РОССИЙСКАЯ АКАДЕМИЯ НАУК

СИБИРСКОЕ ОТДЕЛЕНИЕ

АВТОМЕТРИЯ

№ 5 2001

УДК 517.8 : 519.72

В. В. Савченко

(Нижний Новгород)

ТЕОРЕТИКО-ИНФОРМАЦИОННОЕ ОБОСНОВАНИЕ ЛИНЕЙНЫХ ОЦЕНОК ПРОГНОЗИРОВАНИЯ

Обсуждается проблема оптимальности линейных оценок прогнозирования в условиях априорной неопределенности. Исследуются возможности адаптивного подхода и нового класса методов, основанных на авторегрессионной модели стационарного гауссовского процесса. Предлагается информационный критерий для их тестирования по эффективности на ограниченных интервалах наблюдений.

Введение. Задача прогнозирования случайных временных рядов относится к числу центральных задач статистического анализа во многих областях его практического применения, особенно в экономике и управлении. Поэтому не вызывает удивления неугасающий долгие годы интерес к ней со стороны не только исследователей – представителей разных научных школ и направлений, но и широкого круга специалистов-практиков. В связи с бурным развитием и распространением цифровой вычислительной техники, а также современных информационных сетей исследования в этом актуальном направлении приняли в последнее время наиболее интенсивный характер. Как результат, все чаще на первый план выступает проблема выбора наилучшего метода прогнозирования из огромного числа предлагаемых разными авторами. Ее решению в расчете на методы линейного вида, по-видимому, наиболее перспективные в практическом отношении, и посвящена данная статья. Исследование проведено на основе универсального теоретико-информационного подхода в развитие результатов и выводов предыдущей работы автора [1]. При этом главное внимание уделяется новому классу адаптивных методов [2], основанных на авторегрессионной (АР) модели наблюдений и обладающих высокими динамическими свойствами.

Линейные оценки прогнозирования. Рассмотрим стационарный гауссовский процесс $X(t)$ с нулевым значением математического ожидания $\mathbf{M}[X(t)]=0$ и $(n\times n)$-матрицей автокорреляций $\mathbf{K}_{n,n}=\|k(t,l)\|$, где $k(t,l)=\mathbf{M}[X(t)X(l)]=k_x(t-l)$, $t,l=\overline{1,n}$. И пусть этот процесс наблюдается в дискретном времени $t=0,1,\ldots,n-1$ по своей реализации (выборке) $\{x(t)\}$ на ограниченном интервале $t<n$ длиной $n<\infty$. Тогда его прогноз на один шаг или временной дискрет в будущее $\hat{x}(n)=y(n)$ есть функция в общем случае $m\geq 1$ последовательных наблюдений, предшествующих моменту $t=n$: $y_m(n)=y[x(n-1),\ldots,x(n-m)]$. Здесь m обозначает порядок оценки прогнозирования, причем $m < n$. В частном случае линейной функции $y(\ldots)$ получаем линейную оценку прогнозирования общего вида:

$$y_m(n) = a_m(1)x(n-1) + a_m(2)x(n-2) + \ldots + a_m(m)x(n-m) = \sum_{i=1}^{m} a_m(i)x(n-i), \tag{1}$$

где $\mathbf{a}_m = \mathrm{col}[a_m(i)]$ – m-вектор (столбец) постоянных коэффициентов.

Из этого же выражения по индукции [1] будем иметь линейные оценки прогнозирования $y_m(n+1)$, $y_m(n+2)$, ..., $y_m(n+k-1)$ на произвольное число шагов $k > 1$ в будущее.

Известно [2], что при заданном порядке m именно линейная оценка (1) применительно к произвольному гауссовскому процессу характеризуется наивысшей точностью в смысле минимума среднего (статистического) квадрата

$$\sigma_z^2(m) = \mathbf{M}[x(n) - y_m(n)]^2 \to \min \tag{2}$$

величины ошибки прогнозирования

$$z(n) = x(n) - y_m(n) \tag{3}$$

при условии, что вектор коэффициентов $\mathbf{a}_m$ отвечает системе m нормальных уравнений

$$\mathbf{K}_{m,m}\mathbf{a}_m = \mathbf{k}_m.$$

Здесь $\mathbf{K}_{m,m}$ – автокорреляционная матрица анализируемого процесса размера $m \times m$; $\mathbf{k}_m = \mathrm{col}_m[k_x(i)]$ – m-вектор (столбец) коэффициентов его (процесса) автокорреляции $k_x(i) = \mathbf{M}[X(t)X(t+i)]$, $i = \overline{1, n}$. Решая систему нормальных уравнений, в матричном виде будем иметь

$$\mathbf{a}_m^* = \mathbf{K}_{m,m}^{-1}\mathbf{k}_m, \tag{4}$$

где $\mathbf{K}_{m,m}^{-1}$ – обратная $(m \times m)$-матрица автокорреляций, символ $*$ – оптимальность результирующего вектора коэффициентов.

Таким образом, выражения (1) и (4) в совокупности определяют наилучшую оценку прогнозирования из множества всех мыслимых оценок m-го порядка, включая даже нелинейные. К сожалению, практическое значение этого классического результата обычно оказывается существенно ослабленным ввиду актуальной для большинства прикладных задач проблемы априорной неопределенности. В указанных условиях и автокорреляционная матрица $\mathbf{K}_{m,m}$, и вектор коэффициентов автокорреляции $\mathbf{k}_m$, а вслед за ними и вся правая часть выражения (4) заранее строго не определены и поэтому не могут быть использованы при оптимальном решении задачи прогнозирования. Эффективным средством преодоления данной проблемы может служить адаптивный (байесовский) подход, нацеленный на получение оценок прогнозирования со свойством их оптимальности в асимптотике, т. е. при бесконечно возрастающем объеме наблюдений.

Адаптивные оценки прогнозирования. Адаптивный подход в любом своем возможном варианте сводит искомое решение задачи прогнозирования к той же системе выражений (1), (4), но при замене в последнем неизвестных автокорреляционных матрицы $\mathbf{K}_{m,m}$ и вектора $\mathbf{k}_m$ их максимально правдоподобными выборочными оценками $\mathbf{K}_{m,m}(n)$ и $\mathbf{k}_m(n)$ соответственно (n – объем выборки). При этом в зависимости от конкретного вида корреляционных оценок будем иметь разные варианты адаптивной оценки прогнозирования (1). Очевидно, что предпочтение следует отдавать тем ее вариантам, которые используют сильно состоятельные корреляционные оценки. В этом случае результирующая адаптивная оценка прогнозирования будет обладать свойством асимптотической оптимальности, т. е. сходимости при $n \to \infty$ к оптимальному виду (1), (4) «почти наверное» или с вероятностью 1. Задача свелась, таким образом, к выбору наиболее эффективных методов корреляционного анализа. Причем определяющее значение здесь будет иметь их скорость сходимости к оптимальному решению.

Из вышеизложенного следует, что адаптивные методы нового класса, изначально нацеленные на решение проблемы малых выборок в задачах спектрального анализа, представляют первостепенный интерес. Типичным представителем этого класса является метод Берга. Его математическая формулировка может быть записана в рекуррентном виде [2]:

$$\left.\begin{aligned}
&a_q(i) = a_{q-1}(i) + \zeta_q a_{q-1}(q-i), \qquad i = \overline{1,q},\\
&\zeta_q = \frac{1}{n-q}\sum_{t=q}^{n-1} \mathrm{v}_{q-1}(t) u_{q-1}(t-1) / S_{q-1}^2,\\
&S_{q-1}^2 = \frac{1}{2(n-q)}\sum_{t=q}^{n-1} [\mathrm{v}_{q-1}^2(t) + u_{q-1}^2(t-1)],\\
&\mathrm{v}_q(t) = \mathrm{v}_{q-1}(t) - \zeta_q u_{q-1}(t-1),\\
&u_q(t) = u_{q-1}(t-1) - \zeta_q \mathrm{v}_{q-1}(t), \qquad q = \overline{1,m},
\end{aligned}\right\} \tag{5}$$

с инициализацией системой равенств $\mathrm{v}_0(t) = u_0(t) = x(t)$ для всех моментов времени $t = 0, 1, \ldots, n-1$. Ее центральным звеном служит рекурсия Левинсона, которая связывает между собой векторы коэффициентов линейной оценки прогнозирования $\{\mathbf{a}_q\}$ возрастающих порядков $q = 1, 2, \ldots, m$. Финальное (при $q = m$) значение этой рекурсии $\{a_m(i),\ i = \overline{1,m}\}$ совместно с выражением (1) определяет адаптивную линейную оценку прогнозирования m-го порядка со свойством оптимальности (2) в асимптотике, когда объем выборки $n \to \infty$. При этом достигаемая скорость сходимости не может быть улучшена по порядку и отвечает степенному закону $1/n$, $n = 1, 2, \ldots$, [3]. Причем указанное качество присуще большинству новых методов, объединенных общей идеей применения авторегрессионной модели наблюдений.

Авторегрессионная модель. В наиболее общем виде АР-модель m-го порядка ($m \geq 1$) определяется следующим разностным уравнением:

$$x(n) = \sum_{i=1}^{m} a_m(i) x(n-i) + \eta(n) = \mathbf{a}_m^{\mathrm{T}} \mathbf{x}_m(n-1) + \eta(n), \qquad n = 1, 2, \ldots . \tag{6}$$

Здесь $\mathbf{a}_m^{\mathrm{т}} = \mathrm{str}_m[a_m(i)]$ – m-вектор (строка) коэффициентов линейной авторегрессии (т – символ транспонирования векторов); $\mathbf{x}_m(n-1) = \mathrm{col}_m[x(n-i)]$ – m-вектор (столбец), составленный из m последовательных отсчетов процесса авторегрессии в дискретном времени n; $\eta(n)$ – белый гауссовский шум в роли порождающего процесса с нулевым математическим ожиданием $\mathbf{M}[\eta(n)] = 0$ и постоянной дисперсией $\mathbf{D}[\eta(n)] = \mathbf{M}[\eta^2(n)] = \sigma_\eta^2$. Его автокорреляция имеет вид дискретной δ-функции Кронекера: $\mathbf{M}[\eta(t)\eta(l)] = \delta(t,l)\sigma_\eta^2$, где $\delta(t,l) = 0$ при $t \neq l$ и $\delta(t,l) = 1$ при $t = l$. В зависимости от выбранных параметров $\mathbf{a}_m$ и σ_η^2 рассматриваемая АР-модель (6) оказывается более или менее адекватной анализируемому случайному процессу $X(t)$. Так, например, в методе Берга вектор АР-коэффициентов определяется финальным значением рекурсии (5) при равенстве $q = m$, а согласованная с ним дисперсия порождающего шума

$$\sigma_\eta^2 = \left[\prod_{q=1}^{m}(1-\zeta_q^2)\right]\sigma_x^2 = \sigma_m^2(n) \tag{7}$$

рассчитывается по набору коэффициентов отражения $\{\zeta_q\}$ [2] при учете заданной дисперсии анализируемого процесса $\sigma_x^2 = \mathbf{D}[X(t)]$. Если последняя априори неизвестна, то в выражение (7) войдет ее состоятельная выборочная оценка S_{q-1}^2 из той же рекурсии (5) при $q = 1$.

Отметим, что равенство (7) выходит далеко за пределы метода Берга и имеет характер нормирующего условия общего вида в том смысле, что при его выполнении средняя мощность моделирующего АР-процесса (6) гарантированно приравнивается к дисперсии σ_x^2 вне зависимости от вида векторов коэффициентов $\mathbf{a}_m$. При этом требуемый в (7) набор коэффициентов отражения $\{\zeta_q\}$ легко пересчитывается из вектора $\mathbf{a}_q$ с помощью рекурсии Левинсона убывающего порядка $q = m, m-1, \ldots$.

В терминах статистической радиотехники [4] уравнение (6) описывает динамику рекурсивного формирующего фильтра m-го порядка с белым шумом $\{\eta(n)\}$ на входе. Перепишем его в инверсном виде

$$\eta(n) = x(n) - \mathbf{a}_m^{\mathrm{т}}\mathbf{x}_m(n-1), \quad n = 1, 2, \ldots. \tag{8}$$

Мы получили уравнение обеляющего фильтра, или декоррелятора, АР-процесса m-го порядка с белым шумом на выходе. Сравним его правую часть с выражением (3) для ошибки прогнозирования $z(n)$ и с учетом определения (1) легко придем к выводу, что с точки зрения решаемой задачи обеляющий фильтр (8) представляет собой фильтр ошибки прогнозирования в его адаптивном, самонастраивающемся (по выборке) варианте. В таком случае оптимизация линейной оценки прогнозирования по критерию (2) по сути означает выбор АР-модели наблюдений (6) с минимальной дисперсией отклика обеляющего фильтра (8) на анализируемый случайный процесс. При этом нельзя забывать, что рассматриваемый процесс $X(t)$ и его АР-модель (6) – это далеко не одно и то же. Сказанное явно вытекает из сравнения их $(n \times n)$-корреляционных матриц.

Перепишем инверсное преобразование (8) для последовательных моментов времени $n, n-1, \ldots, 1$, взятых в порядке их убывания, в матричном виде

$\boldsymbol{\eta}_n(n)=\mathbf{A}\mathbf{x}_n(n)$ или $\mathbf{x}_n(n)=\mathbf{A}^{-1}\boldsymbol{\eta}_n(n)$ и после этого по определению автокорреляционной матрицы АР-процесса (6) при $n \to \infty$ будем иметь

$$\mathbf{R}_{n,n}=\mathbf{M}[\mathbf{x}_n(n)\mathbf{x}_n^{\mathrm{T}}(n)]=\mathbf{M}[\mathbf{A}^{-1}\boldsymbol{\eta}_n(n)\boldsymbol{\eta}_n^{\mathrm{T}}(n)\mathbf{A}^{-\mathrm{T}}]=$$

$$=\mathbf{A}^{-1}\mathbf{M}[\boldsymbol{\eta}_n(n)\boldsymbol{\eta}_n^{\mathrm{T}}(n)]\mathbf{A}^{-\mathrm{T}}=\sigma_\eta^2(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}. \tag{9}$$

Здесь $\boldsymbol{\eta}_n(n)=\mathrm{col}_n[\eta(n+1-i)]$ – n-вектор (столбец), составленный из последовательных отсчетов порождающего шума в моменты $n, n-1, \ldots, 1$; $\mathbf{A}$ – $(n \times n)$-матрица коэффициентов инверсного преобразования верхней треугольной структуры:

$$\mathbf{A}=\begin{bmatrix} 1 & -a_m(1) & -a_m(2) & \ldots & -a_m(m) & 0 & 0 & \ldots \\ 0 & 1 & -a_m(1) & \ldots & -a_m(m-1) & -a_m(m) & 0 & \ldots \\ 0 & 0 & 1 & \ldots & -a_m(m-2) & -a_m(m-1) & -a_m(m) & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix}.$$

Хорошо видно, что в общем случае не обязательно выполнение равенства $\mathbf{R}_{n,n}=\mathbf{K}_{n,n}$, т. е. анализируемый процесс $X(t)$ и его АР-модель m-го порядка, строго говоря, имеют разные корреляционные свойства, хотя, в силу равенства (7), диагональные элементы их корреляционных матриц и совпадают. Поэтому разными в общем случае будут и дисперсии откликов фильтра ошибки прогнозирования (8) на эти случайные процессы. Равенство между ними достигается только в идеальном варианте, когда вектор АР-коэффициентов $\mathbf{a}_m$ отвечает уравнению (4). При данном условии из выражения (9) получим [3]

$$\sigma_\eta^2=[\boldsymbol{\Gamma}_n^{\mathrm{T}}\mathbf{R}_{n,n}^{-1}\boldsymbol{\Gamma}_n]^{-1}=[\boldsymbol{\Gamma}_{m+1}^{\mathrm{T}}\mathbf{R}_{m+1,m+1}^{-1}\boldsymbol{\Gamma}_{m+1}]^{-1}\Big|_{\mathbf{a}_m=\mathbf{a}_m^*}=$$

$$=[\boldsymbol{\Gamma}_{m+1}^{\mathrm{T}}\mathbf{K}_{m+1,m+1}^{-1}\boldsymbol{\Gamma}_{m+1}]^{-1}=\sigma_x^2-\mathbf{k}_m^{\mathrm{T}}\mathbf{K}_{m,m}^{-1}\mathbf{k}_m=\sigma_{z,\min}^2,$$

где $\Gamma_{n(m+1)}=\mathrm{col}_{n(m+1)}[1,0,\ldots,0]$ – вектор-столбец размера n (или $m+1$), составленный из единицы на первой позиции и нулей на всех остальных позициях; $\sigma_{z,\min}^2=\min\sigma_z^2(m)$ – предельно достижимое значение дисперсии ошибки прогнозирования (3) как минимальное собственное число автокорреляционной матрицы $\mathbf{K}_{m+1,m+1}$ конечного порядка $m+1$ [4].

Напротив, в любом другом варианте АР-процесса (6) с нормированной средней мощностью на уровне $\sigma_x^2=\mathrm{const}$ дисперсия его порождающего шума $\sigma_\eta^2<\sigma_z^2(m)$ может существенно отличаться от достигаемой дисперсии ошибки прогнозирования. Чем выше точность осуществляемого прогнозирования и, следовательно, выше качество используемой АР-модели наблюдений, тем меньше отношение дисперсий $\sigma_z^2(m)/\sigma_\eta^2 \geq 1$ отличается от единицы. И наоборот, анализируя это отношение по его величине, можно для каждой конкретной оценки прогнозирования сделать обоснованные выводы о ее надежности. Подробно эта идея развивается далее на основе строгого теоретико-информационного подхода.

Информационный критерий. Сведя решаемую задачу прогнозирования (1), (2) к выбору оптимальной АР-модели (6), воспользуемся при анализе качества последней универсальным критерием минимума информационного рассогласования в метрике Кульбака – Лейблера [5]:

$$I_n[f_R \mid f_K] = \int \ldots \int \ln\left[\frac{f_K(\mathbf{x}_n)}{f_R(\mathbf{x}_n)}\right] f_K(\mathbf{x}_n) d\mathbf{x}_n \to \min_R .$$

Здесь $f_K(\mathbf{x}_n)$, $f_R(\mathbf{x}_n)$ – плотности n-мерных гауссовских распределений с автокорреляционными матрицами $\mathbf{K}_{n,n}$ и $\mathbf{R}_{n,n}$ соответственно. Путем стандартных вычислений получим

$$I_n[f_R \mid f_K] = \frac{1}{2} \ln \frac{|\mathbf{R}_{n,n}|}{|\mathbf{K}_{n,n}|} - \frac{n}{2} + \frac{1}{2} \operatorname{tr}[\mathbf{R}_{n,n}^{-1} \mathbf{K}_{n,n}], \qquad (10)$$

где символ $|\bullet|$ – определитель квадратной $(n \times n)$-матрицы, а $\operatorname{tr}(\cdot)$ – ее след. Отметим, что при равенстве автокорреляционных матриц $\mathbf{K}_{n,n} = \mathbf{R}_{n,n}$ будем иметь $I_n[f_R \mid f_K] = 0$. Во всех других случаях выполняется строгое неравенство $I_n[f_R \mid f_K] > 0$, причем значение данного показателя монотонно возрастает при увеличении различий между рассматриваемыми распределениями в теоретико-информационном смысле. На это же важное качество информационного рассогласования $I_n[f_R \mid f_K]$ указывает и второе известное определение его [6] как величины взаимной энтропии упорядоченной пары распределений $f_R(\mathbf{x}_n)$ и $f_K(\mathbf{x}_n)$. При увеличении взаимной энтропии используемая АР-модель становится все менее адекватной анализируемому случайному процессу $X(t)$.

Раскроем правую часть выражения (10) в зависимости от параметров нашей модели (6) – дисперсии порождающего шума σ_η^2 и вектора АР-коэффициентов $\mathbf{a}_m$:

$$I_n[f_R \mid f_K] = \frac{n}{2} \ln \frac{\sigma_\eta^2}{\sigma_0^2} - \frac{n}{2} + \frac{1}{2} \operatorname{tr}[\mathbf{M}(\mathbf{x}_n \mathbf{x}_n^{\mathrm{T}}) \mathbf{R}_{n,n}^{-1}] =$$

$$= \frac{n}{2} \ln(2\pi\sigma_\eta^2) + \frac{1}{2} \mathbf{M}[\mathbf{x}_n^{\mathrm{T}} \mathbf{R}_{n,n}^{-1} \mathbf{x}_n] - H_n^*(X) =$$

$$= \frac{n}{2} \ln(2\pi\sigma_\eta^2) + \frac{1}{2\sigma_\eta^2} \mathbf{M}[\mathbf{x}_n^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \mathbf{A} \mathbf{x}_n] - H_n^*(X) =$$

$$= \frac{n}{2} \ln(2\pi\sigma_\eta^2) + \frac{1}{2\sigma_\eta^2} \mathbf{M}\left[\sum_{i=0}^{n-1} z^2(n-i)\right] - H_n^*(X) =$$

$$= \frac{n}{2}\left[\ln(2\pi\sigma_\eta^2) + \frac{\sigma_z^2(m)}{\sigma_\eta^2}\right] - H_n^*(X),$$

где $\sigma_0^2 = \text{const}$ – минимальное собственное число автокорреляционной матрицы $\mathbf{K}_{n,n}$; $H_n^*(X) = \frac{1}{2}\ln[(2\pi e)^n |\mathbf{K}_{n,n}|] = \frac{n}{2}[\ln(2\pi\sigma_0^2)] + \frac{n}{2}$ – энтропия неизвестного истинного распределения. Переходя от этого выражения к удельной (на один отсчет данных) величине информационного рассогласования, получим

$$I_n[f_R \mid f_K] = \frac{1}{n} I_n[f_R \mid f_K]_{n\to\infty} = \frac{1}{2}\left[\ln(2\pi\sigma_\eta^2) + \frac{\sigma_z^2(m)}{\sigma_\eta^2}\right] - h^*(X),$$

где $h^*(X) = \frac{1}{2}\ln(2\pi\sigma_0^2) + \frac{1}{2} = \text{const}$ – удельная энтропия случайного процесса $X(t)$. Отсюда мгновенно следует искомый информационный критерий оптимальности АР-модели наблюдений (6) и вместе с ней линейной оценки прогнозирования (1) вида

$$\Delta h(m,n) = \ln\sigma_\eta^2(m,n) + \frac{\sigma_z^2(m,n)}{\sigma_\eta^2(m,n)} \to \min. \tag{11}$$

Здесь $\sigma_\eta^2(m,n)$, $\sigma_z^2(m,n)$ – зависимости дисперсий порождающего шума $\{\eta(n)\}$, а также ошибки прогнозирования $\{z(n)\}$ от двух исходных параметров адаптивной оценки: порядка m и объема анализируемой (обучающей) выборки n.

Проблема состоит в том, что обе указанные зависимости должны быть раскрыты в явном виде, что не всегда выполнимо даже в теории. Особенно это относится к зависимости дисперсии ошибки прогнозирования $\sigma_z^2(m,n)$, строгое определение которой наталкивается все на ту же проблему априорной неопределенности. Поэтому при практических вычислениях вместо строгого определения информационного критерия (11) мы в большинстве случаев будем вынуждены применять его асимптотически оптимальную формулировку. Последняя по своему виду повторяет выражение (11), но при условии замены в нем неизвестных истинных зависимостей $\sigma_\eta^2(m,n)$ и $\sigma_z^2(m,n)$ их состоятельными выборочными оценками. Причем их вид в каждом случае будет зависеть от особенностей конкретного метода адаптации. Например, при применении метода Берга (5), (7) мы можем воспользоваться для указанных целей двумя установленными в нем зависимостями:

$$\left.\begin{aligned} \sigma_\eta^2(m,n) &= \sigma_m^2(n); \\ \sigma_z^2(m,n) &= S_m^2. \end{aligned}\right\} \tag{12}$$

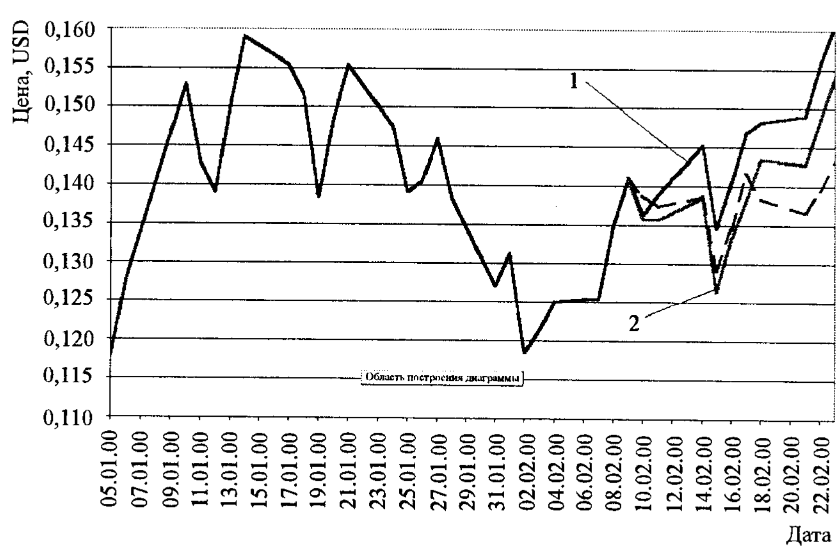
Отметим, что в асимптотике при $n \to \infty$, когда выполняется равенство (9), предложенный критерий (11) принимает упрощенный вид:

$$\Delta h(m,n)\big|_{n\to\infty} = \ln\sigma_z^2(m) + 1 \to \min,$$

эквивалентный классическому критерию минимума среднего квадрата ошибки прогнозирования. Практически это означает, что при больших объемах наблюдений $n \gg 1$ оба рассматриваемых критерия приводят к примерно одинаковым результатам. Совсем иная ситуация возникает в тех случаях, когда объем наблюдений n сопоставим по своей величине с порядком m линейной оценки прогнозирования. Предложенный информационный критерий оптимальности (11) здесь получает наибольшие преимущества по сравнению с критерием. Проиллюстрируем сказанное конкретным примером из практики экономического анализа и прогнозирования.

Прогнозирование динамики фондового рынка страны. Российский фондовый рынок, или рынок корпоративных ценных бумаг: акций, векселей и иных финансовых инструментов, может быть охарактеризован целым рядом своих конъюнктурных показателей. Среди них наиболее информативным многие аналитики полагают котировку акций на закрытие торговой сессии или, проще, цену закрытия. Ее динамика для акций РАО «Единые энергетические системы России» в течение первых пяти недель 2000 г. по состоянию на 9 февраля показана жирной линией на рисунке. Представленные данные получены с открытого сайта РТС [7] и характеризуют собой цены Российской торговой системы в долларах США.

На интервалах роста котировок акций говорят об улучшении конъюнктуры рынка. Напротив, на интервалах спада котировок – об ее ухудшении. Например, резкий спад рыночных цен на 18–19 января 2000 г. отразил реакцию большинства инвесторов, в основном иностранцев, на кризис в Государственной Думе. Понятно, что с точки зрения любого потенциального инвестора задача прогнозирования динамики рыночной конъюнктуры представляет первостепенный интерес. Два разных варианта ее решения в режиме реального времени в отсчете от 9 февраля 2000 г. на две недели, или десять торговых дней, в будущее отображены на том же рисунке двумя сплошными кривыми линиями 1 и 2. При этом в обоих вариантах использовался метод Берга с параметрами $m=160$, $n=220$ (кривая 1) и $m=140$, $n=250$ (кривая 2). Для сравнения штриховой линией показана истинная динамика рыночных цен по состоянию на 23 февраля 2000 г. Видно, что ее характер, даже в мелких деталях, почти безошибочно предопределен во втором варианте прогноза. И этот результат точно согласуется с информационным критерием оптимальности



(11), (12): $\Delta h(140,250) = 2{,}09 < \Delta h(160,220) = 2{,}44$. Отметим, что по критерию минимума среднего квадрата ошибки прогнозирования (12) второй прогноз, напротив, был бы признан менее эффективным, так как $S^2_{140}(250) = 0{,}163 > S^2_{160}(220) = 0{,}109$. И данный факт имеет очевидное объяснение. При высоком порядке ($m = 160$) и относительно небольшом объеме наблюдений ($n = 220$) результирующая АР-модель нацелена главным образом на прогнозирование тренда как доминирующей закономерности в динамике анализируемого временного ряда. Как результат, текущая оценка прогнозирования имеет в этом случае характер медленно меняющегося, т. е. низкочастотного процесса (кривая 1 на рисунке) с относительно небольшой величиной среднего квадрата ошибки $\sigma^2_z(n)$. При уменьшении порядка m до 140 и одновременном увеличении объема n до 250 характер результирующей оценки прогнозирования (кривая 2) меняется в сторону усиления в ней влияния слабых высокочастотных составляющих спектра. Ее вид в таком случае более точно отражает в себе будущие колебания анализируемого процесса, хотя, может быть, и ценой некоторого повышения среднего квадрата ошибки.

Таким образом, благодаря применению предложенного информационного критерия (11), нам удалось заблаговременно, еще 9 февраля 2000 г., предвосхитить более высокое качество второго из рассмотренных вариантов линейной оценки прогнозирования и этим фактом способствовать правильному выбору лучшего варианта.

Заключение. Существует несколько веских причин нарастающего интереса к новым методам авторегрессионного анализа, таким как методы Берга, Юла – Уолкера, ковариационный и др. Среди них на первое место большинство исследователей справедливо ставит высокие динамические свойства данного класса адаптивных методов. Однако при этом важно учитывать, что высокие динамические свойства отнюдь не означают собой автоматически высокую скорость сходимости адаптивной оценки (1) к оптимальному варианту прогноза. Необходимым условием такой сходимости является правильный или, говорят, адекватный выбор параметров авторегрессионной модели, лежащей в основе новых методов. Проблема состоит в том, что в условиях априорной неопределенности и при относительно небольших объемах выборочных данных само понятие адекватность нуждается в специальном обосновании. Довольно общий подход к его определению в терминах информационной метрики Кульбака – Лейблера и составляет главное содержание настоящей статьи. А основной ее результат – это строгое выражение (11) для искомого критерия оптимальности линейных оценок прогнозирования, рассчитанное на ограниченный объем наблюдений. При этом, как и следовало ожидать, в асимптотике информационный критерий (11) плавно переходит в критерий минимума среднего квадрата ошибки [1], поскольку при $n \to \infty$ проблем с адекватностью статистических моделей не возникает в принципе.

Предложенный критерий может быть применен как при оптимизации, так и при тестировании по выборке линейных оценок прогнозирования, основанных на разных методах авторегрессионного анализа.

СПИСОК ЛИТЕРАТУРЫ

1. **Савченко В. В.** Прогнозирование социально-экономических процессов на основе адаптивных методов спектрального оценивания // Автометрия. 1999. № 3. С. 99.

2. **Марпл-мл. С. Л.** Цифровой спектральный анализ и его приложения. М.: Мир, 1990.
3. **Савченко В. В.** Адаптивные методы нелинейного спектрального оценивания на основе принципа минимакса энтропии: Автореф. дис. ... д-ра техн. наук. Н. Новгород: НГТУ, 1993.
4. **Савченко В. В.** Проверка однородности выборочных данных в задачах спектрального оценивания // Радиотехника и электроника. 1999. **44,** № 1. С. 65.
5. **Кульбак С.** Теория информации и статистика. М.: Наука, 1967.
6. **Деврой Л., Дьерфи Л.** Непараметрическое оценивание плотностей: L_1-подход. М.: Мир, 1988.
7. **http**://www.rtsnet.ru

Нижегородский государственный
лингвистический университет,
E-mail: svv@lunn.sci-nnov.ru

Поступила в редакцию
10 апреля 2000 г.