
Подружко А.А., Подружко А.С.

Институт системного анализа РАН, г. Москва

ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРОВ И ХАРАКТЕРИСТИК ТРЕНДОВ

В работе представлен метод интервальной оценки параметров и характеристик трендов, основанный на вариационных принципах. Рассмотрены типовые модели трендов (линейного, параболического, экспоненциального и парного), получены объективно надежные оценки их общих характеристик на основе интервальных образов регрессий.

Введение

Задачи оценки параметров трендов относятся к категории классических в статистике. Их решению посвящено большое число публикаций как научного, так и прикладного характера [1,2,3]. Методической базой таких решений, как правило, служит хорошо известный метод наименьших квадратов (м.н.к.), во многих случаях приводящий к удовлетворительным результатам. Однако общие свойства получаемых при этом решений и оценок далеко не всегда отвечают предъявляемым требованиям, в частности, по показателям надежности [4,5]. Особенно это проявляется на коротких рядах статистики, при анализе фрагментов квазистационарных процессов. Подобного рода осложнения возникают в случае принятия одномоментных решений на основе прогноза, в задачах ситуационного управления процессами с запаздывающей реакцией, планировании с учетом фактора риска и др. Основная причина возникающих затруднений кроется в самой схеме получения оценок, предполагающей усреднение на всем объеме статистики и использование вероятностных показателей качества результатов.

В настоящей работе рассматривается принципиально иной вариационный подход к оценке параметров и характеристик трендов на основе интервального образа регрессии. Получаемые при этом результаты позволяют более корректно подходить к формированию надежных решений с учетом особенностей моделируемых ситуаций и актуальных целей управления.

1. Параметры и общие характеристики трендов

Тренд обычно определяют как «закономерность, характеризующая общую долгосрочную тенденцию изменений временного ряда» [3]. Это не только направление и скорость развития процесса, а так же его экстремальные и асимптотические особенности, минимумы или максимумы, устойчивые или предельные состояния и др.

Для описания трендов могут привлекаться разнообразные функции, отвечающие общим представлениям о наблюдаемом процессе и решаемым задачам: линейные и нелинейные, однозначные и многозначные. Идентификация моделей обычно сводится к оценке их параметров на основе заданной статистики с использованием м.н.к. процедур для линейных функций и методов математического программирования в случае нелинейных. Некоторые из этих параметров имеют вполне понятную содержательную интерпретацию. Однако относить их к категории характеристик тренда следует с осторожностью и обязательно с учетом точности полученного приближения модели в целом. Тренд, как «долгосрочная тенденция», заведомо предполагает стационарность и надежность оценок показателей развития процесса. Эти свойства во многом определяют состоятельность характеристик трендов и возможности их практического использования. Основная трудность состоит в выборе способа описания этих свойств, формировании показателей для проверки условий состоятельности и сопоставления результатов моделирования. К числу общих характеристик трендов в первую очередь следует отнести:

- точку рождения тренда - задающую наиболее правдоподобное положение начало отсчета моделируемого тренда;
- признаки стационарности модели тренда позволяющие объективно и, однозначно отделить стационарную ситуацию от нестационарной;
- горизонт прогноза тренда ограничения по состоятельности прогноза во времени;
- точку схождения – пересечение границ парной регрессии и др.

Практический интерес к таким показателям очевиден. Решение подобных задач известными методами представляется затруднительным и требует нестандартного подхода.

Ниже будут рассмотрены некоторые схемы и приемы оценки характеристик, позволяющие преодолеть эти методологические затруднения, применительно к наиболее распространенным моделям трендов. При этом особое внимание будет уделено надежности оцениваемых показателей.

2. Линейные модели трендов

Это наиболее часто используемая модель тренда, представляемого в виде линейной регрессии $y(t) = a_0 + a_1 t$, где t – текущее время, a_0 – смещение, a_1 – собственно характеристика тренда, убывающего ($a_1 < 0$) или возрастающего ($a_1 > 0$) на заданном промежутке времени. Оценка значения показателя роста (убывания) регрессии a_1 является основной задачей при выявлении линейного тренда. Если задан ряд наблюдений процесса $\mathbf{Y} = \{y_n\}$ в моменты времени $\{t_n\}$, $n = 1, 2, \dots, N$, $t_1 < t_2 < \dots < t_N$, то в рамках м.н.к. приближения значения параметров линейной регрессии определяются следующим образом:

$$\mathbf{A}_1 = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = (\mathbf{M}_1^T \mathbf{M}_1)^{-1} \mathbf{M}_1^T \mathbf{Y}, \quad \mathbf{M}_1 = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \dots & \dots \\ 1 & t_N \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix}, \quad (1)$$

где \mathbf{M}_1 – матрица переменных для линейной регрессии, \mathbf{A}_1 – параметры этой регрессии, а \mathbf{Y} – вектор наблюдений. Заметим, что при оценке параметров здесь используются все точки статистики одновременно. Точность приближения определяется в первую очередь рассеянием исходных данных (дисперсией s^2) относительно линии регрессии

$$s^2 = (\mathbf{Y}^T (\mathbf{E} - \mathbf{M}_1 (\mathbf{M}_1^T \mathbf{M}_1)^{-1} \mathbf{M}_1^T) \mathbf{Y}) / (N-2) \quad (2)$$

где \mathbf{E} – единичная матрица. Кроме того, показатель точности зависит от погрешностей оценки самих параметров регрессии \mathbf{A}_1 . Общая величина дисперсии s_p^2 при этом определяется следующим образом [1,3]

$$s_p^2(t) = s^2 \left(1 + \frac{1}{N} + \frac{(t-t_{cp})^2}{\sum_{n=1}^N (t_n - t_{cp})^2} \right), \quad t_{cp} = \frac{1}{N} \sum_{n=1}^N t_n. \quad (3)$$

Минимальное значение дисперсии достигается в средней точке интервала t_{cp} , а максимальное – в крайних точках, t_1 и t_N . При заданной доверительной вероятности $P_{дов}$ в соответствии с правилами статистики находится коэффициент Стьюдента t_α при $(N-2)$ степенях свободы, а затем ширину доверительного интервала $V = 2t_\alpha s_p$. При $P_{дов} \sim 0,95 \div 0,99$ и $N \sim 10 \div 20$ коэффициент t_α

находится в пределах от 2,0 до 3,2. Соответственно, величина доверительного интервала $V \sim (4,0 \div 6,4) s_p$.

В экономических и финансовых прогнозах для описания неопределенности чаще используется понятие волатильности (изменчивости), совпадающее по смыслу с доверительным интервалом и которое может быть расширенно на случай иных методов оценки такого интервала [4]. Воспользуемся этим показателем для проверки состоятельности параметров трендов. Очевидно, чем больше скорость нарастания процесса a_1 и меньше волатильность V , тем более наблюдаем тренд. И наоборот, чем меньше значение a_1 и больше неопределенность V , тем менее надежны оценки его параметров. В риск-анализе для определения и надежности предложено использовать отношение волатильности к величине оцениваемого показателя [4]. В статистике аналогичный показатель называют коэффициентом вариации K_v [1,2]. Чем меньше значение коэффициента K_v , тем более надежны полученные оценки, и наоборот, чем больше его значение, тем они менее надежны, а решения, принимаемые на их основе, более рискованны. Предельное значение K_v можно установить, исходя из следующих соображений. Если волатильность измерения (оценки) превышает само значение измеряемой величины, то очевидно, что результат заведомо не достоверен или просто лишен смысла (отрицателен). Поэтому условию состоятельности измерения, наблюдения или оценки соответствует естественное ограничение: $K_v < 1$. Возможны более жесткие условия состоятельности и меньшие предельные значения K_v .

Воспользуемся этим показателем для выяснения условий наблюдаемости трендов. Представим полученную регрессию в виде интервального образа $[y(t)]$, где в качестве координаты линейная регрессия $y(t)$ с волатильностью, равной доверительному интервалу

$$[y(t)] = a_0 + a_1 t + t_{\alpha} s_p(t) [-1, 1]. \quad (4)$$

Определим интервал параметра тренда $[a_1]$ как отношение приращения $[y(t)]$ на заданном промежутке времени $[t_1, t_N]$ к величине этого промежутка

$$[a_1] = \frac{[y(t_N)] - [y(t_1)]}{t_N - t_1}. \quad (5)$$

В соответствии с правилами интервальной арифметики и с учетом того, что $s_p(t_1) = s_p(t_N) = s_{\max}$ получаем:

$$[a_1(t)] = a_1 + \frac{t_{\alpha} s_{\max}}{t_N - t_1} ([-1, 1] - [-1, 1]) = a_1 + \frac{2 t_{\alpha} s_{\max}}{t_N - t_1} [-1, 1] = a_1 + V_A [-1, 1].$$

(6)

где a_1 и s_p ранее полученные оценки, а V_A – показатель волатильности тренда. Условие состоятельности оценки тренда теперь можно представить в следующем виде

$$K_V = \frac{V_A}{|a_1|} \frac{2t \pm s_{\max}}{|a_1| (t_N - t_1)} < 1 \quad (7)$$

или

$$\frac{s_{\max}}{|a_1|} \frac{t_N - t_1}{2t \pm} \quad (8)$$

Если это условие нарушается, то оценка тренда заведомо несостоятельна, а тренд не наблюдаем, хотя оценки параметров линейной регрессии могут быть получены из (1). Повышение надежности оценки достигается за счет увеличения длины ряда, но только при выполнении определенных условий стационарности. Основная трудность заключается в проверке этого условия применительно к тренду, как модели заведомо нестационарного процесса. В работе [5] предложен простой и естественный критерий стационарности. Регрессия считается стационарной в конечной точке ряда t_N из N элементов, если все ее параметры, включая ширину доверительного интервала, не отличаются от соответствующих значений, вычисленных в предыдущей точке t_{N-1} для $(N-1)$ точек того же ряда. Иными словами, регрессия стационарна, если расширение ряда на одну точку не влечет изменений ее параметров. В рамках м.н.к. приближения воспользоваться этим критерием в строгом смысле практически не удастся поскольку, как уже отмечалось ранее, при оценке параметров регрессии используются все точки ряда. Расширение ряда неизбежно влечет изменение размерности матрицы переменных M_1 и связанных с ней оценок. Строгие решения на основе предложенного критерия могут быть найдены вариационными методами.

Будем полагать, что все измерения безошибочны, а расхождения между оценочными и фактическими значениями показателей относятся на счет скрытых факторов. Регрессия при этом строится в виде интервального образа – ленточной функции постоянной ширины $[y(t)] = y(t) + V [0, 1]$ с координатой $y(t)$ и волатильностью V . Основным условием ее идентификации здесь является требование полноты охвата статистики. Для заданного ряда наблюдений Y параметры интервальной регрессии $[y(t)] = a_0 + a_1 t + V [0, 1]$ определяются

как решение следующей оптимизационной задачи по критерию наилучшего приближения

$$\begin{aligned} & \text{Min } V & (9) \\ & \{A_1, V\} \\ & M_1 A_1 \leq Y, \quad M_1 A_1 + VI \geq Y, \quad V \geq 0, \end{aligned}$$

где I – вектор из единиц, размерностью по числу точек наблюдений N . Заметим, что здесь одновременно варьируются значения параметров A_1 и показатель волатильности регрессии V . Решению задачи (9) (за исключением тривиальных и вырожденных случаев) соответствует три активных ограничения (по числу варьируемых переменных) и столько же точек достаточной статистики, которые образуют альтернанс $G = \{g_1, g_2, g_3\}$, $g_1 < g_2 < g_3$, $g_{1,2,3} \in \{t_1, t_2, \dots, t_N\}$. Остальные точки временного ряда при этом не информативны. Интервальная регрессия будет стационарна, если последняя точка ряда t_N находится в пределах доверительного интервала и не входит в состав альтернанса $t_N \notin G$. В этом случае она не влияет на выбор параметров регрессии, а, следовательно, саму регрессию можно считать стационарной [5]. Таким образом, вопрос стационарности регрессии в рамках вариационного подхода имеет простое и однозначное решение.

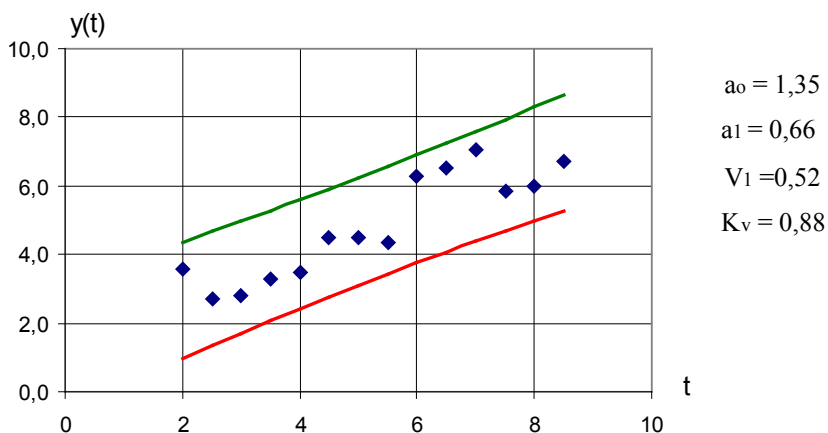
Воспользуемся определением (5) и найдем интервальную оценку основной характеристики тренда $[a_1]$, а также ее коэффициент вариаций

$$[a_1] = a_1 + \frac{V}{t_N - t_1} ([0, 1] - [0, 1]) = a_1 + \frac{V}{t_N - t_1} [-1, 1]. \quad (10)$$

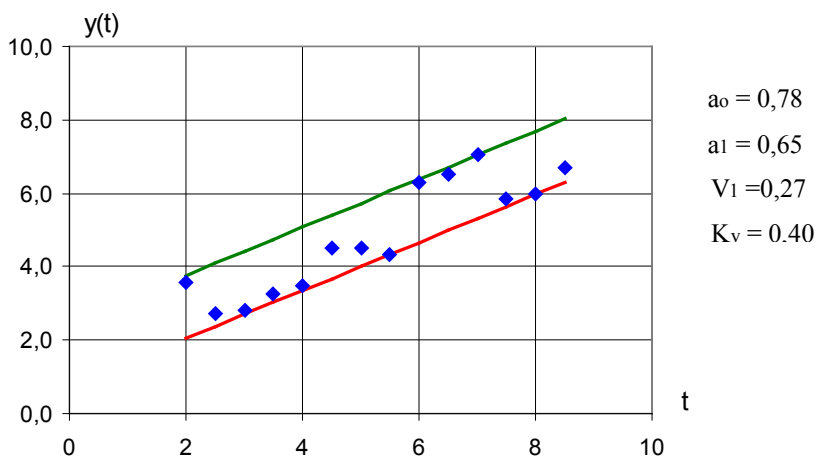
Сопоставим качество решений, полученных двумя методами на конкретном примере (Рис. 1). Оценки основного параметра тренда a_1 , как оказалось, практически совпадают. В то же время надежность оценки, полученной на основе интервального образа, существенно выше, нежели м.н.к. приближение. Построенный интервальный образ регрессии к тому же удовлетворяет условию стационарности. С учетом этих особенностей второй способ оценки тренда предпочтительней.

Исходный временной ряд

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14
t_n	2,0	2,5	3,0	3,5	4,0	4,5	5,0	5,5	6,0	6,5	7,0	7,5	8,0	8,5
y_n	3,58	2,70	2,81	3,26	3,50	4,51	4,50	4,34	6,29	6,51	7,04	5,83	5,97	6,70



М.Н.К. приближение, доверительный интервал регрессии



Интервальный образ регрессии

рис. 1

В полной мере преимущества вариационного подхода раскрываются при оценке и анализе непараметрических характеристик трендов, связанных со свойствами регрессионной модели в целом. К их числу в первую очередь следует отнести ожидаемое «время жизни» тренда, т.е. оценку горизонта

прогноза по линейной регрессии. Решение этой задачи подробно рассмотрено также в работе [5]. Значение горизонта определяется по точкам альтернанса и в общем случае предполагает три вида оценок: надежную, рискованную и предельную (границу состоятельности прогноза).

В качестве другой характеристики тренда рассмотрим момент его рождения T . В ряде случаев этот показатель становится одним из определяющих, в частности, при управлении процессами с запаздывающей реакцией, аналогичных доменному производству в металлургии. Основная проблема заключается в выборе информативного признака, позволяющего надежно регистрировать момент рождения тренда по некоторой переходной характеристике. Известными методами решить эту задачу довольно трудно. В качестве альтернативы рассмотрим процедуру, основанную на аппроксимации переходной траектории двумя регрессиями. При этом постараемся остаться в рамках задач линейного программирования.

Ограничимся случаем вогнутых траекторий, когда скорость изменения процесса в начале траектории существенно меньше, нежели в конце наблюдений. Исходные данные представим в виде ряда из N наблюдений (окна) $Y = \{y(t_n)\}$ в точках $t_n = T_0 + \Delta t (n-1)$ с началом отсчета в момент T_0 и шагом Δt . Размер окна N обычно выбирается в пределах от 6 до 10 точек [3,5]. Для этого ряда Y будем строить две линейные огибающие $(a_0 + a_1 t)$ и $(b_0 + b_1 t)$, удовлетворяющие условию полноты охвата статистики

$$y_n - (a_0 + a_1 t_n) = c_n \geq 0, \quad y_n - (b_0 + b_1 t_n) = d_n \geq 0, \quad n = 1, 2, \dots, N \quad (11)$$

и такие, что $a_1 < b_1$. Ограничим величину отклонений сверху

$$c_n \leq e, \quad d_n \leq e, \quad n = 1, 2, \dots, N \quad (12)$$

и будем минимизировать значение этой верхней границы

$$\min e, \quad (13)$$

варьируя параметры регрессий a_0, a_1, b_0, b_1 . Кроме того, необходимо внести асимметрию в свойства регрессий, позволяющую привязать их к различным частям траектории с разными трендами. Для вогнутой траектории это можно представить в виде следующей системы ограничений

$$2c_n \leq c_{n+1} + c_{n+2}, \quad 2d_n \geq d_{n+1} + d_{n+2}, \quad n = 1, 2, \dots, (N-2). \quad (14)$$

Для первой регрессии отклонения от траектории должны только возрастать (в среднем), а для второй – убывать. Ограничения (11), (12), (14) вместе с критерием (13) образуют задачу линейного программирования. Ее решение позволяет найти точку рождения тренда T на пересечении полученных огибающих из следующего интервального уравнения:

$$a_0 + a_1 T + e_1 [0,1] = b_0 + b_1 T + e_2 [0,1], \quad (15)$$

где e_1 и e_2 - показатели точности приближения по каждой огибающей на соответствующих фрагментах ряда Y : для первой из них $e_1 = \max c_n$ при $t_n \leq T$ и для второй $e_2 = \max c_n$ при $t_n > T$. Нижняя граница T_{\min} и волатильность V_T точки рождения при этом определяются как

$$T_{\min} = \frac{a_0 - b_0 - e_2}{b_1 - a_1}; \quad V_T = \frac{e_1 + e_2}{b_1 - a_1}; \quad b_1 > a_1 \quad (16)$$

Состоятельность результата может быть оценена коэффициентом вариации

$$K_T = \frac{V}{T_{\min}} = \frac{e_1 + e_2}{a_0 - b_0 - e_2} < 1 \quad (17)$$

Представленная здесь схема выявления точки рождения тренда во многих случаях дает удовлетворительный результат (Рис.2). Безусловно, она не единственная. Однако другие модели рождения трендов будут, скорее всего, нелинейными, со всеми вытекающими проблемами сходимости и точности, которые существенно осложняют решение такого рода задач.

3. Интервальные модели нелинейных трендов.

Это большой класс моделей, широко используемых в технико-экономических и других приложениях. Следует заметить, что вопросам качества идентификации таких моделей в публикациях уделялось крайне мало внимания. В первую очередь это связано со сложностью вычисления их параметров и показателей точности приближения. В данной работе мы ограничимся двумя видами нелинейных моделей, на которых продемонстрируем

возможности вариационного подхода для оценки их параметров и общих характеристик.

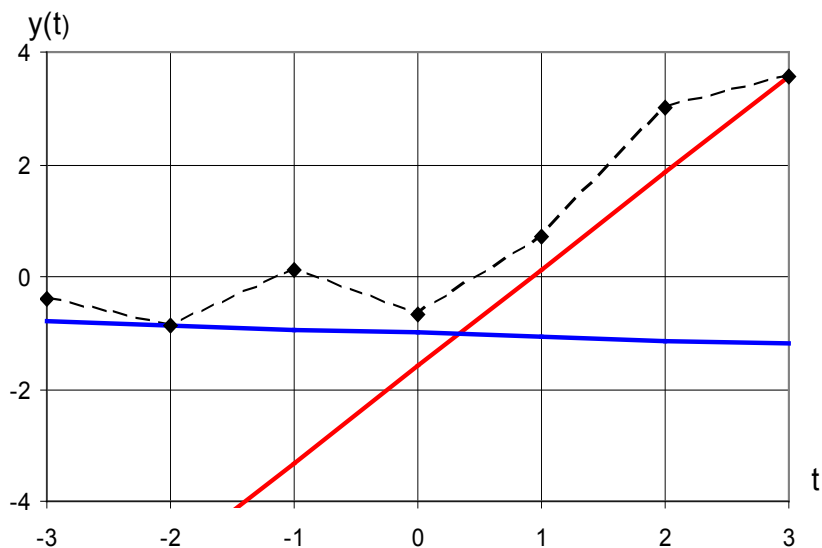


Рис. 2. Оценка момента рождения тренда по огибающим

В качестве первой из них рассмотрим параболическую регрессию, выпуклую вверх. Интервальный образ такой модели определяется из решения оптимизационной задачи (9), но с другой матрицей переменных

$$M_2 = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \dots & \dots & \dots \\ 1 & t_N & t_N^2 \end{bmatrix}, \quad (18)$$

построенной на основе заданного временного ряда $\{t_n\}$, $n = 1, 2, \dots, N$. Интервальная парабола при этом имеет следующий вид:

$$[y(t)] = at^2 + bt + c + V[0,1]; \quad a < 0, \quad (19)$$

где все параметры, включая волатильность V , результат решения задачи (9). Наряду с параметрами особый интерес представляет оценка надежности ожидаемой (или прошедшей) точки максимума $t_{\max} = -b / 2a$ на параболе. Значение максимума при этом находится в пределах следующего интервала

$$[y_{\max}] = c - \frac{b^2}{4a} + V[0,1] \quad (20)$$

Соответственно, все точки траектории, попадающие в этот интервал (20), могут быть отнесены к категории экстремальных. Оценим возможное расхождение τ относительно t_{\max} для интервальной модели (19):

$$a(t_{\max} + \tau)^2 + b(t_{\max} + \tau) + c = c - b^2/4a - V. \quad (21)$$

Подставим значение t_{\max} в последнее уравнение и решим его относительно τ

$$\tau^2 = \frac{V}{|a|} \quad (22)$$

Теперь интервал значений экстремальной точки t_{\max} можно определить следующим образом

$$[t_{\max}] = -\frac{b}{2a} + \sqrt{\frac{V}{|a|}} [-1, +1] \quad (23)$$

Заметим, что ширина этого интервала меняется в зависимости от волатильности по закону квадратного корня. Условие состоятельности оценки t_{\max} в параметрах интервальной параболы выглядит следующим образом

$$K_{\max} = \frac{\tau}{t_{\max}} = \frac{2}{b} \sqrt{V|a|} < 1 \quad (24)$$

Однако это условие не единственное. Необходимо также учесть ограничения по горизонту прогнозирования, если конечная точка наблюдений находится на восходящей ветви параболы, при $t_{\max} > t_N$. Подробно эти вопросы рассмотрены в [5].

Оценки горизонта прогнозирования строятся на основе анализа альтернанса G , получаемого при решении задачи идентификации (9). Для модели параболы число точек альтернанса равно четырем; $G = \{g_1, g_2, g_3, g_4\}$, $g_{1,2,3,4} \in \{t_n\}$, $t_1 \leq g_1 < g_2 < g_3 < g_4 \leq t_N$. Нижняя h_1 и верхняя h_2 границы горизонта прогнозов при этом определяются следующим образом:

$$\begin{aligned} h_1 &= \min \{(g_2 - g_1), (g_3 - g_2), (g_4 - g_3)\}, \\ h_2 &= \max \{(g_1 - t_1), (g_2 - g_1), (g_3 - g_2), (g_4 - g_3), (t_N - g_4)\}. \end{aligned} \quad (25)$$

Если выполняется условие стационарности ($t_N > g_4$), то можно говорить о надежном H_1 и рискованном H_2 горизонтах прогноза: $H_1 = h_1 - t_N + g_4 \geq 0$, $H_2 = h_2 - t_N + g_4 \geq 0$, $H_2 \geq H_1$. При этом, если точка экстремума находится за пределами горизонта прогнозирования, т.е. $(t_{\max} - t_N) > H_2$, экстремум не наблюдаем. Если $H_1 < (t_{\max} - t_N) < H_2$, то точку экстремума следует отнести к сомнительным. И лишь при $(t_{\max} - t_N) < H_1$ экстремальную точку можно считать в полной мере состоятельной. Таким образом, в рамках интервальных представлений неизбежно приходим к трехзначной логике принятия решений. Соответственно, в терминах ситуационного управления можно говорить о режиме ожидания в первом случае, фазе подготовки и превентивных решений во втором, а также фазе активных действий в третьем случае.

В качестве второго примера рассмотрим модель экспоненциального процесса, характерного для инерционных объектов. При этом будем анализировать реакцию на ступенчатое воздействие, полагая ее затухающей. Интервальную модель такого процесса можно представить в следующем виде

$$[y(t)] = A (1 - \exp(-\alpha(t - t_0))) + V [0,1], \quad (26)$$

где $[y(t)]$ – интервальная координата, $A > 0$ – относительная величина скачка, t_0 – момент его возникновения, α – показатель затухания, а V – волатильность восстановленной траектории.

Пусть задан фрагмент траектории в виде наблюдений $Y = \{y_n\}$ в моменты времени $\{t_n\}$, $n = 1, 2, \dots, N$, с началом отсчета в точке $t_1 > t_0$. В рамках ситуационного управления обычно интересуют две характеристики, определяющие характер решений: момент воздействия t_0 и предельное значение координаты $y(\infty) = A$. Параметры интервальной модели (26) будем искать как решение следующей оптимизационной задачи:

$$\begin{aligned} \min V & \\ \{A, \alpha, t_0, V\} & \\ y_n \leq A [1 - \exp(-\alpha(t_n - t_0))] \leq y_n + V, & \quad n = 1, 2, \dots, N, \\ A > 0, \quad \alpha \geq 0, \quad t_0 \geq 0, \quad V \geq 0. & \end{aligned} \quad (27)$$

Данная задача относится к категории нелинейных. Однако ее решение не вызывает серьезных осложнений и может быть найдено в стандартной

вычислительной среде типа SOLVER, EXCEL.

Предельное состояние траектории $[y(\infty)] = A + V [0,1]$ оценивается точностью приближения модельной траектории V и при условии $KA = V/A < 1$ или $V < A$. В этом же случае имеет смысл оценка момента воздействия $[t_0] = t_0 + Vt [0,1]$ в прошлом. Волатильность Vt и показатель надежности K_t этой характеристики могут быть найдены из следующего соотношения

$$A [1 - \exp(-\alpha V_t)] = V. \quad (28)$$

Соответственно

$$V_t = -\frac{1}{\alpha} \ln \left(1 - \frac{V}{A} \right) \approx \frac{V}{A\alpha} ; K_t = \frac{V_t}{t_0} = \frac{V}{A\alpha t_0} = \frac{K_A}{\alpha t_0} < 1 \quad (29)$$

Последнее приближение справедливо при $V \ll A$. Обратим внимание, что показатели надежности K_A и K_t здесь взаимосвязаны. Причем, при $\alpha t_0 > 1$ определяющим является K_A , в случае $\alpha t_0 < 1$ - второй из них K_t . Полученных характеристик переходного процесса $[y(\infty)]$ и $[t_0]$ обычно достаточно для обоснования надежных решений по нейтрализации или компенсации возмущений.

4. Парные регрессии

Под парной регрессией понимается система из двух взаимосвязанных регрессий, позволяющих выделить на плоскости переменных некоторую область, отвечающую заданному критерию и ограничениям. В биржевых моделях этому понятию соответствуют флажки Доу, полосы Болинджера, линии Демарка и др.[6]. В общем случае речь идет о паре окаймляющих регрессий, определяющих изменение границ выделенной по заданным признакам области и на основе доступной статистики.

Подобные задачи в полной мере соответствуют основным положениям вариационного подхода в регрессионном анализе. В данной работе ограничимся моделью флажка Доу в виде двух сходящихся линий $x(t) = a_0 + a_1 t$ и $z(t) = b_0 + b_1 t$, $b_0 > a_0$ и $a_1 > b_1$, образующих конус на плоскости. Для заданного ряда наблюдений $Y = \{y_n\}$ в точках $\{t_n\}$, $n = 1, 2, \dots, N$, выбор границ подчиним следующей системе ограничений:

$$a_0 + a_1 t_n \leq y_n ; b_0 + b_1 t_n \geq y_n ; n = 1, 2, \dots, N, \quad (30)$$

которые можно интерпретировать как условие полноты охвата статистики. В качестве критерия подбора границ выберем площадь выделенной области (трапеции) S в пределах интервала наблюдений $[t_1, t_N]$. Минимум этого критерия

$$\min_{\{a_0, a_1, b_0, b_1\}} S \sim \min [2(b_0 - a_0) + (b_1 - a_1) (t_N - t_1)] \quad (31)$$

при соблюдении ограничений (30) соответствует наилучшему приближению модели флажка.

Полученное решение позволяет не только выявить вид флажка, но и вычислить его общие характеристики, в частности, определить положение точки схождения границ t^* из очевидного условия: $a_0 + a_1 t^* = b_0 + b_1 t^*$. Однако, как и ранее, при обосновании состоятельности оценки t^* необходимо учесть ограничения по надежности прогноза в окрестности точки схождения. Аналогичная процедура идентификации в принципе применима к любым разновидностям парных регрессий.

Литература

1. Фишер Ф. Проблема идентификации в эконометрии. Перевод с английского и предисловие Г.Г. Пирогова – М.: Статистика, 1978.
2. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрии – М.: Юнити, 1998
3. Четыркин Е.М. Статистические методы прогнозирования. – М.: Статистика, 1977
4. Шоломицкий А.Г. Теория риска. Выбор при неопределенности и моделирование риска. – ГУ ВШЭ, 2005
5. Подружко А. А., Подружко А.С. Интервальное представление надежных регрессий. // Динамика неоднородных систем. Вып. 10 – М.: Комкнига, 2006.
6. Берзон Н.И., Буянова Е.А., Кожесвикова М.А., Чаленко А.В. Фондовый рынок. – М.: Вита-Пресс, 1998.