УДК 330.46(045)

ГИБРИДНЫЕ И СЕЛЕКТИВНЫЕ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ФИНАНСОВЫХ ИНДЕКСОВ В РАМКАХ РАНДОМИЗИРОВАННОЙ КОЛЛОКАЦИИ

Бабешко Людмила Олеговна, *д-р экон. наук, профессор Департамента анализа данных, принятия решений и финансовых технологий, Финансовый университет, Москва, Россия*
babeshko_ls@mail.ru

Ясакова Анна Михайловна, *аспирант Департамента анализа данных, принятия решений и финансовых технологий, Финансовый университет, Москва, Россия*
yasakova.ann@gmail.com

Работа нацелена на повышение точности прогнозов финансовых индексов, которые отражают ситуацию на рынке в целом и являются важнейшими индикаторами российской экономики. Повышение точности достигается построением гибридных прогнозов, базовый набор которых включает модели рандомизированной коллокации и тривиального прогнозирования. Механизм рандомизации в коллокационном подходе, основанный на селективной процедуре комбинированного прогнозирования, предназначен для обоснованного выбора между моделями чистой и параметрической коллокации. Оценки весовых коэффициентов гибридных прогнозов строятся по прогнозам из базового списка моделей в классе линейных процедур, удовлетворяющих стандартным требованиям оптимальности: несмещенность ошибок прогнозирования и минимизация их дисперсий. Алгоритмы разработанных моделей реализованы в программной среде *R* и апробированы на данных индекса РТС за 2016 г.
Ключевые слова: рандомизированная коллокация; комбинированные прогнозы; селективная модель; гибридная модель; весовые коэффициенты.

HYBRID AND SELECTIVE MODELS OF FINANCIAL INDEX FORECASTING IN THE RANDOMIZED COLLOCATION FRAMEWORK

Babeshko Ludmila O., *ScD (Economics), full professor of the Data Analysis Department, Financial University, Moscow, Russia*
babeshko_ls@mail.ru

Yasakova Anna M., *post-graduate student of the Data Analysis Department, Financial University, Moscow, Russia*
yasakova.ann@gmail.com

The study aims to improve the accuracy of forecasting financial indices that reflect the general market situation and are important indicators of the Russian economy. The enhanced accuracy is achieved by making hybrid forecasts that basically include a set of randomized collocation and trivial forecast models. The randomization mechanism in the collocation approach based on the selective procedure of combined forecasting is used to make a justified choice between the models of pure and parametric collocation. The weighted coefficients of hybrid forecasts are evaluated based on forecasts of the basic model list in the class of linear processes that meet the standard optimality requirements: unbiased prediction errors and minimization of their variances. Algorithms of developed models are implemented in the R software environment and tested on the RTS index data for 2016.
Keywords: randomized collocation; combined forecasts; selective model; hybrid model; weighted coefficients.

Типы моделей прогнозирования, основанные на комбинированных методах

Фондовый рынок является сферой, где финансовые инструменты используются для мобилизации сбережений в экономике и их конвертации в инвестиционные ресурсы, направляемые в наиболее эффективные мероприятия экономического развития. При прогнозировании характеристик финансовых инструментов необходимо учитывать их волатильность. Изменениям подвергаются не только уровни временных рядов показателей, но и их динамические свойства. Выбор одной модели для построения прогнозов приводит, как правило, к большим дисперсиям оценок ее параметров и соответственно ошибкам прогнозов. Вот почему для повышения точности прогноза применяются комбинированные методы [1, 2].

Модели прогнозирования, основанные на комбинированных методах, подразделяются на модели селективного типа и гибридные [3]. В моделях селективного типа в качестве критерия отбора одной модели из базового набора используются такие показатели, как средняя квадратическая ошибка прогнозирования, информационные критерии (Акаике, Шварца, Хеннана — Куина), коэффициент Тейла.

В гибридных моделях при формировании комбинированного прогноза учитываются веса индивидуальных прогнозов, полученных по моделям, включенным в базовый список. В данной работе комбинированные модели как селективного, так и гибридного типа, основаны на коллокационном подходе прогнозирования финансовых индексов.

Прогнозирование финансовых индексов в рамках стационарной модели логарифмической прибыли

Значение финансового индекса в текущий момент S_t определяется его начальным значением S_0 и коэффициентом наращения:

$$S_t = S_0 \cdot e^{H_t}, \tag{1}$$

где

$$H_t = h_0 + \ldots + h_i + \ldots + h_t$$

— логарифмическая прибыль за период t;

$$h_i = \ln(S_i / S_{i-1}) \text{ при } i > 0,\ h_i = 0 \text{, при } i = 0$$

— логарифмическая прибыль в момент $i \geq 0$.

Последовательность значений логарифмической прибыли $\left(h_i\right)_{i \geq 1}$ предполагается стационарной с характеристиками:

$$E\left\{h\right\} = m,\ Cov\left\{h_i, h_{i+k}\right\} = C_{hh}\left(\tau\right).$$

Значение индекса на некоторый момент в будущем ($t = n + k$), в соответствии с формулой (1), равно:

$$S_{n+k} = S_0 \cdot e^{H_{n+k}} = S_0 \cdot e^{H_n} \cdot e^{\Delta H} = S_n \cdot e^{\Delta H}, \tag{2}$$

где S_n — последнее наблюдаемое значение, n — число наблюдений,

$$\Delta H = \sum_{i=n+1}^{n+k} h_i \tag{3}$$

— приращение логарифмической прибыли за период упреждения k. Таким образом, для прогноза финансового индекса (2) на момент $t = n + k$,

$$\hat{S}_{n+k} = S_n \cdot e^{\Delta \hat{H}} = S_n \cdot \exp\left\{\Delta \hat{H}\right\}, \tag{4}$$

необходимо оценить величину приращения логарифмической прибыли $\Delta \hat{H}$. Для оценки линейного функционала (3) в работе используется рандомизированный алгоритм коллокации, основанный на проверке нулевой гипотезы $H_0 : m = 0$ против альтернативной $H_1 : m \neq 0$ [4, 5].

Комбинированное прогнозирование финансовых индексов в рамках коллокационных моделей

При справедливости нулевой гипотезы $H_0 : m = 0$ прогноз приращения логарифмической прибыли выполняется при помощи модели чистой коллокации (оптимальный средний квадратический прогноз Колмогорова–Винера):

$$\Delta \hat{H} = C_{\Delta H, h} \cdot C_{hh}^{-1} \cdot h, \tag{5}$$

где $h = (h_1, h_2, \ldots, h_n)^T$ — известные значения уровней динамического ряда;

C_{hh} — автоковариационная функция уровней ряда;

$C_{h,\Delta H}$ — вектор взаимных коваріаций значений h_i, $i = 1, \ldots, n$ стационарного динамического ряда и значения линейного функционала ΔH.

Точность оценки (5) характеризуется автоковариационной матрицей ошибок прогнозов

$$C_{ee} = C_{\Delta H} - C_{\Delta H,h} C_{hh}^{-1} C_{h,\Delta H}, \quad (6)$$

где $C_{\Delta H}$ — автоковариационная матрица оценок приращения логарифмической прибыли.

В случае справедливости альтернативной гипотезы $H_1 : m \neq 0$ оценка линейного функционала логарифмической прибыли (3) выполняется посредством модели параметрической коллокации:

$$\Delta\hat{H} = \hat{m} \cdot k + C_{\Delta H,h} \cdot C_{hh}^{-1} \cdot (h - I \cdot \hat{m}), \quad (7)$$

где $\hat{m}$ — оценка математического ожидания стационарного случайного процесса логарифмической прибыли, полученная по выборочным данным $h = (h_1, h_2, \ldots h_n)^T$. Дисперсия ошибки оценки (7) вычисляется по формуле

$$\sigma_e^2 = k^2 \cdot \sigma_{\hat{m}}^2 + \sigma_{\Delta H}^2 - q \cdot C_{h,\Delta H} + q \cdot I \cdot \sigma_{\hat{m}}^2 \cdot I^T \cdot q^T -$$

$$2k \cdot \sigma_{\hat{m}}^2 \cdot I^T \cdot q^T, \quad (8)$$

где $\sigma_{\hat{m}}^2$ — дисперсия оценки $\hat{m}$;

$q = C_{\Delta H,h} \cdot C_{hh}^{-1}$.

Для практической реализации алгоритмов коллокации используются аналитические модели ковариационных функций [6]:

$$C(\tau) = \sigma^2 \cdot e^{-\alpha \cdot |\tau|}, \quad (9)$$

$$C(\tau) = \sigma^2 \cdot \frac{\sin \alpha \cdot \tau}{\alpha\tau}, \quad (10)$$

$$C(\tau) = \sigma^2 \cdot e^{-\alpha \cdot |\tau|} \cos\beta \cdot \tau, \ \alpha > 0, \quad (11)$$

где σ^2 — дисперсия стационарного процесса;

α, τ — параметры.

Параметры моделей (9)—(11) оцениваются методом существенных параметров, в качестве которых используется набор характеристик ковариационной функции:

$\sigma^2 = C(0)$ — дисперсия процесса $(h_i)_{i \geq 1}$,

$\tau_{0,5}$ — радиус корреляции — значение аргумента функции $C(\tau)$, при котором ее значение равно половине дисперсии: $C(\tau_{0,5}) = C(0)/2$,

τ_0 — наименьший положительный нуль функции $C(\tau)$, если она имеет нули. Связь параметров ковариационных функций с существенными параметрами устанавливается исходя из их спецификаций.

Прогнозные модели (4)–(8), построенные на основе аппроксимирующих функций (9)–(11), имеют одинаковую структуру, но отличаются значениями параметров, поэтому будут давать различные результаты. Для выбора одной из них, в рамках селективного подхода, в качестве критерия отбора используются дисперсии ошибок прогноза, которые вычисляются по формулам (6) и (8) рандомизированного алгоритма.

При построении комбинированных прогнозов, в рамках гибридного подхода, в качестве базового набора прогнозных моделей используются модели рандомизированной коллокации: прогнозы, полученные при помощи ковариационных функций (9)–(11), комбинируются путем выбора оптимальных весов [7].

Для определения оптимальных весов $g = \left(g_1, g_2, \ldots, g_m\right)^T$ комбинированного прогноза

$$\hat{Y} = \sum_{i=1}^{m} g_i \hat{Y}_i = g_1\hat{Y}_1 + g_2\hat{Y}_2 + \ldots + g_m\hat{Y}_m, \quad (12)$$

представим структуру его составляющих в виде суммы

$$\hat{Y}_i = Y + e_i, \ i = 1, \ldots, m, \quad (13)$$

где Y — n-мерный вектор истинных значений финансового индекса;

e_i — n-мерный вектор ошибок прогноза i-й модели базового набора,

$$E\left\{e_i\right\} = 0, \ i = 1, \ldots, m. \quad (14)$$

Запишем выражение для ошибки комбинированного прогноза (13):

Таблица 1

Значения индекса РТС за период с 11.01.2016 по 07.03.2016 г.

Дата	Индекс РТС	Дата	Индекс РТС
11.01.2016	11 067,9	08.02.2016	11 540,37
12.01.2016	11 131,43	09.02.2016	11 383,00
13.01.2016	11 089,72	10.02.2016	11 435,65
14.01.2016	11 069,65	11.02.2016	11 303,70
15.01.2016	10 588,08	12.02.2016	11 407,53
18.01.2016	10 685,95	15.02.2016	11 504,98
19.01.2016	10 849,47	16.02.2016	11 604,72
20.01.2016	10 752,65	17.02.2016	11 726,85
21.01.2016	11 074,11	18.02.2016	11 985,36
22.01.2016	11 369,15	19.02.2016	11 883,72
25.01.2016	11 368,1	22.02.2016	12 091,38
26.01.2016	11 276,47	24.02.2016	11 910,29
27.01.2016	11 569,19	25.02.2016	11 953,80
28.01.2016	11 770,39	26.02.2016	12 042,04
29.01.2016	11 874,94	29.02.2016	12 201,68
01.02.2016	11 762,08	01.03.2016	12 220,54
02.02.2016	11 644,88	02.03.2016	12 135,21
03.02.2016	11 541,63	03.03.2016	12 320,71
04.02.2016	11 854,41	04.03.2016	12 465,37
05.02.2016	11 801,67	07.03.2016	12 680,63

Источник: Информационно-аналитическая система *Bloomberg Professional*, (дата обращения: 13.04.2016).

$$e_p = \hat{Y} - Y = g_1\hat{Y}_1 + g_2\hat{Y}_2 + \ldots + g_m\hat{Y}_m - Y =$$

$$= g_1(Y + e_1) + g_2(Y + e_2) + \ldots + g_m(Y + e_m) - Y =$$

$$= \sum_{i=1}^{m} g_i Y - Y + \sum_{i=1}^{m} g_i e_i = \left(\sum_{i=1}^{m} g_i - 1\right) Y + \sum_{i=1}^{m} g_i e_i, \quad (15)$$

и определим ее математическое ожидание с учетом (14):

$$E\left\{e_p\right\} = \left(\sum_{i=1}^{m} g_i - 1\right) E\left\{Y\right\} + \sum_{i=1}^{m} g_i E\left\{e_i\right\} = \\ = \left(\sum_{i=1}^{m} g_i - 1\right) E\left\{Y\right\}. \quad (16)$$

Как следует из (16), требование несмещенности ошибки комбинированного прогноза

$$E\left\{e_p\right\} = E\left\{\hat{Y} - Y\right\} = 0$$

сводится к условию нормировки коэффициентов

$$\sum_{i=1}^{m} g_i = 1, \quad (17)$$

при котором несмещенная ошибка комбинированного прогноза равна

$$e_p = e \cdot g, \quad (18)$$

где e — $(n \times m)$ — матрица, столбцами которой являются векторы ошибок прогнозов, полу-

Таблица 2

Результаты прогнозирования

Дата	Значение индекса РТС	Модель (0)	Модель (9)	Модель (10)	Модель (11)	Комбинированный прогноз
1	2	3	4	5	6	7
08.02.2016	11 540,37	11 801,67	11 803,42	11 807,19	11 787,39	11 818,77
09.02.2016	11 383,00	11 540,37	11 533,68	11 513,62	11 568,28	11 515,69
10.02.2016	11 435,65	11 383,00	11 371,98	11 334,78	11 450,95	11 329,59
11.02.2016	11 303,70	11 435,65	11 399,53	11 305,49	11 528,28	11 328,31
12.02.2016	11 407,53	11 303,70	11 311,15	11 332,35	11 285,64	11 320,51
15.02.2016	11 504,98	11 407,53	11 417,35	11 442,50	11 393,82	11 425,92
16.02.2016	11 604,72	11 504,98	11 498,79	11 491,99	11 475,85	11 510,8
17.02.2016	11 726,85	11 604,72	11 645,93	11 722,06	11 516,51	11 782,74
18.02.2016	11 985,36	11 726,85	11 749,83	11 800,32	11 699,75	11 784,5
19.02.2016	11 883,72	11 985,36	11 982,00	11 974,98	11 993,22	11 972,08
22.02.2016	12 091,38	11 883,72	11 878,30	11 866,69	11 874,13	11 885,82
24.02.2016	11 910,29	12 091,38	12 114,41	12 226,81	12 087,64	11 979,37
25.02.2016	11 953,80	11 910,29	11 921,29	11 952,91	11 850,50	11 968,36
26.02.2016	12 042,04	11 953,80	11 957,10	11 963,36	11 919,79	11 995,32
29.02.2016	12 201,68	12 042,04	12 032,81	12 003,46	12 081,93	12 010,51
01.03.2016	12 220,54	12 201,68	12 193,40	12 167,68	12 257,27	12 151,63
02.03.2016	12 135,21	12 220,54	12 216,55	12 207,81	12 251,54	12 183,48
03.03.2016	12 320,71	12 135,21	12 148,86	12 184,10	12 049,36	12 228,65
04.03.2016	12 465,37	12 320,71	12 310,37	12 227,50	12 318,53	12 451,38
07.03.2016	12 680,63	12 465,37	12 442,49	12 284,87	12 456,07	12 698,38

ченных при помощи моделей базового набора. Дисперсия ошибки прогноза (18)

$$Var\left\{e_p\right\} = Cov\left(e_p, e_p\right) = g^T C_{ee} g \tag{19}$$

определяется квадратичной формой с матрицей C_{ee} (матрицей взаимных ковариаций ошибок прогнозов по моделям базового набора).

Таким образом, задача определения оптимальных весовых коэффициентов (оптимальных в смысле несмещенности ошибок прогноза и минимальности их дисперсий) — это задача на условный экстремум с целевой функцией вида

$$L\left(g, \lambda\right) = g^T C_{ee} g - 2\lambda\left(g^T I - 1\right), \tag{20}$$

где λ — множитель Лагранжа;

I — единичный вектор-столбец.

Необходимые условия экстремума первого порядка для функции (20) приводят к следующей системе уравнений:

$$\begin{cases} \dfrac{\partial L}{\partial g^T} = 2C_{ee} g - 2\lambda I = 0 \\ \dfrac{\partial L}{\partial \lambda} = 2(g^T I - 1) = 0 \end{cases}. \tag{21}$$

Решением системы (21) является вектор g, который минимизирует дисперсию ошибок комбинированного прогноза (19) и удовлетворяет ограничению (17):

Таблица 3

Коэффициенты комбинированного прогноза

Номер модели	g_i
0	−4,4537
8	8,8383
9	−2,3584
10	−1,0262

Таблица 4

Оценки средних квадратических ошибок прогнозов базовых моделей и их комбинации

Число прогнозов	Модель (0)	Модель (1)	Модель (2)	Модель (3)	Комбинированный прогноз
20	150,98	149,41	173,11	174,18	118,94

$$g = \left(I^T C_{ee}^{-1} I\right)^{-1} C_{ee}^{-1} I. \quad (22)$$

Для построения матрицы C_{ee}, выполняется оценка значений эмпирических ковариационных и взаимных ковариационных функций ошибок прогнозов, которые затем аппроксимируются моделями (9)–(11).

Реализация алгоритма комбинированного прогнозирования

Описанный алгоритм комбинированного прогнозирования реализован в векторно-ориентированной среде *R* и апробирован на данных индекса РТС за 2016 г. (*табл. 1*).

При настройке коллокационных моделей в качестве обучающей выборки выбрано скользящее окно, включающее 20 наблюдений, период упреждения — 1 шаг.

В базовый набор включены модели тривиального прогнозирования (столбец 3 *табл. 2*) и модели рандомизированной коллокации с применением ковариационных функций (9)—(11) (столбцы 4–6 *табл. 2*). Результаты комбинированного прогноза, полученного в рамках гибридной модели, приведены в столбце 7 *табл. 2*. Весовые коэффициенты для гибридного прогноза (12), вычисленные по формуле (22), представлены в *табл. 3*.

Средние квадратические ошибки прогнозов показывают оптимальность процедуры гибридного прогнозирования (*табл. 4*) по сравнению с индивидуальными прогнозами, включенными в базовый набор.

Весовые коэффициенты, приведенные в *табл. 3*, совпадают с оценками параметров регрессионной модели при наличии ограничений (17), в которой эндогенной переменной является вектор *Y*, а регрессорами — его индивидуальные прогнозы, построенные по моделям базового набора [8]. Однако практическая реализация комбинированной модели с весами (22) удобнее, чем решение аналогичной задачи в рамках эконометрической модели с ограничениями на параметры [9].

ЛИТЕРАТУРА

1. Bates J.M., Granger C.W.J. The combination of forecasts. Operation Research Quarterly, 1969, vol. 20, No. 4, pp. 451–468.
2. Granger C.W.J. Invited review: combining forecasts — twenty years later. *Journal of Forecasting*, 1989, No. 8, pp. 167–173.
3. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. М.: Финансы и статистика, 2003. 416 с.
4. Бабешко Л.О. Коллокационные модели прогнозирования в финансовой сфере. М.: Экзамен, 2001. 288 с.
5. Бывшев В.А., Бабешко Л.О., Клапко А.О. Прогнозирование динамических рядов финансово-экономической информации рандомизированным алгоритмом коллокации // Управление риском. 2004. № 1. С. 35–39.
6. Бывшев В.А., Бабешко Л.О., Арсеньева Л.В. Алгоритм оценивания основных инвестиционных характеристик финансовых активов при помощи оптимальной статистической процедуры Эйткена // Управление риском. 2000. № 4. С. 31–37.
7. Бабешко Л.О., Ясакова А.М. Комбинированные модели прогнозирования финансовых индексов // Актуальные вопросы в научной работе и образовательной деятельности. Сборник научных трудов по материалам Международной научно-практической конференции: в 10 томах. Тамбов, 2015. Т. 4. С. 18–22.
8. Бабешко Л.О., Ясакова А.М. Гибридные модели с ограничением на параметры в рамках коллокационного подхода // Сборник статей Международной научной конференции, Орловский государственный университет. Воронеж, 2015. С. 74–80.
9. Бабешко Л.О. Эконометрическое прогнозирование по разнородной информации / Л.О. Бабешко. М.: Вега-Инфо, 2016.— 232 с. ISBN 978–5–91590–024–9.

REFERENCES

1. Bates J.M., Granger C.W.J. The combination of forecasts. Operation Research Quarterly, 1969, vol. 20, No. 4, pp. 451–468.
2. Granger C.W.J. Invited review: combining forecasts — twenty years later. *Journal of Forecasting*, 1989, No. 8, pp. 167–173.
3. Lukashin Y.P. An adaptive methods of short-term time series forecasting [Adaptivnye metody kratkosrochnjgo prognozirovaniya]. Moscow, Finansy i statistika — Finance and statistics, 2003, 416 p.
4. Babeshko L.O. Collocation forecasting models in the financial sector [Kollokacionnye modely prognozirovaniya v finansovoy sfere], Moscow, 2001, 288 p.
5. Byvshev V.A., Babeshko L.O., Klapko A.O. Prediction of time series of financial and economic information within randomized collocation [Prognozirovaniye dynamicheskih rjadov finansovo-ekonomicheskoj informacii randomizirovannym algoritmom kollokacii]. *Upravlenie riskom — Risk management*, 2004, No. 1, pp. 18–22.
6. Byvshev V.A., Babeshko L.O., Arsenyeva L.V. Estimator of basic investment characteristics of the financial assets within optimal statistical Aitken procedure [Algoritm ocenivanija osnovnyh investicionnyh haracteristic finansovyh aktivov pri pomoschi optimal'noj statisticheskoj procedury Aitkena]. *Upravlenie riskom — Risk management*, 2000, No. 4, pp. 31–37.
7. Babeshko L.O., Yasakova A.M. Combined forecasting models of financial indices [Kombinirovannye modely prognozirovaniyja finansovyh indeksov: sbornik statei Mezhdunarodnoi nauchnoi konferensii]. Tambov, 2015, pp. 18–22.
8. Babeshko L.O., Yasakova A.M. Hybrid models with restriction on the parameters within the collocation approach [Gibridnye modeli s ogranichenijami na parametry v ramkah kollokaciionnogo podhoda: sbornik statei Mezhdunarodnoi nauchnoi konferensii]. Orel, State University. Voronezh, 2015, pp. 74–80.
9. Babeshko L.O. Econometric forecasting for heterogeneous information [Ekonometricheskoye prognozirovaniye po raznorodnoy informatsii]. Moscow, Vega-Info, 2016. 232 c. ISBN 978–5–91590–024–9.