевой $u(m)$ переменных (рис. 4) демонстрирует согласованность моделей линейной регрессии и полигауссовского приближения нелинейной регрессии. Результаты моделирования централизованного и нормированного ВР $u(m)$ перевозок авиапассажиров и его прогнозирования на 12 месяцев вперед с помощью оптимальной регрессионной зависимости (14) представлены на рис. 5.

5. ДИНАМИЧЕСКАЯ ОДНОМЕРНАЯ РЕГРЕССИЯ

Полигауссовское приближение (13) ковариационной модели нелинейной регрессии рационально применять для моделирования динамической зависимости целевой переменной $x_N(t)$ от факторов $\bar{x}(t)$

$$\tilde{x}_N(t) = \sum_{j=1}^J g_j(t) \tilde{x}_N^{(j)}\{\bar{x}(t)\}, \quad \sum_{j=1}^J g_j(t) = 1, \quad (15)$$

где $t_0 \leq t \leq t_J$ – время; $0 \leq g_j(t) \leq 1$ – субъективные вероятности, с которыми локальные нелинейные регрессионные модели

$$\tilde{x}_N^{(m)}\{\bar{x}(t)\} = a_N^{(m)} + \sum_{k=1}^K \Pr_m\{k|\bar{x}(t)\} \bar{w}_{m,k,N}^T \bar{y}_{m,k}(t), \quad m = 1, 2, \dots, J$$

адекватно аппроксимируют целевую переменную $x_N(t)$ на текущем временном интервале $t_{j-1} \leq t \leq t_j$; $\bar{y}_{m,k}(t) = \Lambda_{m,k}^{-1/2} \{\bar{x}(t) - \bar{A}_{m,k}\}$ – центрированный и нормированный вектор-столбец факторов на m -м временном интервале. Динамическую зависимость (15) целевой переменной от факторов уместно, на наш взгляд, назвать гирляндой Гаусса.

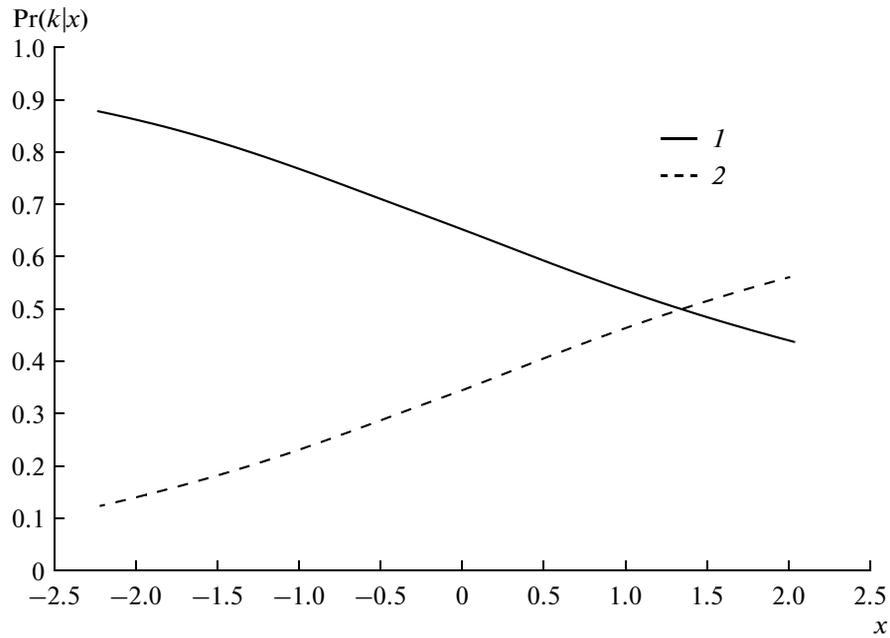


Рис. 3. Апостериорные вероятности классов оптимальной полигауссовской модели ПРВ фактора: $k = 1$ (1) и 2 (2).

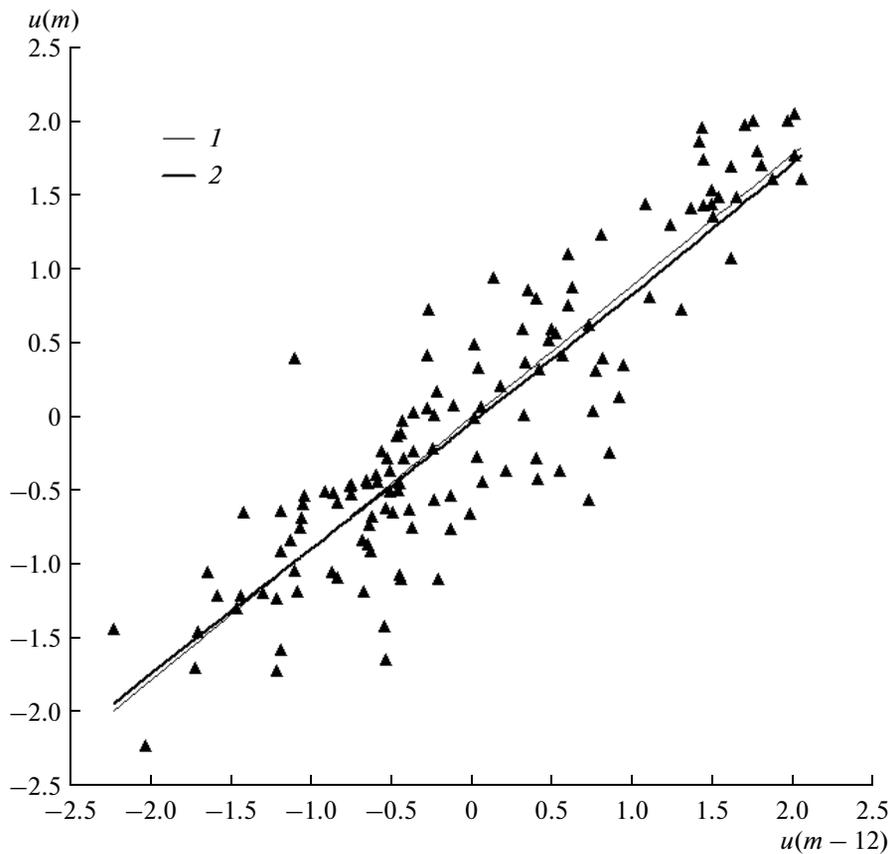


Рис. 4. Диаграмма рассеяния факторной и целевой переменных: 1 – линейная регрессия; 2 – полигауссовская модель нелинейной регрессии.

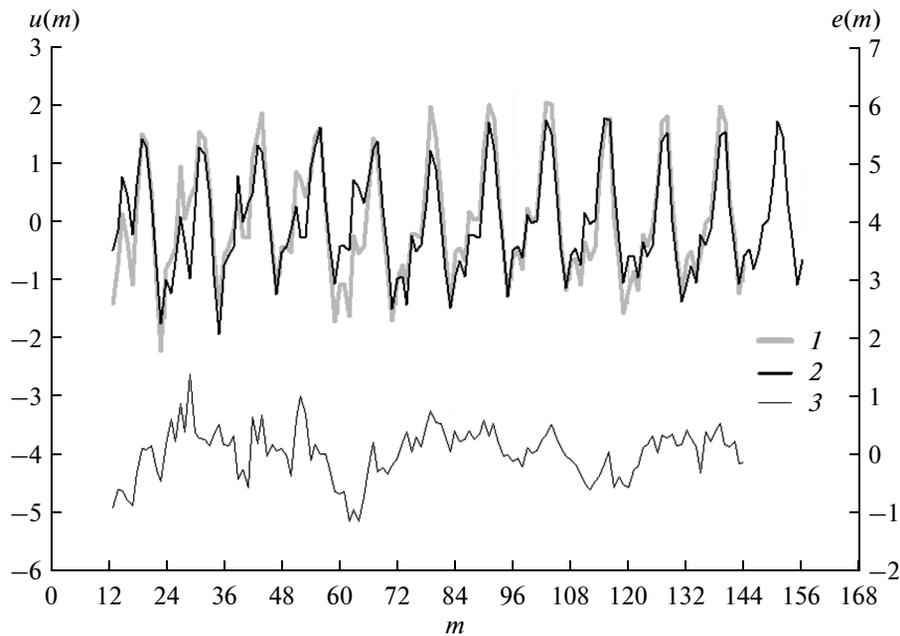


Рис. 5. Полигауссовская ковариационная модель нелинейной регрессии: 1 – центрированный и нормированный ВР $u(m)$; 2 – прогноз на 12 месяцев; 3 – ошибка $e(m)$ моделирования данных.

Такого рода регрессионная модель хорошо согласуется с архитектурой сети иерархического смещения мнений экспертов (Hierarchical Mixture of Experts, НМЕ) [13] (рис. 6). Динамическая НМЕ-сеть состоит из двух уровней иерархии. Нижний содержит J МЕ-сетей экспертов, формирующих нелинейные регрессионные гиперповерхности $\tilde{x}_{j,N}\{\bar{x}(t)\}$, $j = 1, 2, \dots, J$ в пространстве факторов \bar{x} . Каждая из МЕ-сетей адекватно аппроксимирует данные в фиксированном интервале времени $[t_{j-1}, t_j]$. Сеть шлюза верхнего уровня иерархии агрегирует выходные реакции МЕ-сетей, реализуя мягкий переход от модели к модели с помощью их взвешивания субъективными вероятностями $g_j(t)$ по временным интервалам. Здесь также уместна аналогия с представленными в [25, гл. 8] автоассоциативными моделями NLPCA-подхода к анализу данных.

Рассмотренный выше подход применялся для моделирования динамической регрессионной зависимости курса доллара США, устанавливаемого Центральным банком Российской Федерации (Банк России), от котировок нефти марки Brent. Диаграмма рассеяния этих активов в период с 04.04.12 по 16.01.13 представлена на рис. 7. Очевидно, что практически важным является вопрос: в какой мере уровень изменчивости курса доллара, устанавливаемый Банком России, согласуется с динамикой цен на нефть?

Ответ на поставленный вопрос дает двумерная ($N = 2$) гирлянда Гаусса

$$\tilde{x}_2(t) = \sum_{j=1}^3 g_j(t) \tilde{x}_2^{(j)}\{x_1(t)\}, \quad \sum_{j=1}^3 g_j(t) = 1, \quad (16)$$

$$\tilde{x}_2^{(j)}\{x_1(t)\} = a_2^{(j)} + \sum_{k=1}^{K_j} \text{Pr}_j\{k|x_1(t)\} \frac{b_{1,2}^{(j)} x_1(t) - a_{k,1}^{(j)}}{\sigma_{k,1}^{(j)}} \frac{1}{\sigma_{k,1}^{(j)}}$$

Гирлянда сформирована для трех временных интервалов ($J = 3$): первый с 04.04.12 по 22.06.12, второй с 22.06.12 по 23.08.12 и третий с 23.08.12 по 16.01.13. В качестве субъективных вероятностей $g_j(t)$ выбирались трапециевидные модели с линейными участками переходных зон от нуля до единицы продолжительностью семь дней, симметрично расположенных относительно дат 22.06.12 и 23.08.12. Выборочные оценки описательных статистик биржевых активов для указанных выше временных интервалов приведены в табл. 2.

В качестве оценок ПРВ фактора x_1 для трех временных интервалов применялись гистограммы сглаженные сдвигом трижды взвешенного ядра Епанечникова. Гистограммы формировались по пяти разрядным интервалам и двумстам суженным подынтервалам. На рис. 8 представлена ASH-оценка (кривая 1) для первого временного интервала. Затем ASH-оценки аппроксимировали смесью трех гауссоид (кривая 2) с помощью модифицированного EM-алгоритма. Парамет-

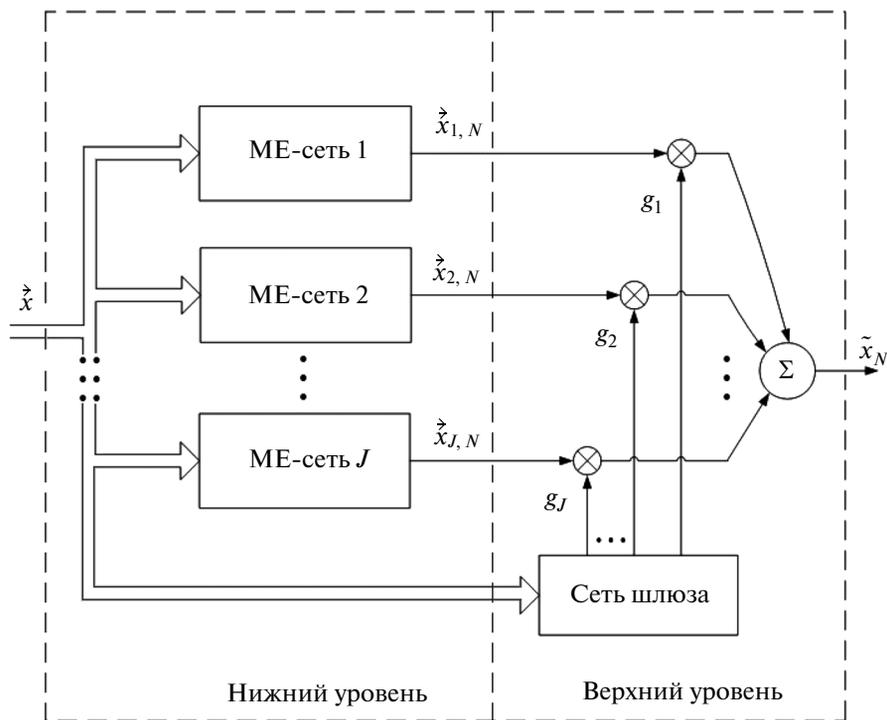


Рис. 6. Сеть иерархического смещения мнений экспертов.

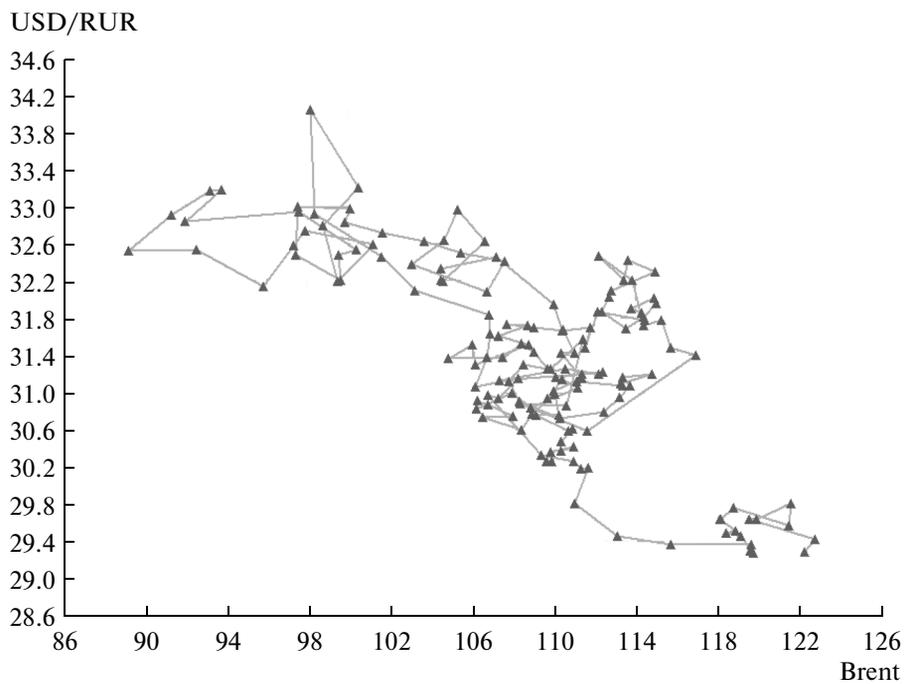


Рис. 7. Диаграмма рассеяния котировок нефти Brent и курса валютной пары USD/RUR.

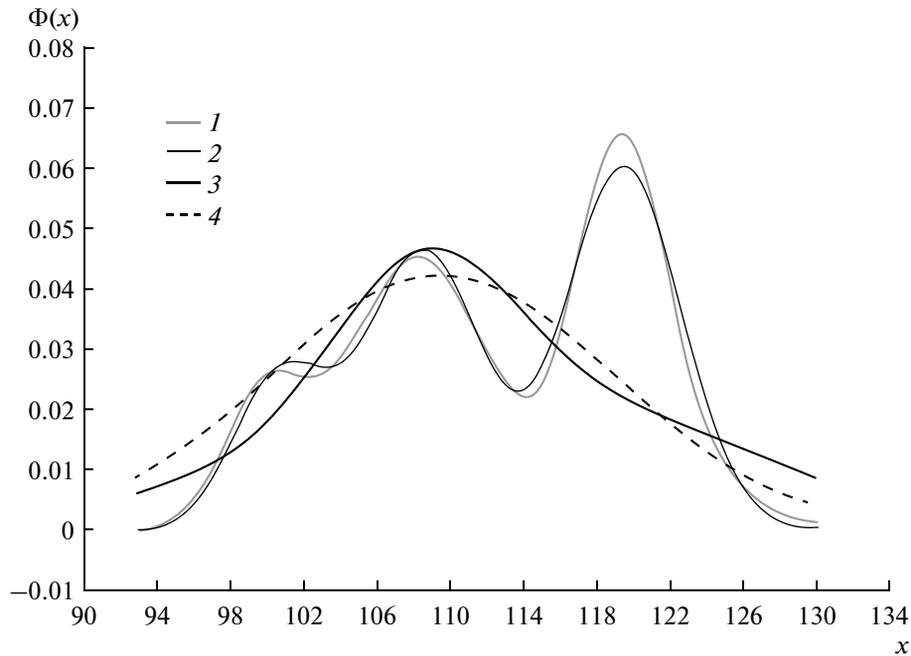


Рис. 8. Оценки ПРВ котировок нефти Brent: 1 – ASH-оценка распределения; 2 – полигауссовская аппроксимация; 3 – оптимальная полигауссовская модель; 4 – гауссоида соответствующая линейной регрессии.

рам этих полигауссовских описаний отвечают завышенные значения СКО ошибки аппроксимации данных $\sigma^{(1)} = 0.6678$, $\sigma^{(2)} = 0.1951$ и $\sigma^{(3)} = 0.0675$ для трех временных интервалов соответственно.

Последующая настройка параметров регрессионных моделей (16) для трех временных интервалов обеспечила наименьшие величины СКО $\sigma^{(1)} = 0.0538$, $\sigma^{(2)} = 0.0412$ и $\sigma^{(3)} = 0.0549$ соответственно. Квазиоптимальные значения параметров этих моделей приведены в табл. 3, а квазиоптимальная трехкомпонентная полигауссовская модель ПРВ фактора x_1 для первого временного интервала представлена на рис. 8 (кривая 3). Эта ПРВ хорошо согласуется с гауссоидой $N(109.62; 9.42)$ (кривая 4), отвечающей модели линейной регрессии. Рис. 9 иллюстрирует зависимость апостериорных вероятностей $\text{Pr}_1\{k|x_1(t)\}$ ($k = 1, 2, 3$) от значений фактора x_1 на первом временном интервале для квазиоптимальной трехкомпонентной полигауссовской смеси (16). Модели линейной регрессии и ковариационного приближения нелинейной регрессии для первого временного интервала представлены на рис. 10. Результирующую гирлянду Гаусса, аппроксимирующую в среднем динамику поведения диаграммы рассеяния активов Brent–USD/RUR, иллюстрирует рис. 11. Соотношение временных рядов в виде курса доллара США, устанавливаемого Банком России, и его регрессионной модели в за-

висимости от дневных котировок нефти марки Brent демонстрирует рис. 12. Важно отметить, что модель обеспечивает меньший уровень изменчивости котировок валютной пары по сравнению с курсом Банка России.

6. ДВУМЕРНАЯ РЕГРЕССИЯ

Для трехмерных данных ($N = 3$) полигауссовское приближение (13) ковариационной модели нелинейной регрессии приобретает вид

$$\tilde{x}_3(x_1, x_2) = a_3 + \sum_{k=1}^K \frac{\text{Pr}(k|x_1, x_2)}{1 - \rho_{k,1,2}^2} \left\{ \left(\frac{b_{1,3}}{\sigma_{k,1}} - \rho_{k,1,2} \frac{b_{2,3}}{\sigma_{k,2}} \right) \times \right. \\ \left. \times \frac{x_1 - a_{k,1}}{\sigma_{k,1}} + \left(\frac{b_{2,3}}{\sigma_{k,2}} - \rho_{k,1,2} \frac{b_{1,3}}{\sigma_{k,1}} \right) \frac{x_2 - a_{k,2}}{\sigma_{k,2}} \right\}, \quad (17)$$

Таблица 2. Описательные статистики котировок нефти Brent и валютной пары USD/RUR для трех временных интервалов

Интервал	j	$a_1^{(j)}$	$a_2^{(j)}$	$b_{1,2}^{(j)}$
04.04.12–22.06.12	1	109.62	30.50	-12.3572
22.06.12–23.08.12	2	104.46	32.39	-2.6444
23.08.12–16.01.13	3	110.52	31.21	2.0627

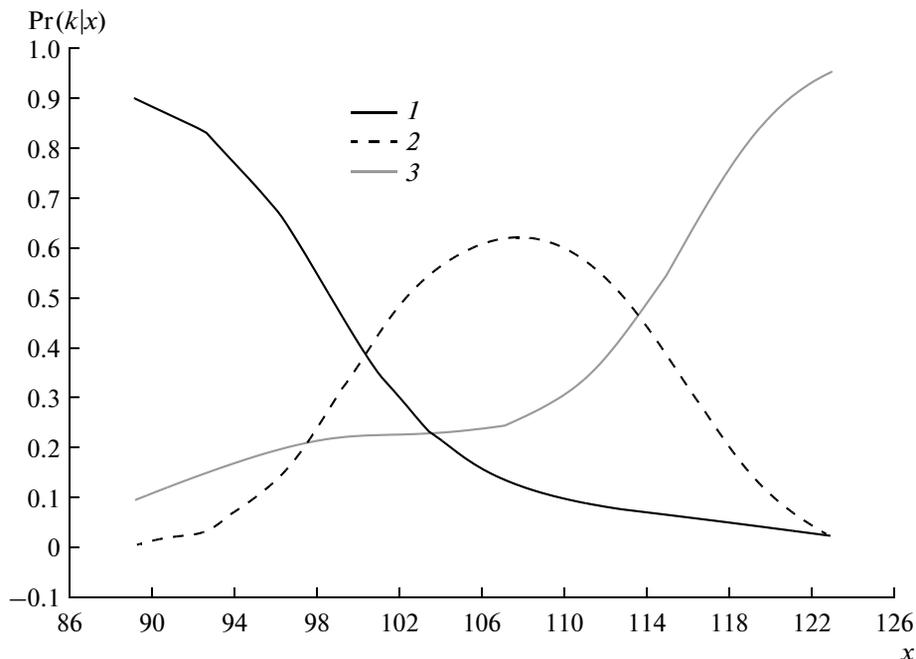


Рис. 9. Апостериорные вероятности классов оптимальной полигауссовской модели ПРВ котировок нефти Brent: $k = 1$ (1), 2 (2) и 3 (3).

где $\sigma_{k,1}$, $\sigma_{k,2}$ и $\rho_{k,1,2}$ – средние квадратичные отклонения и коэффициент корреляции k -й парциальной гауссовской ПРВ факторов x_1 и x_2 . Очевидно, что при $K = 1$ равенство (17) эквивалентно уравнению двумерной линейной регрессии.

Байесовское приближение (17) применялось для формирования нелинейной регрессионной

зависимости целевого ВР $v(m)$, $m = 1, 2, \dots, 252$ объемов ежемесячных продаж в США транспортных средств (рис. 13, кривая 3) от двух факторных ВР – продаж продукции химического производства (кривая 1) и нефтепродуктов (кривая 2) в период с января 1971 г. по декабрь 1991 г. [29]. Ниже приведены выборочные оценки безусловных статистик ВР, необходимые для реализации модели (17):

Таблица 3. Оптимальные параметры нелинейной регрессии нефти Brent на валютную пару USD/RUR

k	$p_k^{(j)}$	$a_{k,1}^{(j)}$	$\sigma_{k,1}^{(j)}$
$j = 1, \sigma^{(1)} = 0.0538$			
1	0.1657	100.84	9.0
2	0.3616	108.60	5.0
3	0.4727	117.60	10.0
$j = 2, \sigma^{(2)} = 0.0412$			
1	0.2689	99.27	6.0
2	0.3633	105.16	4.4
3	0.3678	108.0	8.0
$j = 3, \sigma^{(3)} = 0.0549$			
1	0.2135	108.11	1.300
2	0.4225	110.24	1.061
3	0.3640	112.50	1.467

Статистика	a_3	$b_{1,3}$	$b_{2,3}$
Оценка	11.743	13.390418	28.437772

Диаграмма рассеяния объясняющих ВР, параметром которой является время, демонстрирует наличие явно выраженной кластерной структуры в динамике данных (рис. 14). Поэтому на первом этапе формирования ковариационной модели нелинейной регрессии (17) были выбраны пять ($K = 5$) классов данных, соответствующие пяти временным интервалам: 01.01.71–31.12.74; 01.01.75–30.06.79; 01.07.79–31.08.83; 01.09.83–31.12.87; 01.01.88–31.12.91 с количеством отсчетов ВР 48, 54, 50, 52, 48 в каждом.

Параметры двумерной полигауссовской аппроксимации (6) ПРВ факторов (рис. 15а), полученные с помощью EM-алгоритма, приведены в табл. 4. Значения СКО ошибки $e(m)$ аппроксимации целевого ВР регрессионной моделью (17) свидетельствуют о неприемлемо больших погрешностях моделирования данных.

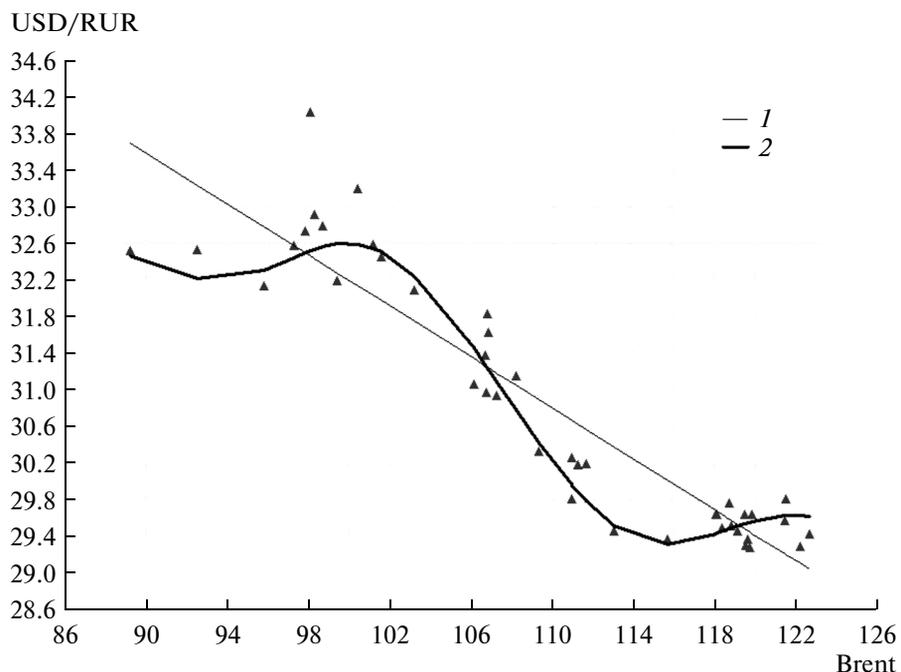


Рис. 10. Диаграмма рассеяния Brent–USD/RUR в период с 04.04.12 по 22.06.12: 1 – линейная регрессия; 2 – полигауссовская модель нелинейной регрессии.

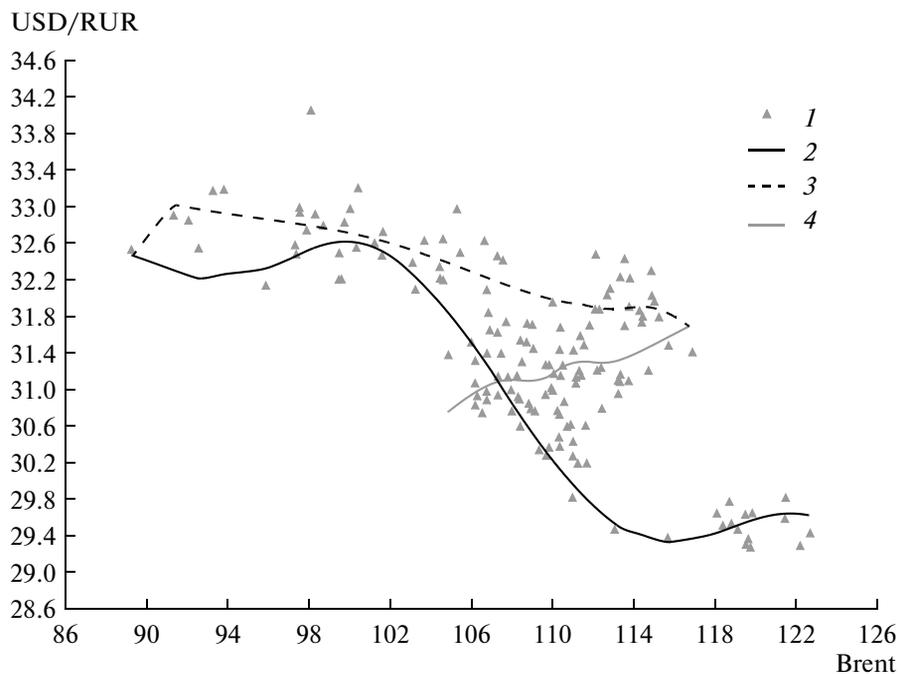


Рис. 11. Гирлянда Гаусса: 1 – диаграмма рассеяния Brent–USD/RUR; 2, 3, 4 – полигауссовская модель нелинейной регрессии на трех временных интервалах соответственно.

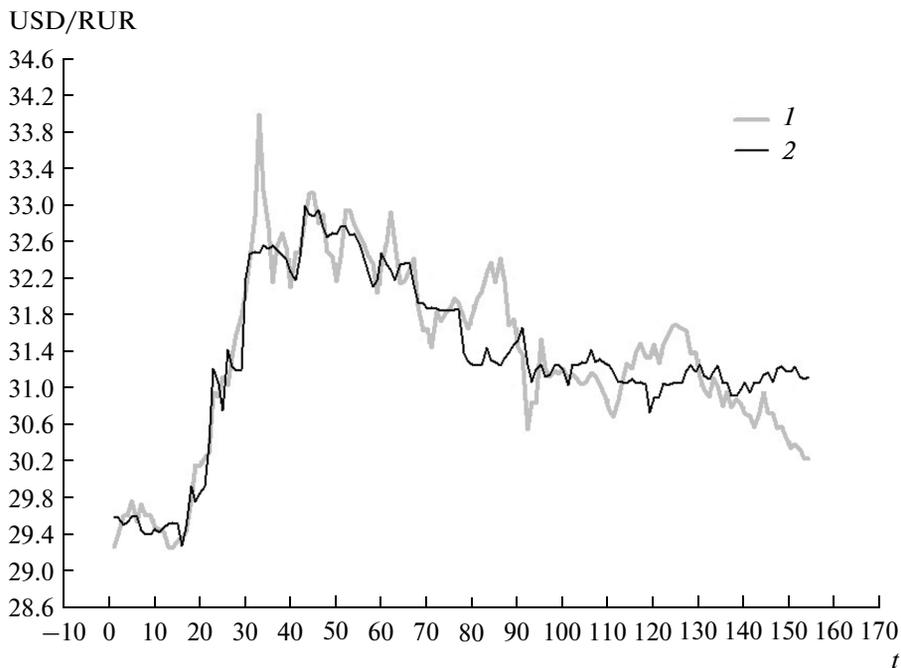


Рис. 12. Полигауссовская ковариационная модель нелинейной регрессии: 1 – курс USD/RUR Банка России в период с 04.04.12 по 16.01.13; 2 – модель нелинейной регрессии.

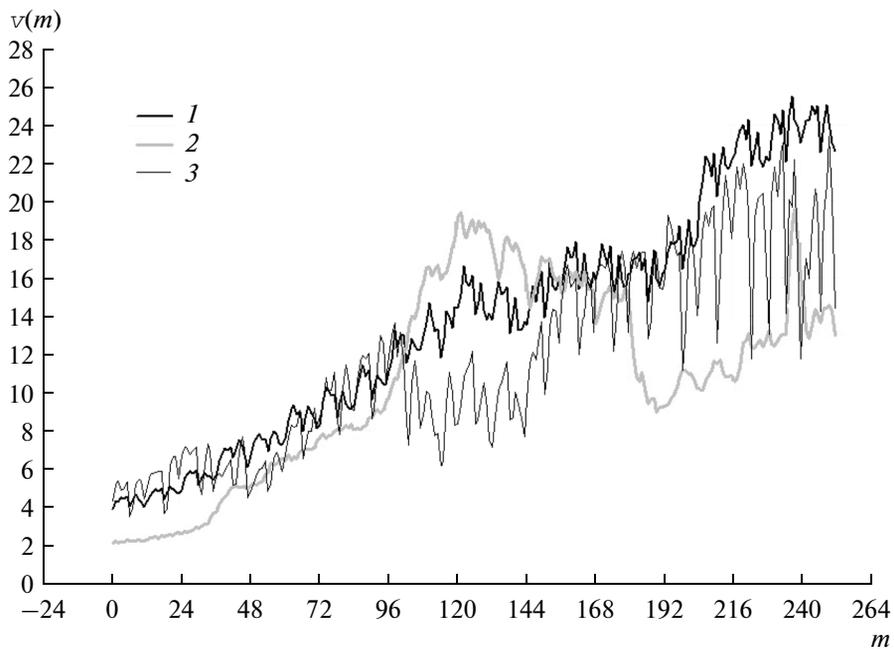


Рис. 13. ВР $v(m)$ ежемесячных продаж нефтепродуктов и сопутствующих товаров: 1 – продукция химического производства; 2 – нефтепродукты; 3 – транспортные средства.

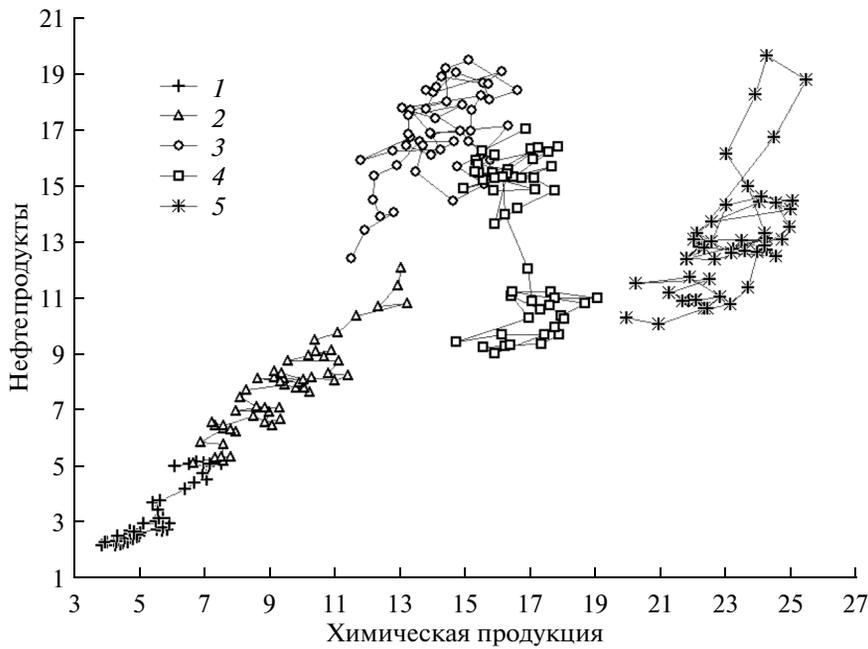


Рис. 14. Кластерная структура динамики факторных ВР.

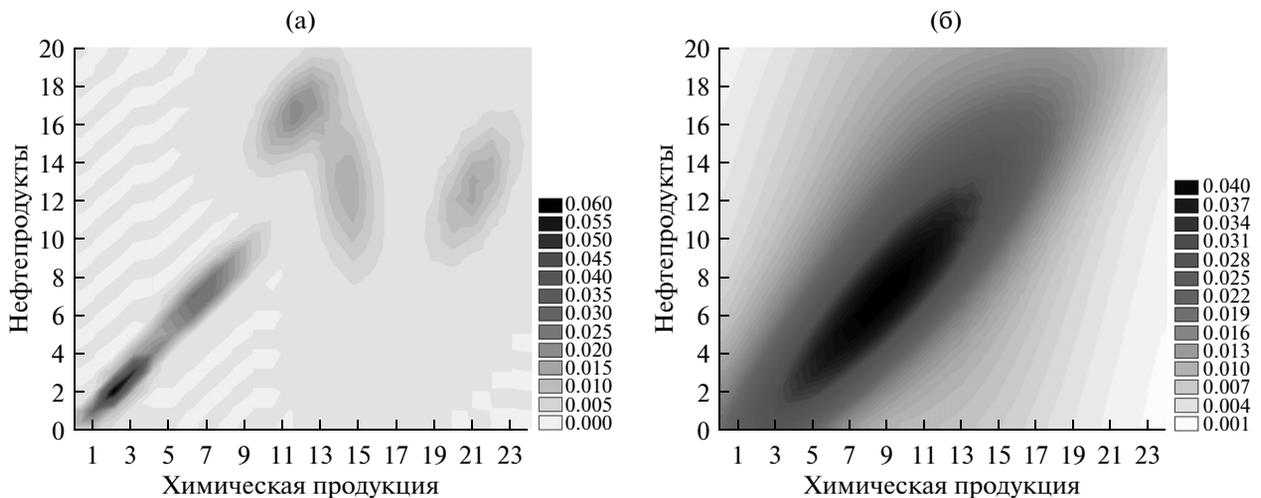


Рис. 15. Полигауссовская модель ПРВ факторных ВР: (а) исходные данные; (б) квазиоптимальная регрессия.

Последующая настройка математического ожидания (МО) a_3 целевого ВР, а также МО $a_{k,1}$, $a_{k,2}$ и СКО $\sigma_{k,1}$, $\sigma_{k,2}$, $k = 1, 2, \dots, 5$, для классов факторных ВР на порядок уменьшает СКО ошибки $e(m)$ регрессионной модели (табл. 5). Соответствующую этим параметрам полигауссовскую аппроксимацию (б) ПРВ факторов иллюстрируют рис. 15б. На рис. 16а–16д представлены зависимости апостериорных вероятностей $\text{Pr}(k|x_1, x_2)$ классов $k = 1, 2, \dots, 5$ от значений факторов x_1 и x_2

для исходных данных (слева) и квазиоптимальной регрессионной модели (справа).

На рис. 17а представлено квазиоптимальное регрессионное приближение (17) в случае пяти классов (кривая 1) для ВР $v(m)$ ежемесячных продаж транспортных средств (кривая 2), а также ВР $e(m)$ ошибок и методическая погрешность регрессии, полученная рекуррентным сглаживанием [30] с помощью модели Мак'Лейна наименьших квадратов, взвешенных расстоянием (Distance

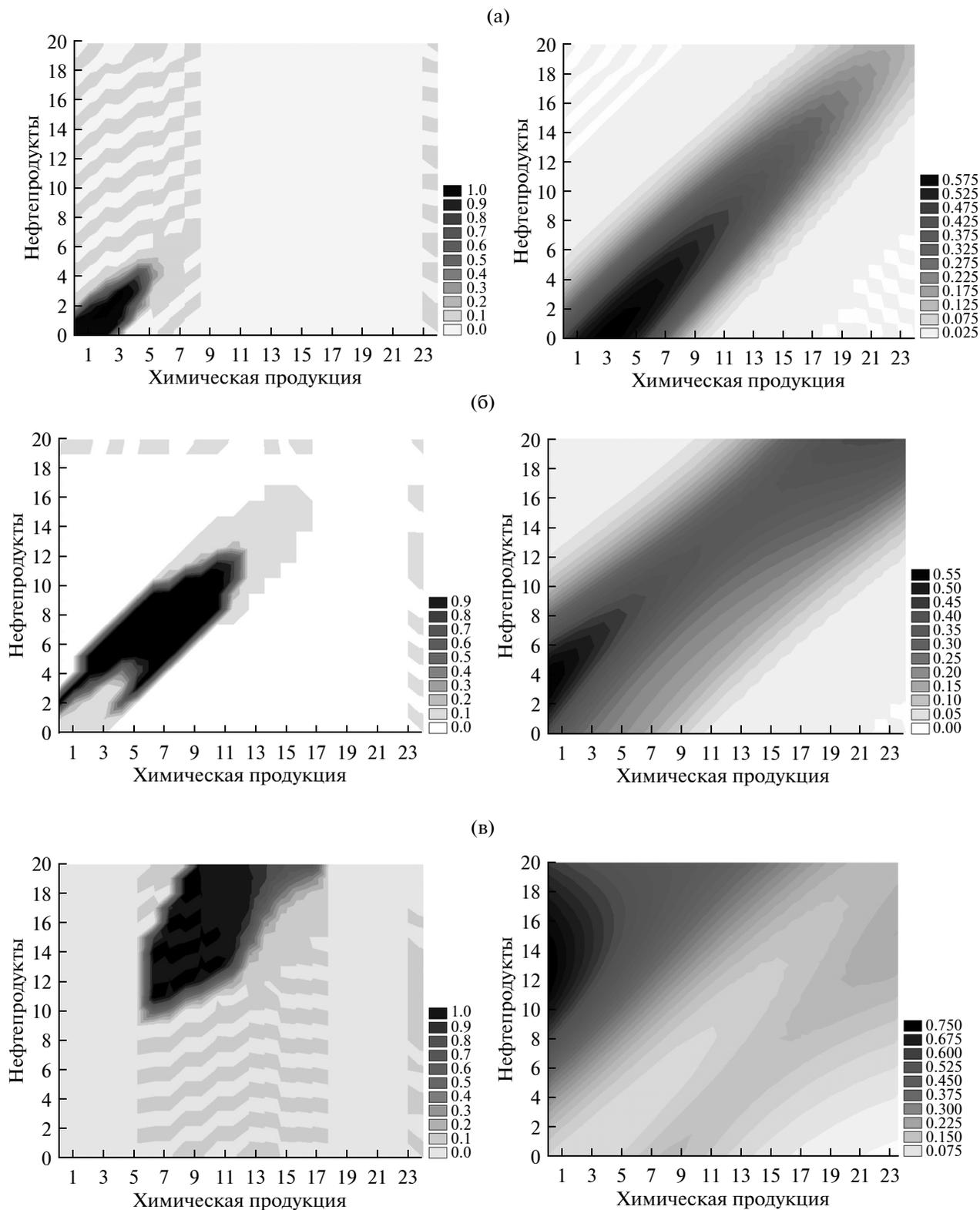


Рис. 16. Апостериорные вероятности для исходных данных (слева) и квазиоптимальной регрессии (справа): а) $\Pr(1|x_1, x_2)$, б) $\Pr(2|x_1, x_2)$, в) $\Pr(3|x_1, x_2)$, г) $\Pr(4|x_1, x_2)$, д) $\Pr(5|x_1, x_2)$.

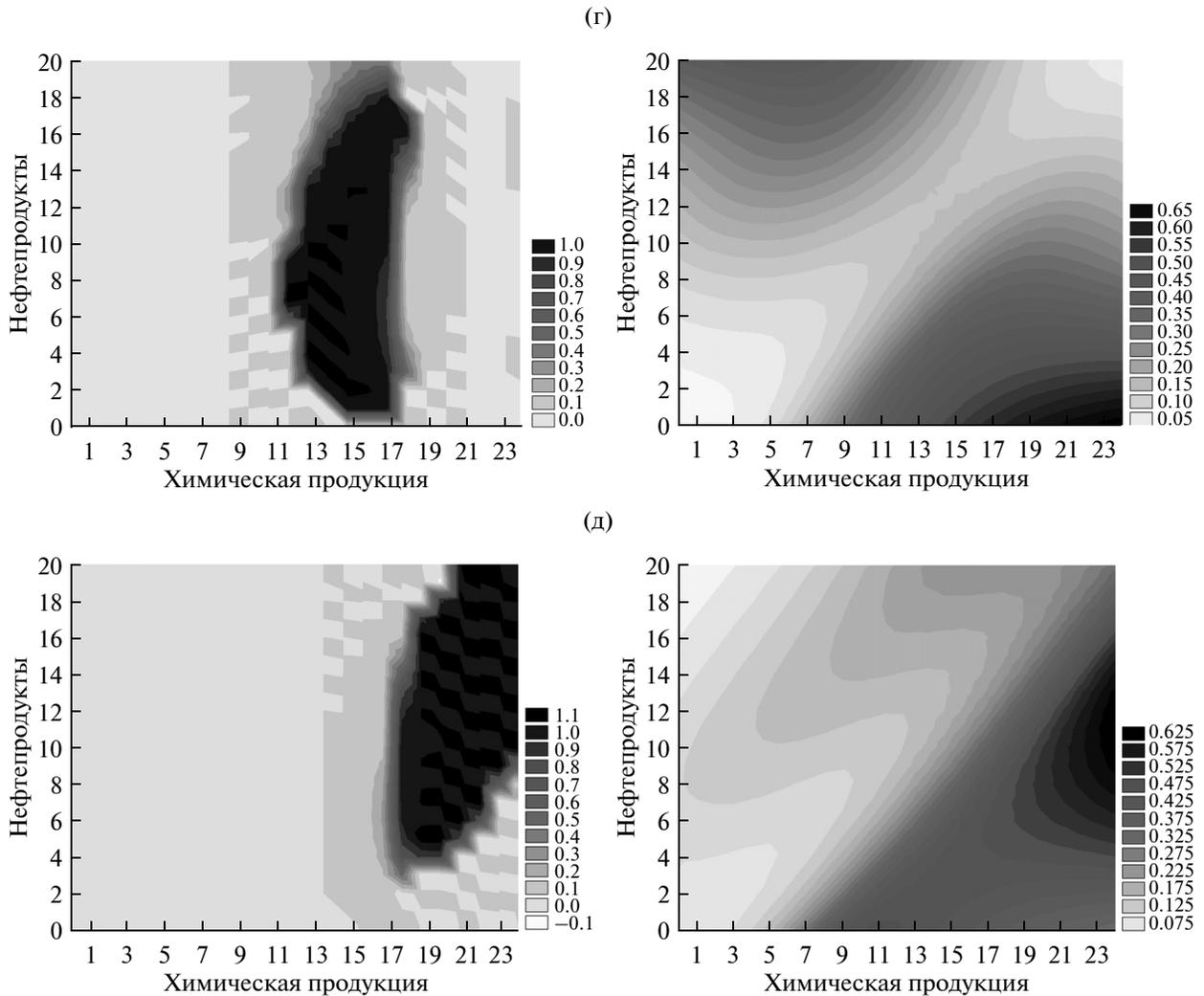


Рис. 16. Окончание.

Weighted Least Squares, DWLS) с весовой функцией, предложенной в [31].

Методическую погрешность удается уменьшить за счет увеличения количества классов регрессионного приближения (17). В частности,

на втором этапе моделирования были выбраны семь классов данных, соответствующих семи временным интервалам: 01.01.71–31.01.74; 01.02.74–30.11.78; 01.12.78–30.06.79; 01.07.79–31.08.83; 01.09.83–31.01.86; 01.02.86–31.12.87;

Таблица 4. Параметры полигауссовской модели совместной ПРВ продукции химического производства и нефтепродуктов

k	1	2	3	4	5
ρ_k	0.1905	0.2143	0.1984	0.2063	0.1905
$a_{k,1}$	3.1153	7.7041	16.8021	13.1280	13.0088
$a_{k,2}$	5.3692	9.41980	14.1693	16.7129	23.0941
$\sigma_{k,1}$	1.0211	1.5837	1.5959	2.7113	2.1229
$\sigma_{k,2}$	0.9989	1.6075	1.2430	0.9731	1.2911
$\rho_{k,1,2}$	0.9217	0.9168	0.5745	-0.2630	0.6244
$\sigma^{(k)}$	20.7316	18.4281	19.9023	19.1368	20.7316

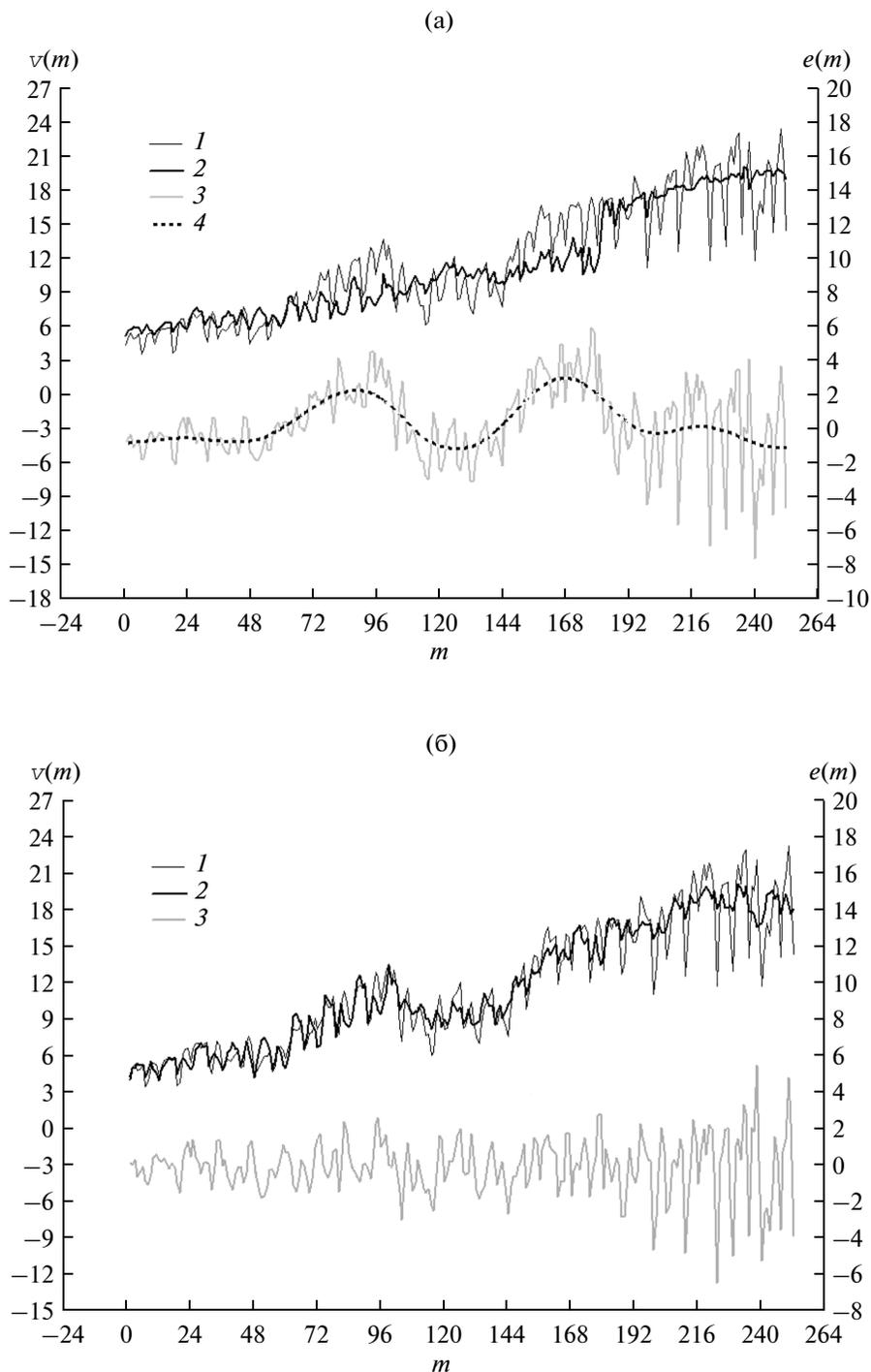


Рис. 17. Полигауссовская ковариационная модель нелинейной регрессии для количества классов $K = 5$ (а) и 7 (б): 1 – ВР $v(m)$ ежемесячных продаж транспортных средств; 2 – квазиоптимальная регрессия; 3 – ВР $e(m)$ ошибки регрессии; 4 – методическая ошибка регрессии.

01.01.88–31.12.91 с количество отсчетов ВР 37, 58, 7, 50, 29, 23, 48 в каждом. Настройка параметров двумерной полигауссовской аппроксимации (б) ПРВ факторов более чем в 20 раз уменьшает СКО ошибки $e(m)$ регрессионной

модели (табл. б) по сравнению с начальным приближением, полученным с помощью ЕМ-алгоритма. Соответствующие этим параметрам результаты моделирования ВР $v(m)$ представлены на рис. 17б.

Таблица 5. Квазиоптимальные параметры полигауссовской модели совместной ПРВ факторных ВР для $K = 5$ и $MO a_3 = 11.1565$

Параметр	k				
	1	2	3	4	5
$a_{k,1}$	5.7633	8.8597	14.2818	11.1588	8.4557
$a_{k,2}$	9.9330	10.8327	12.0439	14.2059	15.0111
$\sigma_{k,1}$	6.1267	8.7102	9.5754	14.9122	11.6757
$\sigma_{k,2}$	5.9935	8.8410	7.4578	5.3520	7.1009
$\sigma^{(k)}$	1.6552	1.4713	1.5890	1.5278	1.6552

Таблица 6. Квазиоптимальные параметры полигауссовской модели совместной ПРВ факторных ВР для $K = 7$ и $MO a_3 = 10.5693$

Параметр	k						
	1	2	3	4	5	6	7
p_k	0.0868	0.0900	0.0278	0.4399	0.1151	0.0913	0.1505
$a_{k,1}$	3.6485	12.9298	10.6327	16.8021	10.0232	6.6546	8.4557
$a_{k,2}$	6.8875	16.3619	12.1274	14.1693	10.6582	11.1220	15.0111
$\sigma_{k,1}$	2.4546	9.4436	3.0634	7.9795	3.3802	3.6667	9.9774
$\sigma_{k,2}$	3.5605	9.4429	3.6747	6.2149	3.71693	4.5668	6.0681
$\rho_{k,1,2}$	0.7853	0.8969	0.8841	0.6345	0.1692	0.4437	0.6244
$\sigma^{(k)}$	2.1050	1.3429	11.1267	1.5577	2.6857	3.3864	1.6226

ЗАКЛЮЧЕНИЕ

В работе получено байесовское приближение ковариационной модели нелинейной регрессии. Структуру модели реализует ассоциативная машина в виде сети смещения мнений экспертов. Каждый эксперт формирует главную поверхность, наилучшую в локальной области пространства факторов. Для текущей точки этого пространства сеть шлюза оценивает достоверности главных поверхностей в терминах их апостериорных вероятностей. Выходной линейный нейрон машины агрегирует мнения экспертов, формируя средневзвешенную апостериорную оценку регрессионной поверхности.

Описание распределения факторов конечной смесью эллипсоидально симметричных плотностей и применение EM-алгоритма идентификации параметров парциальных распределений обеспечивает возможность адаптации регрессионной модели к изменяющимся характеристикам факторов. В рамках полигауссовской аппроксимации ПРВ факторов главные поверхности естественным образом преобразуются в локально взвешенные гиперплоскости.

Результаты вычислительных экспериментов показывают, что для минимизации СКО ошибки аппроксимации данных регрессионной моделью параметры главных поверхностей, полученные с помощью EM-алгоритма, необходимо корректировать в сторону увеличения эффекта сглаживания факторов.

СПИСОК ЛИТЕРАТУРЫ

1. *Scott D.W., Stephan R.S.* Multi-dimensional Density Estimation // Handbook of Statistics. 2005. V. 24. P. 229.
2. *Расин Дж.* // Квантиль. 2008. № 4. С. 7.
3. *Надарая Э.А.* // Теория вероятностей и ее применения. 1964. Т. 19. Вып. 1. С. 147.
4. *Watson G.S.* // Sankhya. Indian J. Statist. 1964. V. A26. P. 359.
5. *Надарая Э.А.* Непараметрическое оценивание плотности вероятностей и кривой регрессии. Тбилиси: Изд-во Тбил. ун-та, 1983.
6. *Алексеев В.Г.* // Автоматика и телемеханика. 1988. № 7. С. 81.
7. *Cleveland W.S.* // J. Amer. Statistical Association. 1979. V. 74. № 368. P. 829.

8. *Scott D.W.* Multivariate Density Estimation: Theory, Practice and Visualization. N.Y.: John Wiley & Sons Inc, 1992.
9. *Heiler S.* // A Course in Time Series Analysis / Eds. Pena D., Tiao G.C., Tsay R.S. N.Y.: John Wiley & Sons Inc, 2001.
10. *Воронцов К.В.* Лекции по алгоритмам восстановления регрессии. 2009 (<http://www.MachineLearning.ru/wiki/images/a/aa/Voron-ML-Regression.pdf>).
11. *Анатольев С.А.* // Квантиль. 2009. № 7. С. 37.
12. *Ту Дж., Гонсалес Р.* Принципы распознавания образов. М.: Мир, 1978.
13. *Хайкин С.* Нейронные сети: полный курс. М.: ИД "Вильямс", 2006.
14. *Bhuiyan S.M.A., Atton-Okine N.O., Barner K.E. et al.* // Adv. Adaptive Data Analysis. 2009. V. 1. № 2. P. 309.
15. *Деврой Л., Дьерфи Л.* Непараметрическое оценивание плотности. L1- подход. М.: Мир, 1988.
16. *Breiman L., Meisel W., Purcell E.* // Technometrics. 1977. V. 19 № 2. P. 135.
17. *Marchette D.J.* // Wiley Interdisciplinary Reviews: Computational Statistics. 2009. V. 1. № 1. P. 106.
18. *Priebe C.E., Marchette D.J.* // Pattern Recognition. 1993. V. 26. № 5. P. 771.
19. *Лабунец Л.В.* // РЭ. 2000. Т. 45. № 10. С. 1214.
20. *Лабунец Л.В., Лебедева Н.Л., Чижов М.Ю.* // РЭ. 2012. Т. 57. № 6. С. 638.
21. *Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: Классификация и снижение размерности: Справочное издание. М.: Финансы и статистика, 1989.
22. *Де Гроот М.* Оптимальные статистические решения. М.: Мир, 1974.
23. *Фукунага К.* Введение в статистическую теорию распознавания образов. М.: Наука, 1979.
24. *Tipping M.E., Bishop C.M.* // Neural Computation. 1999. V. 11. P. 443.
25. *Principal Manifolds for Data Visualization and Dimension Reduction / Eds. Gorban A.N., Kegl B., Wunsch D.C., Zinovyev A.Y.* // Lecture Notes in Computational Science and Engineering. Berlin; Heidelberg: Springer Verlag, 2008. V. 58.
26. *Бокс Дж., Дженкинс Г.* Анализ временных рядов. Прогноз и управление: М.: Мир, 1974. Вып. 2.
27. *Лабунец Л.В., Лукин Д.С., Червяков А.А.* // РЭ. 2012. Т. 57. № 12. С. 1289.
28. *Лабунец Л.В., Лабунец Е.Л., Лебедева Н.Л.* // Аудит и финансовый анализ. 2014. № 3. С. 450.
29. *Hyndman R.J.* Time Series Data Library (URL: <http://datamarket.com/data/list/?q=provider:tsdl>).
30. *Лабунец Л.В., Лабунец Н.Л., Чижов М.Ю.* // РЭ. 2011. Т. 56. № 12. С. 1468.
31. *Лабунец Л.В., Лебедева Н.Л., Чижов М.Ю.* // Вестник Рос. нового ун-та. 2013. № 4. С. 66.