

ФИНАНСОВЫЕ РЫНКИ

Е. В. Дорохов

канд. экон. наук, преподаватель кафедры статистики Московского государственного университета им. М. В. Ломоносова

МОДЕЛИРОВАНИЕ ДИНАМИКИ ОТЕЧЕСТВЕННОГО РЫНКА АКЦИЙ

Прогнозирование индекса РТС с помощью методов стохастической аппроксимации

Традиционно рассматриваемые в статистике прикладные методы прогнозирования, как правило, изначально были разработаны для анализа экономической деятельности (Хольт, Уинтерс) или в рамках теории управления, применяющейся в технике (Бокс и Дженкинс). В то же время в течение нескольких лет инженеры, исходя из основополагающей работы (Калман, Бьюси, 1961), следовали независимым путем, развивая рекуррентные методы для мультисенсорных систем и используя их в следящих механизмах или системах автоматического регулирования. В действительности возникающие при этом проблемы во многих отношениях аналогичны тем, что встают при прогнозировании более привычными экономико-статистическими методами, хотя при этом и существуют важные различия:

— в технике часто имеется большая априорная информация, которая дает возможность получать хорошие оценки параметров системы, или, альтернативно, имеется обширная информация относительно других систем того же типа, которая служит той же цели;

— входная информация обычно многомерна;

— иногда текущий «прогноз» нельзя проверить наблюдением (например, при слежении за движущимся объектом из некоторой неподвижной точки наблюдения поступающими сигналами могут быть значения углов, отражения радара и т. д.; цель при этом состоит в том, чтобы получить прогноз положения и скорости объекта, который непосредственно наблюдать нельзя).

Тем не менее ряд исследований подтвердил правомерность применения методов прогнозирования, развитых в инженерной теории управления, к задачам изучения развития временных рядов статистических показателей экономической и финансовой природы. Рассмотрим возникающие при этом вопросы в приложении к проблеме исследования динамики индекса фондового рынка России.

В наиболее общей формулировке мы будем рассматривать прикладную задачу восстановления некоторых характеристик (параметров) объекта по имеющимся экспериментальным данным. В случае, если процесс измерения искажается вероятностными «шумами», т. е. решается задача наблюдения в условиях случайных возмущений, принято говорить о применимости *теории стохастического оценивания*. Процесс выбора характеристик (параметров) модели из заданного класса для наилучшего описания результатов представляет собой одно из

довольно общих определений понятия оценивания. На практике процесс оценивания часто удается связать с какой-нибудь количественной характеристикой качества оценивания и при выборе оценок, естественно, стараться минимизировать отрицательное влияние погрешностей.

Остановимся на основных положениях теории стохастического оценивания, важных с точки зрения решения задачи прогнозирования¹.

На первом плане в статистической теории связи стоит проблема различения сигналов и восстановления сообщений. Эта проблема тесно связана с оценкой параметров, от которых могут зависеть принимаемые сигналы. Хотя первоначально преобразование сигналов и оценивание их параметров изучалось со специальными целями, вскоре было установлено, что эти проблемы естественно укладываются в рамки статистики. Важнейшей категорией при этом является понятие *фильтрации* — процесса выделения сигнала из его аддитивной смеси с шумом². Задача фильтрации сигнала состоит в построении оценки значения сигнала (случайного процесса) в некоторый момент времени по наблюдениям связанного с ним другого случайного процесса³. В частности, широкое применение находят линейные фильтры, задаваемые линейным преобразованием стационарного случайного процесса как функции параметра времени.

Под *оптимальной фильтрацией* понимаются алгоритмы обработки реализаций случайных процессов, направленные на максимальное (в смысле некоторого критерия) подавление помех, зашумляющих (обычно аддитивно) полезный сигнал. Фундамент теории оптимальной фильтрации составляют метод Винера-Колмогорова (Граничин, 2003) и его рекуррентные модификации на нестационарный случай, известные под общим названием фильтра Калмана-Бьюси (Браммер, Зиффлинг, 1982). Теория Винера-Колмогорова в существенной степени базируется на методе наименьших квадратов. Оценивание параметров в этой

¹ Первым важным результатом в этой области явилась работа К. Гаусса 1821 г., в которой он предложил рекуррентный вариант процедуры метода наименьших квадратов, позволяющий корректировать ранее вычисленную оценку с учетом вновь поступивших дополнительных измерений без необходимости повтора всех предшествующих вычислений.

Фундамент современного состояния теории оценивания заложен Р. Фишером в 1920—1930-х гг. прошлого века (Fisher, 1935). Тщательно рассматривая основания теории оценивания, он избавил ее от жестких ограничений, существовавших с момента появления работ К. Гаусса. Обобщения его теории привели, в частности, к развитию современных методов непараметрического и робастного оценивания, в которых точная природа распределения вероятностей оцениваемых случайных величин не предполагается известной.

Одновременно с формализацией и развитием математической статистики проводились исследования в далеких, казалось бы, от нее областях. В частности, специалисты по линиям связи имели дело с задачей синтеза устройств, позволяющих эффективно обнаруживать присутствие или отсутствие сигнала, наблюдаемого на фоне помехи. Быстрое развитие теории связи привело к необходимости учета воздействия помех на распространение и прием сигналов. Первые попытки уменьшить нежелательное воздействие помех были связаны с введением методов расчета фильтров, позволяющих оценить спектр мощности полезного сигнала. Эти попытки делались в перспективном направлении, но их сдерживала неразвитость теории фильтрации. Математические основы этой теории только закладывались: в начале 1930-х гг. А. Хинчин и Н. Винер создали теорию гармонического анализа случайных функций, центральное место в которой занимает теорема о спектральном представлении стационарных процессов. Основы теории оптимальной фильтрации стационарных процессов были заложены в работах А.Н. Колмогорова и Н. Винера. Винер показал, в частности, что теория оценивания может быть применена для синтеза электрического фильтра, который обеспечивает наилучшее выделение сигнала при наличии стационарной помехи. Основной акцент он делал не столько на рассмотрение частотных спектров сигналов, сколько на их обработку как стохастических процессов. Центральным местом теории оптимальной фильтрации Винера-Колмогорова является уравнение Винера-Хопфа, решение которого непосредственно связано с синтезом оптимального фильтра.

² Собственно говоря, благодаря сути указанной задачи — выделению сигнала из «смеси» сигнала и случайного шума — и возник термин «фильтрация».

³ Если статистические характеристики случайных процессов известны, то, как правило, пытаются построить оценку с наименьшей среднеквадратической ошибкой.

теории происходит на основе обработки последовательно поступающих входных данных, являющихся некоторой траекторией стохастического процесса.

Важную роль в теории оптимальной фильтрации играют *стохастические дифференциальные уравнения* вида:

$$dx = a(x, \tau)dt + \sigma(x, \tau)dw_t,$$

где $dw_t = w(\omega, \tau)$ — винеровский случайный процесс¹.

Решением этого стохастического дифференциального уравнения на промежутке $[t, T]$ называется случайный процесс $x(\omega, \tau)$, удовлетворяющий следующим условиям:

все траектории случайного процесса $x(\omega, \tau)$ непрерывны на промежутке $[t, T]$;

для любого $\tau \in [t, T]$ при достаточно малом $\Delta\tau$:

$$x_{\tau+\Delta\tau}(\omega) - x_\tau(\omega) \approx a(x_\tau(\omega), \tau)\Delta\tau + \sigma(x_\tau(\omega), \tau)(w_{\tau+\Delta\tau}(\omega) - w_\tau(\omega)).$$

Фильтр Калмана представляет собой алгоритм для вычисления оптимальной в среднеквадратическом смысле оценки случайного процесса x_t , определяемого стохастическим дифференциальным уравнением

$$dx_t = a(t)x_t dt + b(t)dw_t,$$

по наблюдениям другого случайного процесса y_s в моменты $s \in [0, t]$. При этом случайный процесс y_t допускает стохастический дифференциал

$$dy_t = A(t)x_t dt + B(t)dv_t$$

(иными словами x_t — сигнал, y_t — наложившийся на него случайный шум).

Этот подход, по существу представляющий собой рекуррентный вариант метода наименьших квадратов, в частных случаях исследовался ранее и другими авторами, но именно с работ Р. Калмана и Р. Бьюси в начале 1960-х гг. началось широкое развитие методов теории *рекуррентного (последовательного) оценивания*, в рамках которой задача оптимальной фильтрации получила существенное продвижение. Подчеркнем, что рекуррентные процедуры оценивания (фильтр Калмана-Бьюси) оказались применимыми и в случае нестационарных процессов.

Работы Калмана по рекуррентному оцениванию появились в связи с необходимостью оптимального оценивания вектора состояния линейных нестационарных систем. Оценивание производилось по наблюдениям за зашумленным компонентом вектора состояния. При этом в теоретическом плане существенным моментом является линейная зависимость наблюдаемого процесса от оцениваемого параметра (линейная фильтрация). Вместе с тем многие практические задачи приводят к нелинейной зависимости данных наблюдения от оцениваемого параметра. Этот раздел теории оценивания — *нелинейная фильтрация* — развит значительно меньше, основные идеи были выдвинуты в 1960 г. Р. Л. Стратоновичем (Стратонович, 1966). Предложенная им рекуррентная процедура оценивания в линейном случае преобразуется в фильтр Калмана-Бьюси.

¹ Случайный процесс $w(\omega, \tau)$, определенный на промежутке $t \geq t_0$, называется *винеровским случайным процессом (процессом броуновского движения)*, если выполнены следующие условия:

приращения случайного процесса $w(\omega, \tau)$ независимы;

приращения $w_t(w) - w_s(w)$, $t > s$ имеют нормальное распределение с нулевым средним и $\sigma^2 = t - s$;

все траектории случайного процесса $w(\omega, \tau)$ непрерывны на промежутке $t \geq t_0$.

Совершенствование и доступность вычислительной техники стимулировали развитие рекуррентных схем оценивания и в других классических разделах статистики. Так, получили широкое признание процедуры стохастической аппроксимации Роббинса-Монро (Robbins, Monro, 1951) и Кифера-Вольфовица (Kiefer, Wolfowitz, 1952). Ряд исследователей показали широкую применимость рекуррентных стохастических алгоритмов в задачах оценивания, идентификации, распознавания образов, оптимизации управления (Цыпкин, 1968; Kushner, Yin, 1997).

Методы *стохастической аппроксимации*, к числу которых относятся и хорошо известные в статистике адаптивные методы, предназначены для решения широкого класса параметрических и непараметрических задач статистического оценивания и основаны на последовательном уточнении оценки при увеличении числа наблюдений (то есть новое значение оценки представляет собой поправку к уже имеющейся оценке, основанную на новом наблюдении) (Вазан, 1972). Применительно к параметрическим задачам стохастическая аппроксимация часто называется *рекуррентным оцениванием*. Рекуррентное оценивание применяют также для оценивания параметров линейной регрессии, в случае независимых шумов, для непрерывных по времени процессов (Фомин, 1984).

Суть алгоритма методов стохастической аппроксимации заключается в следующем. На основе имеющихся начальных данных классическими методами строится предварительная модель (это могут быть и система дифференциальных уравнений, и оптимизационная задача, и т. д.). При поступлении новой информации модель проходит проверку на предмет соответствия предварительно наложенным на нее ограничениям. Вид последних определяется реальным ходом моделируемого процесса и дополнительными условиями, не позволяющими отклонениям расчетных значений от фактических выходить за определенные рамки доверительного коридора¹. Если модель не отвечает предъявляемым требованиям, в силу вступает блок адаптации, имитирующий естественный процесс приспособления экономического «организма» к условиям существования². Алгоритм работает до исчерпания массива информации.

Проиллюстрируем применение методов рекуррентного оценивания в задаче прогнозирования динамики индекса фондового рынка. Классические статистические модели (регрессионные, ARIMA и т. п.), основанные на восстановлении условных распределений будущих значений процесса относительно прошлых, не обладают универсальностью по отношению к входной информации. Так, если имеет место существенно нестационарное (не сводимое к стационарному) изменение параметров, то применение подобных моделей становится невозможным. Такого рода нестационарность вызвана влиянием внешних неподконтрольных факторов, и на определенных временных этапах она играет доминирующую роль в формировании, например, изучаемых индексов. Это особенно характерно для нестабильной экономики, в условиях которой статистические показатели зачастую претерпевают скачкообразное изменение.

Рассмотрим процесс изменения индекса РТС. Предположим, что значения индекса постоянно колеблются около своего тренда вследствие аддитивного воздействия на него случайных факторов, и будем считать, что приращения индек-

¹ Точность описания фактических значений с помощью модели отражает ошибка аппроксимации. Очевидно, сведение этой ошибки к нулю — идеальный вариант. Поэтому допускается требование, чтобы эта ошибка не превышала некоторого допустимого значения. Причем этим допустимым значением может быть и среднее абсолютное отклонение, и средний квадрат отклонений, и границы, определенные с помощью *t*-статистики, и другие критерии, применяемые в зависимости от апостериорно выявленного характера исследуемого процесса.

² Иначе говоря, адаптацию следует проводить только в том случае, когда текущее значение отклонения расчетных значений от фактических превышает допустимое значение. Целью адаптации является изменение параметров модели таким образом, чтобы расчетные значения вновь удовлетворительно описывали реальный ряд значений.

са — независимые величины. Учитывая, что на значение индекса оказывает влияние множество случайных факторов, а воздействие каждого отдельного фактора незначительно, и между собой они, как правило, независимы, на основании центральной предельной теоремы можно утверждать, что в краткосрочном периоде закон распределения приращений процесса изменения индекса близок к нормальному. Будем считать, что трендовая составляющая не оказывает влияния на случайную, и наоборот, случайные изменения не влияют на характер тренда. Тогда выражение для индекса можно записать в разностной форме:

$$p(t + \Delta t) - p(t) = f(t, p(t), z(t))\Delta t + \xi(t, p(t), \Delta t),$$

где $z(t)$ — совокупность параметров, $f(t, p(t), z(t))$ — скорость изменения индекса во времени, а $\xi(t, p(t), \Delta t)$ — флуктуационная составляющая из-за влияния большого числа независимых факторов.

Так как при небольших изменениях t незначительно изменяется и $p(t)$, то флуктуационную составляющую будем считать однородным случайным процессом:

$$\xi(t, p(t), \Delta t) = F(t, p(t))(\omega(t + \Delta t) - \omega(t)),$$

где $F(t, p(t))$ — среднее квадратическое отклонение, $\omega(t + \Delta t) - \omega(t)$ — приращение винеровского процесса.

При этом приращение индекса можно записать в виде:

$$p(t + \Delta t) - p(t) = f(t, p(t), z(t))\Delta t + F(t, p(t))\Delta\omega(t)$$

или в виде стохастического дифференциального уравнения:

$$dp(t) = f(t, p(t), z(t))dt + F(t, p(t))d\omega(t), \quad p(\tau) = p_0. \quad (1)$$

Любая реализация индекса есть решение этого уравнения и имеет общий вид:

$$p(t) = p_0 + \int_{\tau}^t f(t, p(t), z(t))dt + \int_{\tau}^t F(t, p(t))d\omega(t),$$

где интеграл по случайному процессу $d\omega(t)$ является стохастическим.

Для вычисления прогнозных значений необходимо воспользоваться законом распределения этого процесса либо его первыми моментами как числовыми характеристиками. Рассмотрим два способа прогнозирования такого случайного процесса. Прогнозное значение, полученное с помощью уравнения фильтра Калмана-Бьюси, будет оптимальным с точки зрения принципа максимума апостериорной вероятности. В случае, когда можно установить автокорреляционную функцию, прогнозирование целесообразно осуществлять, основываясь на разложении процесса в базисе собственных функций.

Ниже изложены структуры алгоритмов решения задачи прогнозирования в каждом из указанных подходов.

1. *Алгоритм прогнозирования на основе реализации фильтра Калмана-Бьюси.*

Мы предполагаем, что уравнение состояния представлено как (1), и, кроме того, задано уравнение наблюдения в виде:

$$y(t) = p(t) + v(t),$$

где $v(t)$ — гауссовский «белый шум», $y(t)$ — наблюдаемое значение.

Уравнение для оптимальной оценки записывается в виде [7]:

$$\frac{\partial p^*(t)}{\partial t} = f(t, p^*(t), z(t)) + K(t)[y(t) - p^*(t)] + \int_{t-T}^t \frac{E[\varphi(t)y(\tau_2)]}{E[y(t)y(\tau_1)]} H(t, \tau_2)y(\tau_2)d\tau_2, \quad (2)$$

где $p^*(t)$ — значение оценки, $K(t)$ — «коэффициент усиления» фильтра, определяемый из следующего соотношения:

$$K(t) = f(t, p(t), z(t)) \cdot \frac{R(t)}{R(\varphi(t)) + R(t)}. \quad (3)$$

В этом соотношении $R(t) = E[(p^*(t) - p(t))(p^*(t) - p(t))]$ — дисперсия ошибок оценки, для которой справедливо рекуррентное соотношение:

$$R(t+1) = [f(t, p(t), z(t)) - K(t)]^2 R(t) + R_v(t)K^2(t) + R_\varphi(t+1) + 2E \left[\int_{t-T}^t \frac{E[\varphi(t)y(\tau_2)]}{E[y(t)y(\tau_1)]} H(t, \tau_2)y(\tau_2)d\tau_2 \right] (p^*(t) - p(t)), \quad (4)$$

функции $R_v(t) = E[v(t)v(t)] = Q(t)\delta(t-\tau)$ и $R_\varphi(t+1)$ — известны.

Структура алгоритма прогнозирования включает следующие операции:

1. Задать начальные приближения из условия $p(\tau) = p^*(\tau) = p_0, R(\tau) = 0$;
2. Вычислить коэффициент усиления $K(t)$ по формуле (3);
3. Вычислить по формуле (4) значение $R(t+1)$;
4. Получить прогноз индекса $p^*(t+1)$ как численное решение уравнения (2).

Производя операции 2—4 при $t = \tau, \tau+1, \dots$, получим траекторию предсказанных на шаг вперед значений индекса.

2. *Алгоритм прогнозирования на основе применения разложения Карунена-Лоэва к решению уравнения (1).*

В этом случае значения индекса $p(t)$ представляем разложением в базисе собственных функций $\varphi_i(t)$ в виде сходящегося в среднем квадратическом ряда (Давенпорт, Рут, 1960):

$$p(t) = \sum_{i=1}^{\infty} \alpha_i \cdot \frac{\varphi_i(t)}{\sqrt{\lambda_i}}.$$

Тогда наилучшее предсказанное значение для индекса вычисляется по формуле:

$$p^*(t + \Delta t) = \sum_{i=1}^{\infty} \alpha_i \cdot \frac{\varphi_i(t + \Delta t)}{\sqrt{\lambda_i}}, \quad (5)$$

где $\varphi_i(t)$ — собственные функции интегрального уравнения

$$\int_0^T r(t, \tau)\varphi(\tau)d\tau = \lambda \cdot \varphi(t), \quad (6)$$

где λ_i — соответствующие собственные значения, α_i — случайные величины, такие, что математические ожидания

$$E[\alpha_i] = 0, E[\alpha_i \alpha_j] = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

где $r(t, \tau)$ — автокорреляционная функция случайного процесса $p(t)$, оцениваемая по статистическим данным за предыдущие периоды.

Для решения интегрального уравнения (6) можно использовать метод Ритца (Стренг, Фикс, 1977).

Структура алгоритма прогнозирования включает следующие операции:

1. Вычислить автокорреляционную матрицу $r(t, \tau)$;
2. Для полученной матрицы $r(t, \tau)$ определить собственные функции и собственные значения;
3. Применить формулу (5) для получения прогноза.

Существенным моментом при выборе массива исходных данных является необходимость исследования имеющихся наблюдений на предмет однородности, т. е. выяснение вопроса о схожести условий экономической активности на протяжении изучаемого временного интервала, а также о неизменности способов оценивания параметров и инструментов (в противном случае требовались бы выделение периодов однородности и отдельное моделирование каждого из подобных периодов). Кроме того, следует учитывать тесную связь протяженности интервала прогнозирования с временным интервалом, использованным при моделировании. Принято считать, что на каждую прогнозную оценку должно приходиться не менее десяти (в некоторых источниках — трех) исходных наблюдений (Лугачев, Ляпунцов, 1999; Лугачев, Ляпунцов, 2002). Исходя из приведенных соображений при решении задачи прогнозирования развития российского фондового рынка в качестве исходного был взят промежуток времени с января 2003 г. по август 2005 г., а в качестве прогнозного — с сентября по ноябрь 2005 г.¹

Для проверки временного ряда на стационарность можно использовать различные методы. В частности, простейшим из них является визуальный контроль графического представления ряда. В этом случае о вероятной нестационарности будут свидетельствовать очевидные на глаз тренд или периодический компонент, а также возрастание или убывание разброса наблюдений во времени. Другим методом является построение графика выборочной автокорреляционной функции (коррелограммы), которая должна быстро убывать после нескольких первых значений в случае стационарного временного ряда (и, наоборот, убывает достаточно медленно при нестационарности ряда). Кроме автокорреляционной, можно строить график и для частной автокорреляционной функции, которая также должна быстро убывать для стационарного процесса². Наконец, возможно использование формальных тестов на наличие единичного корня (простой и расширенный тесты Дики-Фуллера и др.).

Учитывая, что используемый в задаче прогнозирования исходный временной ряд значений объясняемой переменной (индекса фондового рынка РТС) не является стационарным (нестационарность ряда значений индекса РТС очевидна из рис. 1), был использован ряд преобразований, которые приводят к получению модифицированного ряда с характеристиками, приближенными к характеристикам стационарных рядов. Заметим, что рассмотрены не все возможные способы преобразований, а лишь те, которые приводят к формированию новых рядов, демонстрирующих «хорошие» свойства при их использовании в описанных методах прогнозирования.

¹ Значения индекса российского фондового рынка были получены из архива, хранящегося на FTP-сервере РТС (<ftp://ftp.rtsnet.ru/pub/info/stats/>).

² Содержательно частная автокорреляционная функция представляет собой «чистую корреляцию» между текущим и одним из предыдущих уровнями ряда при исключении влияния всех промежуточных лаговых значений.



Рис. 1. График значений индекса российского фондового рынка (индекса РТС) на интервале моделирования

Иными словами, в качестве исходной информации для прогнозирования задействовались уровни следующих временных рядов, полученных из значений индекса РТС:

- Ряд 1 — ряд значений индекса РТС;
- Ряд 2 — ряда остатков индекса РТС после его аппроксимации алгебраическим полиномом на основе метода последовательных разностей [1];
- Ряд 3 — ряда остатков индекса РТС после его аппроксимации алгебраическим полиномом и 30 гармониками ряда Фурье;
- Ряд 4 — ряд разностей значений индекса РТС;
- Ряд 5 — ряда значений «логарифмической прибыли» (логарифмов цепных темпов роста) индекса РТС;
- Ряд 6 — ряд значений «логарифмической прибыли» индекса РТС после вычитания выборочного среднего;
- Ряд 7 — ряд остатков «логарифмической прибыли» индекса РТС после ее аппроксимации алгебраическим полиномом на основе метода последовательных разностей и гармониками ряда Фурье.

В рамках предложенных в данном разделе алгоритмов было выполнено моделирование для всех рассматриваемых рядов. Затем в рамках каждой построенной модели проводилось прогнозирование на 1, 5 и 50 дней, а также в соответствии с процедурой скользящего экзамена. Суть этой процедуры (известной в зарубежной литературе под названием *jackknife method* — «метод складного ножа») может быть описана следующим образом. Предположим, мы сравниваем q построенных для прогнозирования моделей. Принципиальным признаком этого подхода является требование постоянного обновления модели для определения прогнозной оценки каждой постпрогнозной точки. Будем считать, что выбрано $k = 50$ постпрогнозных точек, находящихся в конце временного ряда: y_{n-k+1}, \dots, y_n . В этом случае процедура скользящего экзамена состоит в следующем:

1. По первым $(n-k)$ наблюдениям исходного ряда оцениваются параметры первой модели в спецификации, определенной для прогнозирования $(n+1)$ -й точки.

2. Построенная модель используется для оценки постпрогноза \hat{y}_{n-k+1} .
3. Оцениваются параметры модели по первым $(n-k+1)$ наблюдениям исходного ряда и рассчитывается \hat{y}_{n-k+2} .
4. Расчеты повторяются до определения \hat{y}_n .
5. Оценивается мера качества прогноза

$$\hat{\sigma}^2(l) = \frac{1}{k-p-1} \sum_{i=n-k+1}^n (y_i - \hat{y}_i)^2.$$

6. Процедура повторяется для всех имеющихся q моделей.
7. Последовательность $\hat{\sigma}^2(1), \dots, \hat{\sigma}^2(q)$ представляет собой значения частного критерия точности прогнозирования для всех рассматриваемых моделей.

Наилучший результат показали:

в случае прогнозирования на 1 день — метод прогнозирования, основанный на фильтре Калмана-Бьюси, на основе Ряда 6, и метод прогнозирования, основанный на использовании разложения Карунена-Лоэва, на основе Ряда 3;

в случае прогнозирования на 5 дней — метод прогнозирования, основанный на фильтре Калмана-Бьюси, на основе Ряда 6, и метод прогнозирования, основанный на использовании разложения Карунена-Лоэва, на основе Ряда 4;

в случае прогнозирования на 50 дней — метод прогнозирования, основанный на фильтре Калмана-Бьюси, на основе Ряда 7, и метод прогнозирования, основанный на использовании разложения Карунена-Лоэва, на основе Ряда 4;

в случае прогнозирования в соответствии с процедурой скользящего экзамена — метод прогнозирования, основанный на фильтре Калмана-Бьюси, на основе Ряда 7, и метод прогнозирования, основанный на использовании разложения Карунена-Лоэва, на основе Ряда 3.

Численные значения прогнозов приведены в табл. 2, графическая иллюстрация — на рис. 11, 12.

Прогнозирование индекса РТС с помощью метода анализа сингулярного спектра

Опыт многих естественных наук свидетельствует о возможности представления случайных процессов, описывающих те или иные природные процессы, с помощью функций, состоящих из нескольких слагаемых:

$$y(t) = y_T(t) + y_n(t) + y_r(t) + \varepsilon(t),$$

где $y_T(t)$ — медленная нерегулярная составляющая, обычно называемая трендом, $y_n(t)$ — периодическая (или сумма периодических) составляющая (в зависимости от области приложений они называются сезонными, суточными и т. п. вариациями), $y_r(t)$ — быстрые нерегулярные малые вариации, в которые обычно включают все, что не укладывается в формальную модель, $\varepsilon(t)$ — чисто случайная составляющая, описываемая случайным процессом определенного типа.

Во многих частных случаях к настоящему времени созданы мощные теории с развитым аппаратом приложений и компьютерными реализациями в виде библиотек и пакетов программ. Так, для функций вида $y(t) = y_T(t) + \varepsilon(t)$ такой теорией является теория аппроксимации или метод наименьших квадратов, для функций вида $y(t) = y_n(t)$ хорошие результаты показывает теория гармонических рядов Фурье. Однако во многих других ситуациях возникают достаточно большие сложности эффективного исследования функций. В различных конкретных приложениях предложено много эвристических приемов, но они, как правило, имеют слабую теоретическую обоснованность. Поэтому и в настоящее время не прекращаются попытки разработки новых методов анализа временных рядов.

Одним из таких методов является *метод анализа сингулярного спектра (SSA)*¹ (Голяндина, 2004(а); Голяндина, 2004(б)). Его теоретическими основами являются, с одной стороны, образующий ядро факторного анализа классический анализ главных компонент, а с другой — быстро развивающаяся в последнее время теория динамических систем.

Процедура исследования одномерного временного ряда длины n начинается с преобразования его в многомерный. Задав шаг лагом $m < n/2$ (параметр m может выбираться в широких пределах — при достаточно больших значениях m и n результаты не будут зависеть от величины лага), значениями исходного ряда последовательно заполняют строки так называемой траекторной матрицы X . При этом первая строка содержит первые m элементов ряда, вторая — со второго элемента по $m+1$ и так далее, пока ряд не исчерпается. После центрирования по столбцам и соответствующей нормировки вычисляется корреляционная матрица $R = XX^T$, сингулярное разложение которой $R = PLP^T$ дает диагональную матрицу собственных чисел L и ортогональную матрицу собственных векторов P матрицы R^2 . В программной реализации этого метода главные компоненты исходной матрицы $Y = XP$ могут быть исследованы и, что особенно важно при дальнейшем восстановлении по ним исходного ряда, визуализированы и упорядочены по возрастанию их вклада в исходный ряд. Это позволяет интерактивно производить непосредственный поиск гармонических компонентов, фильтрацию или сглаживание ряда, выбирая соответствующие значимые компоненты Y_i . Ввиду ортогональности матрицы P можно восстановить матрицу $X = YP^T$, используя при этом выбранные главные компоненты Y_i .

При этом часто оказывается возможным выделить отдельные слагаемые исходного ряда, такие, как медленный тренд общего вида, медленные сезонные составляющие, периодические составляющие (если они есть) и случайные вариации. Это, в свою очередь, позволяет прогнозировать как сам временной ряд, так и тенденции развития различных его составляющих. Метод допускает естественное обобщение на случай анализа многомерных временных рядов.

Иными словами, ключевой особенностью метода SSA является сингулярное разложение траекторной матрицы в сумму элементарных матриц. Каждая элементарная матрица задается тройкой — набором из собственного числа и двух сингулярных векторов (собственного и факторного). В предположении, что исходный временной ряд является суммой нескольких рядов, теоретические результаты позволяют при некоторых условиях определить по виду собственных чисел, собственных и факторных векторов, что это за слагаемые и какой набор элементарных матриц соответствует каждому из них. Суммируя элементарные матрицы внутри каждого набора и переходя затем от результирующих матриц к ряду, мы получаем разложение ряда на аддитивные слагаемые (например, на сумму тренда, периодики и шума или на сумму низкочастотной и высокочастотной составляющих).

Таким образом, целью метода является разложение временного ряда на интерпретируемые аддитивные составляющие. При этом метод не требует стационарности ряда, знания модели тренда, а также сведений о наличии в ряде периодических составляющих и их периодах.

¹ В русскоязычной литературе этот метод часто называется методом «Гусеница».

² В терминологии теории фильтрации собственные векторы корреляционной матрицы выступают в роли передаточных функций соответствующих фильтров. Ширина полосы пропускания зависит от формы передаточной функции фильтра и определяется как видом собственного вектора, так и длиной интервала усреднения, т. е. лагом m . Чем больше m , тем уже полоса фильтра. Выбор нескольких главных компонент эквивалентен параллельному соединению соответствующих фильтров, что позволяет управлять формой спектральной характеристики. При выборе m , которое значительно меньше характерной ритмичности ряда, фактически происходит его сглаживание. Периодические, но не обязательно гармонические составляющие исследуемого ряда образуют графически хорошо различимую пару соседних компонентов Y (типа фигур Лиссажу).

Данный метод имеет определенные аналогии с вэйвлет-анализом и динамическим Фурье-анализом (если ряд состоит из набора строго гармонических компонентов, то фактически осуществляется разложение в ряд Фурье)¹. Наиболее важные преимущества метода для проводимого исследования состоят в том, что:

— базисные функции метода не фиксированы, а порождаются самим исследуемым рядом, так как являются собственными векторами R ;

— имеется возможность восстановления ряда по информативным компонентам, исследование которых доступно интерактивно.

Метод SSA потенциально обладает хорошими перспективами применительно к решению задачи прогнозирования, которое, в отличие от анализа, возможно только при наличии математической модели временного ряда. Модель должна либо порождаться самими данными, либо, по крайней мере, тестироваться на имеющихся данных.

В работах (Главные компоненты временных рядов, 1997; Голяндина, 2004(б)) содержатся доведенные до алгоритмов прогноза временных рядов результаты, связывающие траекторную матрицу с линейными рекуррентными уравнениями. При прогнозировании с помощью метода SSA рассматривается множество временных рядов, описываемых с помощью линейных рекуррентных формул (ЛРФ). При этом порядок ЛРФ может быть, вообще говоря, заранее неизвестен, поэтому класс рядов, управляемых такими ЛРФ, достаточно широк и важен для практических применений.

Ряд, управляемый ЛРФ, естественным образом порождает рекуррентное продолжение, так как каждый его член равен линейной комбинации некоторого количества предыдущих. Поэтому коэффициенты этой линейной формулы (если они известны) могут быть использованы и для продолжения временного ряда. Важно отметить, что необязательно искать ЛРФ минимальной размерности, так как любая ЛРФ, управляющая рядом, приводит к одному и тому же продолжению. Тем самым главной задачей является нахождение порядка и коэффициентов некоторой ЛРФ, управляющей рядом.

Общая идея нахождения ЛРФ состоит в следующем. Базовый метод SSA порождает естественный базис траекторного пространства, на основе которого конструктивно строится управляющая ЛРФ. Если мы применим эту ЛРФ к последним точкам ряда, то получим продолжение ряда.

Та же самая идея используется и в случае продолжения аддитивного компонента f_1 ряда $f = f_1 + f_2$. Здесь мы предполагаем, что компонент f_1 управляется некоторой ЛРФ. Тогда он может быть выделен с помощью метода SSA и продолжен по соответствующей линейной рекуррентной формуле. Предположим, что рассматриваемый ряд f может быть представлен в виде суммы допускающего рекуррентное продолжение ряда f_1 и остаточного ряда f_2 . Если второй ряд является шумом, то мы приходим к стандартной задаче прогнозирования сигнала f_1 при наличии шума f_2 (также можно ставить задачу прогнозирования ряда f_1 , если это тренд или сезонный компонент ряда f). В реальных задачах условия, при которых ряд или компонент ряда допускает точное продолжение, как правило, не выполняются. В этом случае главным предположением для возможности прогнозирования является предположение о том, что при некоторой длине ряда компоненты ряда f_1 и f_2 разделимы с помощью метода SSA. Тогда с помощью SSA мы можем приближенно восстановить ряд f_1 и получить тем самым аппроксимацию как самого ряда f_1 , так и его траекторного пространства. Други-

¹ Учитывая тригонометрическую структуру базисных функций в Фурье-анализе, очевидна периодичность любой функции, разлагаемой в ряд Фурье. Для того чтобы расширить соответствующую теорию на случай непериодических функций, переходят к вэйвлет-анализу, семейство функций ортогонального базиса в котором задано на всей числовой прямой и имеет непериодическую структуру.

ми словами, мы получаем сразу и ЛРФ, приближенно управляющую рядом f_1 , и начальные данные для этой формулы. Следовательно, мы имеем все, чтобы построить приближенное рекуррентное продолжение ряда f_1 , т. е. прогноз.

Воспользуемся методикой, предложенной в рамках метода SSA, применительно к задаче анализа и прогнозирования развития временного ряда, составленного из значений индекса РТС. Для примера описывается применение метода SSA в случае выбора в качестве исходных данных значений Ряда 1.

Визуальный анализ Ряда 1 (см. рис. 1) свидетельствует о наличии тренда, который должен описываться полиномиальной или экспоненциальной функцией, а также периодическое поведение довольно сложно меняющейся формы.

На первом этапе будем решать задачу поиска разложения данного ряда на три компонента — тренд, периодический компонент и шум. Поскольку для реальных (зашумленных) данных точная разделимость практически исключена, нам необходимо воспользоваться результатами о приближенной (асимптотической) разделимости медленно меняющегося тренда и гармоник. Для достижения лучшей точности мы должны выбрать длину лага большой, близкой к половине длины ряда. Выберем $n = 330$ ¹.

Рассмотрим теперь результат сингулярного разложения траекторной матрицы ряда при таком выборе длины лага. На рис. 2 приведено изображение собственных векторов из первых девяти собственных троек сингулярного разложения².

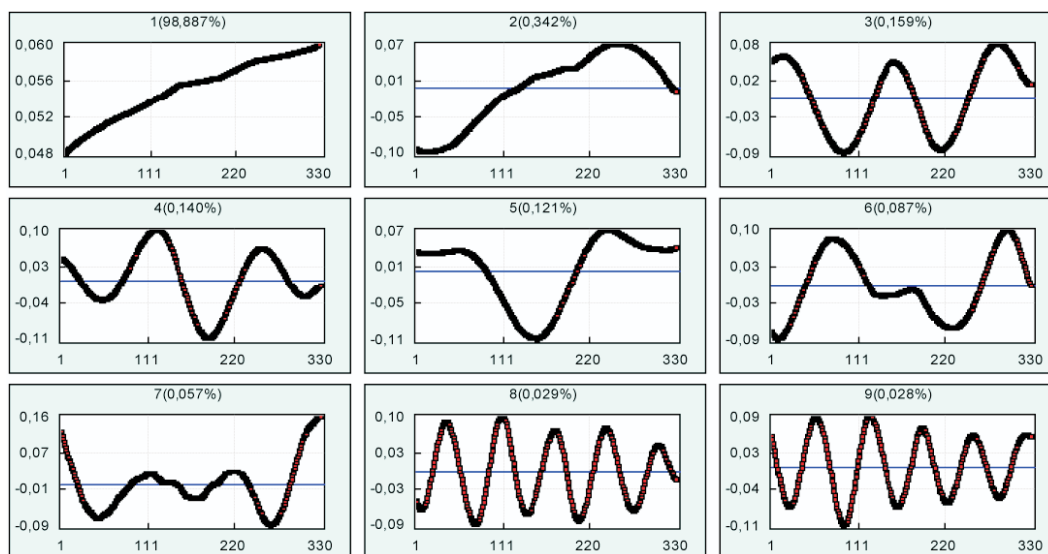


Рис. 2. Одномерные диаграммы собственных векторов

Для идентификации собственных троек воспользуемся результатами о виде собственных векторов, соответствующих тренду и гармоникам при условии их приближенной разделимости (Golyandina, Nekrutkin, Zhigljavsky, 2001).

Начнем с идентификации тренда. Известно, что сингулярные (в частности собственные) векторы имеют в целом такой же вид, как и компонент исходного ряда, которому они соответствуют. Поэтому на одномерных диаграммах собст-

¹ Рассматривались и другие варианты выбора длины лага. Указанный случай обеспечил получение наибольшей точности прогнозирования.

² Заметим, что вид факторных векторов, или правых сингулярных векторов (с четными номерами), мало отличается от вида левых, так как длина лага близка к половине длины ряда. Если бы мы взяли длину лага меньше, то собственные векторы имели бы более регулярный вид по сравнению с факторными, которые содержали бы в себе, например, изменение амплитуд для гармонических компонентов.

венных векторов нужно найти медленно меняющиеся собственные векторы. В данном случае, как представляется, два, а именно первый и второй собственные векторы имеют требуемый вид. Так как получилось, что в данном случае тренд не описывается единственной собственной тройкой, то это означает, что тренд аппроксимируется более сложной кривой, нежели обычная экспонента¹.

Перейдем теперь к идентификации гармонических (возможно, с меняющейся амплитудой) компонентов, порожденных периодическим компонентом исходного ряда. В рассматриваемом ряде выделяется некоторое число самостоятельных (не являющихся гармониками других) периодических компонентов. На рис. 2 видно, что собственные тройки с номерами 3—7, возможно, соответствуют каким-либо гармоникам, так как имеют регулярное периодическое поведение. Однако поскольку гармоника с периодом, большим, чем 2, порождает две собственные тройки (косинус с периодом 2 порождает только одну собственную тройку, сингулярные векторы которой имеют пилообразный вид, а таких на одномерных диаграммах в данном случае не обнаружено), то искать относящиеся к гармоникам пары собственных троек удобнее на двумерных диаграммах². Так как при достаточной большой длине ряда соответствующая пара собственных чисел имеет близкие значения, то достаточно рассматривать двумерные диаграммы собственных векторов из соседних, упорядоченных по собственным значениям, собственных троек. На рис. 3 можно различить регулярные двумерные изображения, образующие двумерные траектории с вершинами, лежащими на кривой, имеющей спиралеобразную форму. Это означает, что соответствующая пара собственных векторов порождена модулированным гармоническим компонентом исходного ряда. Таким образом, получаем, что определенным значениям периода соответствуют пары собственных троек (используем аббревиатуру СТ) СТ3, СТ4 и СТ6, СТ7.

Следует также заметить, что обнаружение высокочастотных составляющих может быть результатом дискретизации данных, а также иметь причиной шум. Поэтому, если для низкочастотных составляющих обычно решается вопрос о том, что это — тренд или периодика, то для высокочастотных составляющих должен ставиться вопрос об их случайности. Например, о случайности может говорить биение амплитуды периодики (см. рис. 2, собственные вектора 8 и 9). В то же время увеличение/уменьшение амплитуды собственных векторов 4, 6 и 7 свидетельствует только о нестационарности амплитуды периодики самого ряда.

Дополнительным свидетельством в пользу правильности идентификации собственных троек выступает рис. 4, на котором изображены логарифмы собственных значений (каждой паре соответствует «ступенька»; начиная с собственного значения 8, в наклоне наблюдается излом, после которого график становится гораздо более пологим, чем в случае собственных значений с начальными номерами)³.

¹ Чем сложнее форма тренда, тем больше его (приближенная) размерность и тем большее число собственных троек ему соответствует.

² Двумерные графики собственных векторов служат для облегчения разбиения главных компонентов по парам. Так как двумерное изображение синуса и косинуса образует единичную окружность, то гармоническая составляющая с целым периодом изображается в виде правильного многоугольника с числом вершин, равным величине периода. При изменении амплитуды многоугольник превращается в спираль. Звездообразный график свидетельствует о наличии периодики с дробным периодом. Заметим, что в поиске составляющей с периодом 2 двумерные графики помочь не могут.

³ В этом случае применимы стандартные критерии метода главных компонент. Кроме того, метод SSA предполагает и другие критерии. Так как одной синусоидальной составляющей соответствует в идеальной ситуации два главных компонента с одинаковым периодом (синус и косинус), соответствующие одному собственному значению, то в реальной ситуации эти два главных компонента соответствуют близким собственным значениям. Поэтому на графике можно увидеть «ступеньку».

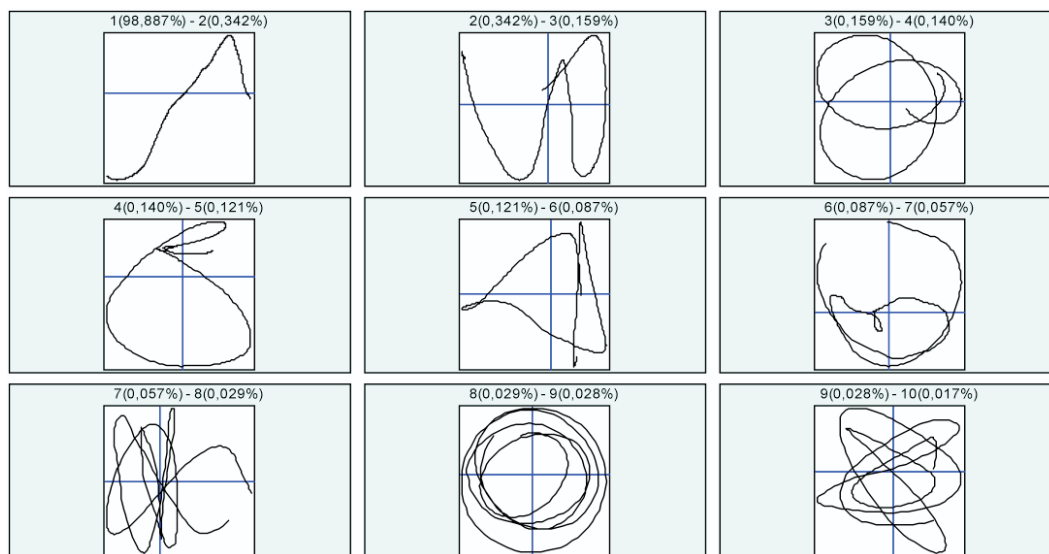


Рис. 3. Двумерные диаграммы собственных векторов

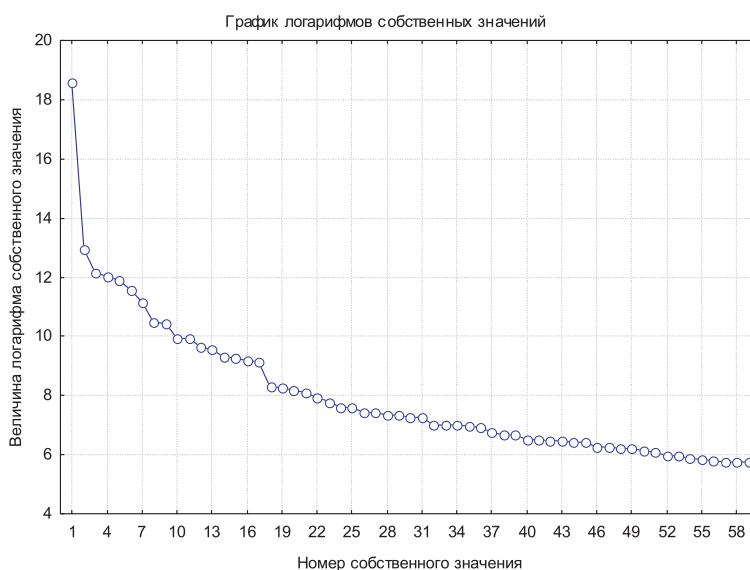


Рис. 4. Логарифмы первых 60 собственных значений

На рис. 5 изображены первые 9 формально восстановленных, по каждой собственной тройке отдельно, компонентов исходного ряда (так называемых элементарных компонентов).

Отдельно следует остановиться на вопросе отделения компонентов, относящихся к сигналу, от шумовых компонентов. Во-первых, нерегулярное поведение сингулярных векторов может говорить о принадлежности их к набору, порожденному шумовым компонентом. Во-вторых, медленное, почти без скачков, убывание собственных значений с некоторого номера также говорит об этом. В-третьих, большой набор собственных троек, порождающих коррелирующие друг с другом восстановленные компоненты, скорее всего, относится к шуму. Рис. 6 показывает, что собственные тройки с номерами, большими 7, как раз образуют такой блок¹.

¹ Для изображения представленной на рисунке корреляционной матрицы предлагается визуальный способ, в котором большие значения обозначаются более темным цветом.

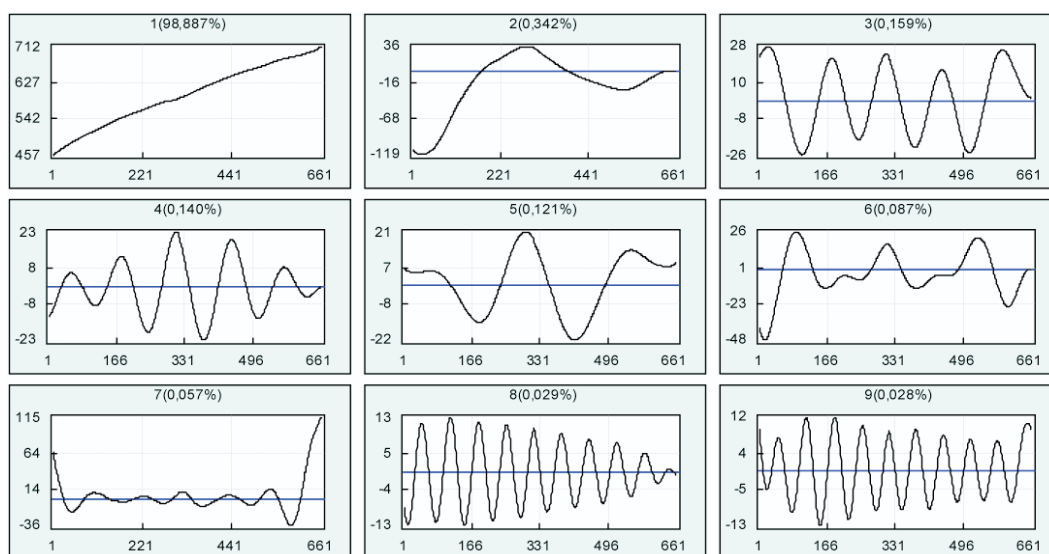


Рис. 5. Элементарные восстановленные компоненты ряда

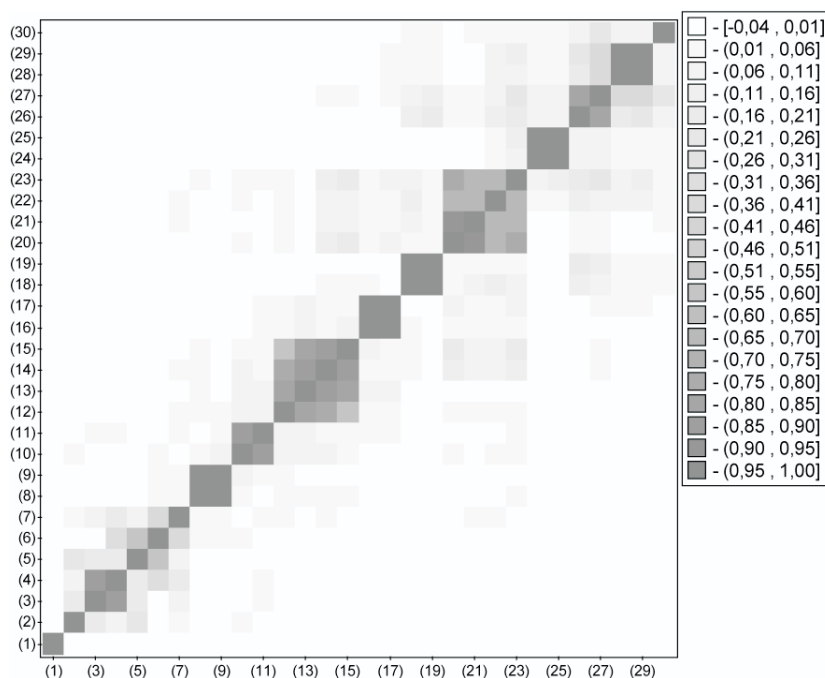


Рис. 6. Матрица корреляций элементарных восстановленных компонентов ряда

Последним подтверждением правильности разделения (естественно, приближенного) сигнала и шума является проверка на принадлежность к шуму компонента ряда, полученного с помощью восстановления по, так сказать, «шумовым» собственным тройкам, стандартными статистическими методами. На рис. 7 изображено разложение исходного ряда на три компонента — тренд (СТ1 — СТ2), периодику (СТ3 — СТ7) и шум (СТ8 — СТ330). Статистические критерии не отвергают того, что третий компонент является реализацией «белого шума».

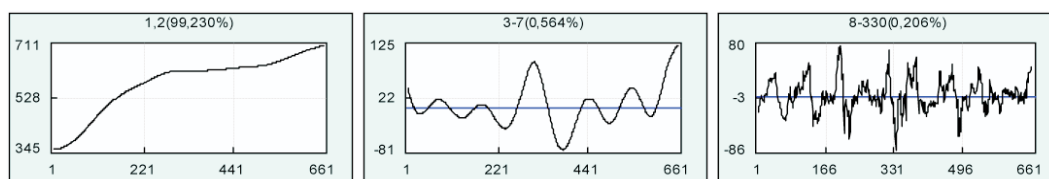


Рис. 7. Разложение ряда на тренд (левый рисунок), периодический компонент (средний рисунок) и шум (правый рисунок)

Таким образом, рис. 7, полученный на основе группировки элементов сингулярного разложения, дает решение поставленной задачи построения разложения исходного ряда на интерпретируемые аддитивные компоненты¹.

После завершения этапа анализа мы можем перейти к решению задачи построения прогноза. Так как нам удалось приближенно отделить сигнал от шумового компонента, и при этом сигналу соответствовало 7 собственных троек, то его траекторное подпространство должно хорошо аппроксимироваться подпространством размерности 7. В свою очередь, подпространство порождает линейную рекуррентную формулу размерности $329 = n - 1$, такую, что восстановленный ряд (в данном случае — сигнал) будет аппроксимироваться рядом, управляемым этой ЛРФ (Голяндина, 2004(а)).

В табл. 1 представлена информация о первых 10 из 329 корней характеристического полинома данной ЛРФ. Корни упорядочены по убыванию их модуля². Первые четыре строки таблицы можно легко проинтерпретировать: строки со 2 по 4 соответствуют комплексно-сопряженным корням, порождающим гармоники с периодами, примерно равными 120, 170 и 385. Первая строка таблицы соответствует вещественному корню с модулем 1,0007. Так как число главных корней должно быть равно как раз семи, то больше главных корней нет. Остальные корни являются побочными, по модулю меньшими единицы. Все корни изображены на рис. 8 на комплексной плоскости в координатах («вещественная часть», «мнимая часть»).

Для проверки качества описания ряда порожденной восстановленным подпространством ЛРФ (будем называть такие формулы полными ЛРФ) можно оценить максимальную ошибку глобальной аппроксимации. Она равняется 80, что составляет порядка 10% от значений ряда.

Специфической особенностью рассматриваемого метода анализа сингулярного спектра при решении задачи прогнозирования значений индекса РТС явилась найденная возможность улучшения качества прогноза в случае разбиения элементарных компонентов на периодическую и шумовую составляющие иным, отличным от построенного выше на основе рекомендаций теории, способом. При этом, однако, несколько возросла максимальная ошибка глобальной аппроксимации, которая составила 141, или 16% от значений ряда.

¹ Неудивительно, что исходный и восстановленный ряд практически совпадают, так как вклад среднего и первых семи главных компонентов составляет 99,79%.

² Если в столбце «Тип» таблицы написано «компл.», то это означает, что строка содержит информацию о двух комплексно-сопряженных корнях и указана величина абсолютного значения мнимой части корня.

Пара корней с одинаковой ненулевой вещественной частью и противоположными по знаку мнимыми частями соответствует синусоидальной составляющей ряда с амплитудой, меняющейся по экспоненте. Причем если модуль корня больше единицы, то амплитуда возрастает, а если меньше единицы, то она убывает. Модуль, равный единице, показывает постоянство амплитуды. Период синуса равен 2π , деленному на величину угла между осью абсцисс и лучом, проходящим через точку корня. Таким образом, есть возможность определять периоды по корням характеристического полинома.

Вещественные положительные корни соответствуют тренду. Корень кратности два, лежащий на единичной окружности, соответствует линейному тренду. Корень, лежащий вне единичного круга, соответствует возрастающей экспоненте, внутри — убывающей.

Таблица 1

**Свойства корней характеристического полинома полной линейной рекуррентной формулы
Ряда 1**

Номер	Re	Im	Модуль	Частота	Период	Тип
1	1,0007	0,0000	1,0007	0,0000	Нет	веществ.
2	0,9992	0,0534	1,0007	0,0534	117,7518	компл.
3	0,9981	0,0367	0,9987	0,0368	170,9639	компл.
4	0,9976	0,0162	0,9978	0,0163	386,6372	компл.
5	0,9932	0,0760	0,9962	0,0764	82,2709	компл.
6	0,9910	0,0973	0,9958	0,0979	64,2085	компл.
7	0,9884	0,1166	0,9953	0,1174	53,5172	компл.
8	0,9859	0,1358	0,9952	0,1369	45,8922	компл.
9	0,9792	0,1733	0,9944	0,1752	35,8713	компл.
10	0,9820	0,1544	0,9941	0,1560	40,2796	компл.

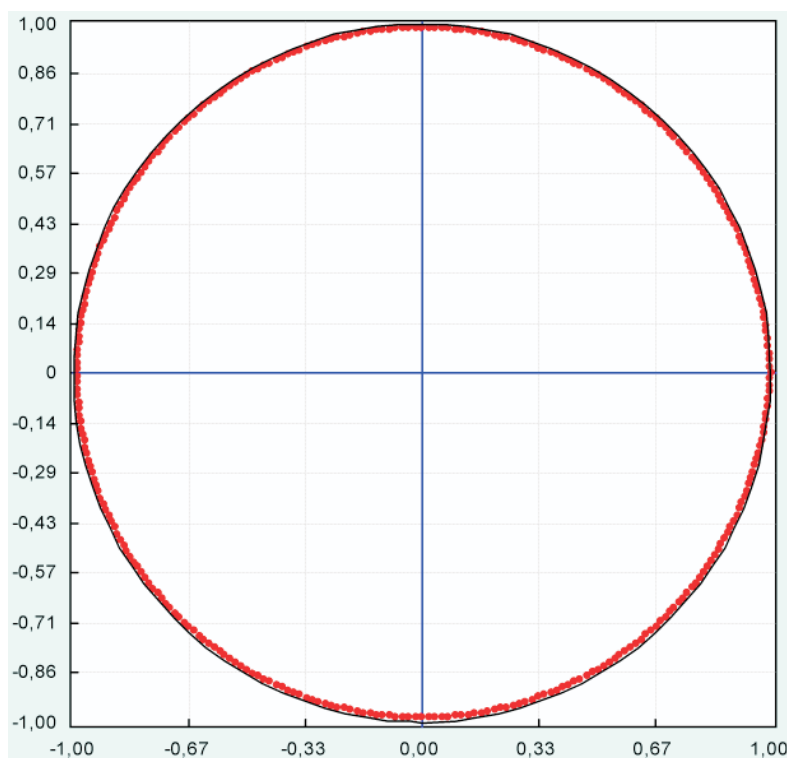


Рис. 8. Главные и побочные корни характеристического полинома полной линейной рекуррентной формулы Ряда 1

Для модификации состава компонентов разложения использовалась матрица корреляций для всех элементарных восстановленных компонентов ряда, представленная на рис. 9.

С помощью визуального анализа матрицы корреляций было выбрано новое разложение исходного ряда на три компонента — тренд (СТ1 — СТ2), периодичку (СТ3 — СТ165) и шум (СТ166 — СТ330). Статистические критерии по-прежнему не отвергают того, что третий компонент является реализацией «белого шума».

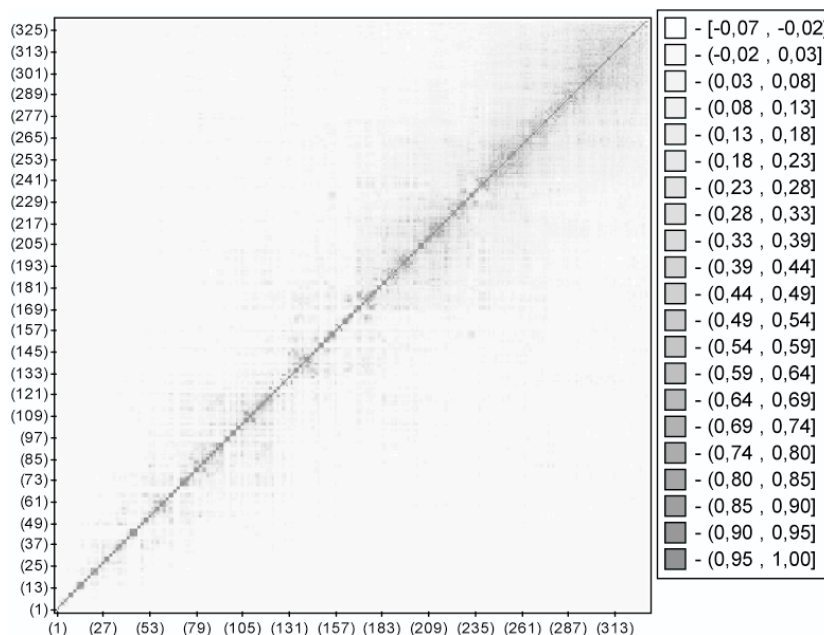


Рис. 9. Матрица корреляций всех элементарных восстановленных компонентов ряда

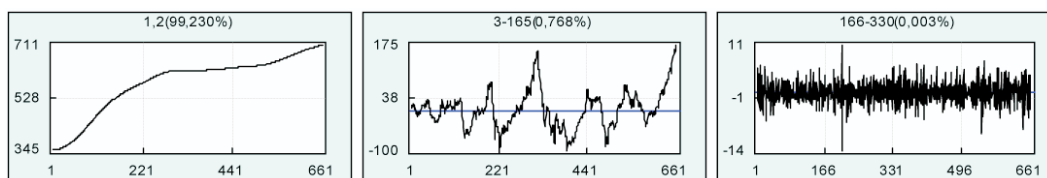


Рис. 10. Разложение ряда на тренд (левый рисунок), периодический компонент (средний рисунок) и шум (правый рисунок)

В рамках модели полной ЛРФ, построенной по подпространству, порожденному первыми, не относящимися к шумовому компоненту собственными тройками, проводилось прогнозирование на 1, 5 и 50 дней, а также применялась процедура скользящего экзамена. Аналогичным образом при проведении исследования было выполнено прогнозирование для всех рассматриваемых рядов.

Наилучший результат показали:

в случае прогнозирования на 1 день — метод анализа сингулярного спектра на основе Ряда 3;

в случае прогнозирования на 5 дней — метод анализа сингулярного спектра на основе Ряда 5;

в случае прогнозирования на 50 дней — метод анализа сингулярного спектра на основе Ряда 7;

в случае прогнозирования в соответствии с процедурой скользящего экзамена — метод анализа сингулярного спектра на основе Ряда 7.

Численные значения прогнозов приведены в табл. 2, графическая иллюстрация — на рис. 11, 12.



Рис. 11. Прогнозирование индекса РТС на 50 дней



Рис. 12. Прогнозирование индекса РТС на 50 дней с помощью процедуры скользящего экзамена

Таблица 2

Результаты прогнозирования индекса РТС на 50 дней

Дата	Индекс РТС	Прогноз методом фильтра Калма- на-Бьюси на ос- нове Ряда 6	Прогноз методом фильтра Калма- на-Бьюси на ос- нове Ряда 7	Прогноз методом фильтра Калма- на-Бьюси с по- мощью процеду- ры скользящего экзамена	Прогноз методом разложения Ка- рунана-Лозва на основе Ряда 4	Прогноз методом разложения Ка- рунана-Лозва с помощью про- цедуры скользя- щего экзамена	Прогноз методом анализа спектра на основе Ряда 5	Прогноз методом анализа спектра с помощью про- цедуры скользя- щего экзамена
01.09.05	894,6	883,2	881,4	880,3	882,1	883,8	875,2	878,3
02.09.05	909,9	884,4	883,6	900,6	882,1	886,8	881,4	883,0
05.09.05	910,1	885,6	888,1	915,2	882,1	890,8	890,9	896,2
06.09.05	882,6	886,8	889,9	905,5	882,1	893,4	895,1	903,0
07.09.05	866,6	888,0	888,7	875,1	882,1	894,7	890,5	898,1
08.09.05	883,1	889,2	887,2	865,4	882,1	896,6	886,9	896,7
09.09.05	905,9	890,4	889,7	889,1	882,1	901,7	894,2	905,5
12.09.05	903,9	891,6	893,8	908,4	882,1	906,8	893,0	910,2
13.09.05	900,2	892,8	896,2	901,4	882,1	910,2	889,9	910,3
14.09.05	896,1	894,0	894,6	898,6	882,1	912,8	898,7	917,9
15.09.05	906,9	895,3	893,5	898,6	882,1	916,1	900,8	921,0
16.09.05	923,0	896,5	895,4	910,0	882,1	920,5	900,5	917,8
19.09.05	946,3	897,7	900,0	926,6	882,1	926,8	893,1	914,7
20.09.05	945,0	898,9	902,1	950,6	882,1	934,5	888,1	913,9
21.09.05	960,5	900,1	901,1	948,2	882,1	942,1	884,1	912,0
22.09.05	966,0	901,3	899,4	962,5	882,1	952,2	881,5	910,2
23.09.05	965,3	902,6	901,6	968,2	882,1	963,1	884,0	911,6
26.09.05	954,8	903,8	905,8	970,1	882,1	973,8	888,9	919,0
27.09.05	984,5	905,0	908,5	960,8	882,1	984,0	892,3	925,2
28.09.05	988,3	906,2	907,1	987,9	882,1	996,0	896,4	930,5
29.09.05	991,1	907,5	905,8	985,8	882,1	1004,2	894,7	930,2
30.09.05	1007,8	908,7	907,4	991,9	882,1	1009,1	901,1	934,0
03.10.05	1038,8	909,9	912,1	1017,2	882,1	1012,1	908,2	943,0
04.10.05	1045,0	911,2	914,5	1045,7	882,1	1012,1	914,2	952,0
05.10.05	1009,6	912,4	913,6	1040,6	882,1	1007,2	913,4	954,0
06.10.05	954,6	913,6	911,8	999,3	882,1	995,5	908,4	950,5
07.10.05	949,1	914,9	913,8	951,9	882,1	980,6	910,4	952,6
10.10.05	970,3	916,1	917,9	954,0	882,1	968,1	917,6	961,8

Продолжение табл. 2

Дата	Индекс РТС	Прогноз методом фильтра Калма- на-Бьюси на ос- нове Ряда 6	Прогноз методом фильтра Калма- на-Бьюси на ос- нове Ряда 7	Прогноз методом фильтра Калма- на-Бьюси с по- мощью процеду- ры скользящего экзамена	Прогноз методом разложения Ка- рунена-Лозва на основе Ряда 4	Прогноз методом разложения Ка- рунена-Лозва с помощью про- цедуры скользя- щего экзамена	Прогноз методом анализа спектра на основе Ряда 5	Прогноз методом анализа спектра с помощью про- цедуры скользя- щего экзамена	
11.10.05	994,9	917,3	921,0	970,7	882,1	958,0	916,7	962,1	969,4
12.10.05	966,9	918,6	919,7	986,1	882,1	948,3	909,4	953,7	984,3
13.10.05	935,2	919,8	918,3	958,1	882,1	936,0	901,4	945,9	957,6
14.10.05	911,2	921,1	919,6	933,8	882,1	924,6	904,6	948,4	934,7
17.10.05	915,4	922,3	924,3	911,2	882,1	915,9	902,7	948,9	912,0
18.10.05	918,2	923,6	927,0	909,3	882,1	910,7	897,9	945,1	906,3
19.10.05	882,1	924,8	926,4	909,7	882,1	907,3	892,3	940,6	907,3
20.10.05	887,4	926,1	924,3	875,4	882,1	901,5	883,7	931,9	876,0
21.10.05	887,9	927,3	926,0	889,2	882,1	900,9	878,9	926,8	885,9
24.10.05	908,6	928,6	930,2	886,4	882,1	900,6	880,3	929,9	886,0
25.10.05	904,4	929,8	933,5	905,8	882,1	903,8	876,7	927,7	904,1
26.10.05	910,4	931,1	932,5	899,9	882,1	905,9	871,9	921,9	899,8
27.10.05	898,5	932,4	931,0	910,6	882,1	911,0	871,9	922,8	912,2
28.10.05	910,8	933,6	932,0	897,1	882,1	915,8	874,3	927,1	900,6
31.10.05	935,0	934,9	936,7	911,7	882,1	924,6	870,3	926,1	912,8
01.11.05	942,1	936,2	939,6	937,4	882,1	935,1	870,7	927,6	939,3
02.11.05	964,4	937,4	939,3	944,4	882,1	945,6	875,7	933,2	946,5
03.11.05	972,4	938,7	937,1	965,1	882,1	957,1	877,6	935,4	964,6
07.11.05	971,6	940,0	938,5	972,6	882,1	967,6	876,5	935,8	973,7
08.11.05	971,2	941,2	942,6	974,7	882,1	975,8	878,8	939,3	974,2
09.11.05	980,3	942,5	946,3	977,8	882,1	983,2	880,0	942,1	975,9
10.11.05	972,2	943,8	945,5	982,2	882,1	989,1	885,0	945,8	980,5
Сумма квадратов ошибок за 50 дней		124494	121452	12219	257487	12986	198540	72838	12742

Результаты прогнозирования индекса РТС

Таким образом, можно говорить о том, что рассмотренные методы построения прогнозов динамики российского фондового рынка сопоставимы по своей эффективности с точки зрения выбранного критерия сравнения. Некоторое предпочтение может быть отдано системам прогнозирования на базе метода, основанного на фильтре Калмана-Бьюси (при прогнозировании с помощью процедуры скользящего экзамена), а также метода анализа сингулярного спектра (для случая единовременного прогнозирования).

Источники

- Айвазян С. А., Мхитарян В. С.* Прикладная статистика и основы эконометрики. М., 1998.
- Браммер К., Зиффлинг Г.* Фильтр Калмана-Бьюси. М., 1982.
- Вазан М.* Стохастическая аппроксимация. М., 1972.
- Главные компоненты временных рядов: метод «Гусеница» / под ред. Д. Л. Данилова, А. А. Жиглявского. СПб., 1997.
- Голяндина Н. Э.* Метод «Гусеница»—SSA: анализ временных рядов. СПб., 2004(а).
- Голяндина Н. Э.* Метод «Гусеница»—SSA: прогноз временных рядов. СПб., 2004(б).
- Граничин О. Н.* Введение в методы стохастической оптимизации и оценивания. СПб., 2003.
- Давенпорт В. Б., Рут В. Л.* Введение в теорию случайных сигналов и шумов / пер. с англ. под ред. Р. Л. Добрушина. М., 1960.
- Калман Р., Бьюси Р.* Новые результаты в линейной фильтрации и теории предсказания // Техническая механика. Сер. Д. 1961. Т. 83. № 1.
- Лукачев М. И., Ляпунов Ю. П.* Методы социально-экономического прогнозирования. М., 1999.
- Стратонович А. Л.* Условные марковские процессы и их применение к теории оптимального управления. М., 1966.
- Стренг Г., Фикс Дж.* Теория метода конечных элементов / пер. с англ. под ред. Г.И. Марчука. М., 1977.
- Теория статистики / под ред. Г. Л. Леонтьевой. М., 2002.
- Фомин В. Н.* Рекуррентное оценивание и адаптивная фильтрация. М., 1984.
- Цыпкин Я. З.* Адаптация и обучение в автоматических системах. М., 1968.
- Fisher R. A.* The design of experiments. Edinburgh, 1935.
- Golyandina N., Nekrutkin V., Zhigljavsky A.* Analysis of Time Series Structure: SSA and Related Techniques. London, 2001.
- Kiefer J., Wolfowitz J.* Statistical estimation on the maximum of a regression function // Ann. Math. Statist. 1952. Vol. 23. P. 462—466.
- Kushner H. J., Yin G. G.* Stochastic approximation algorithms and applications. New York, 1997.
- Robbins H., Monro S.* A stochastic approximation method // Ann. Math. Statist. 1951. Vol. 22. P. 400—407.