

УДК 621.002:681.324

В.В. Голикова, Е.Л. Первухина

**ВЫЧИСЛИТЕЛЬНЫЕ АСПЕКТЫ ПОИСКА ПРИЧИННО-
СЛЕДСТВЕННЫХ ЗАВИСИМОСТЕЙ МЕЖДУ ПАРАМЕТРАМИ
СЛОЖНЫХ СИСТЕМ**

В основе решения многих задач современного машиностроения лежат качественные и количественные зависимости между параметрами выпускаемых и вновь создаваемых изделий. Эти изделия в большинстве случаев представляют сложные системы, состоящие из разнородных элементов, механизмов и узлов, объединенных пространственными, функциональными, информационными отношениями, меняющимися со временем. Поэтому исследование таких систем и выявление указанных отношений и зависимостей, представляет сложную научную проблему. Серьезным недостатком теоретико-аналитического подхода к решению проблемы является необходимость абстрагирования от многообразия свойств и характеристик изделий, а также ограничение по количеству изучаемых параметров и времени получения результатов [1]. При возрастании сложности изделий знания об их свойствах и поведении становятся в значительной степени неопределенными, что также снижает эффективность теоретико-аналитического подхода. В последнее время в практике прикладных исследований большое распространение получил экспериментально-статистический подход [2,3], позволяющий эмпирически исследовать непосредственно изделия, их элементы и строить модели путем статистической обработки результатов измерений параметров.

К настоящему времени разработано достаточное количество методов статистического анализа сложных систем. Однако большая их часть не работоспособна при наличии автокорреляций и других систематических воздействиях и, самое главное, не позволяет установить причинно-следственные зависимости между параметрами. Методы построения векторных авторегрессионных моделей [4,5] для многомерных статистических систем связаны, как правило, с экономическими приложениями и неоправданно редко используются в практике исследования технических и технологических систем.

Среди последних публикаций, раскрывающих методы статистического анализа сложных технических объектов и систем, наибольший интерес представляет монография [3], в которой наиболее полно изложены вопросы формализованного выбора устойчивых структур многофакторных статистических моделей, и работа [6], где результаты факторного эксперимента используют для целей диагностирования таких объектов. Важные выводы указанных работ заключаются в том, что в отсутствие конструкторских или технологических дефектов статистические связи между параметрами технических объектов стабильны. При наличии дефектов или развивающихся повреждений происходит нарушение связей. Очевидно, что прогнозирование момента нарушения связей позволит определить предотказные состояния объектов. Однако в указанных работах рассматриваются только корреляционные связи между параметрами и проводится многофакторный анализ, результаты которого не могут быть в полной мере использованы для решения задачи, поскольку, с одной стороны, в сложных системах полный многофакторный эксперимент невозможен и должен быть заменен детальным экспериментом, как следствие, интересующие параметры могут не войти в статистическое описание системы. С другой стороны, нет необходимости рассматривать множество всех параметров при решении конкретных задач определения статистических взаимосвязей между отдельными параметрами (при наличии априорной информации о физической природе исследуемых объектов, их качественных характеристиках и предварительного исключения возможного влияния неинформативных параметров).

Предлагается алгоритм определения причинно-следственных зависимостей между параметрами сложных технических объектов, рассматриваемых как динамические системы, при описании последовательности равноотстоящих во времени результатов наблюдений за параметрами в виде векторной авторегрессии. Решение основано на методике выделения неизменных во времени комбинаций последовательностей измерений различных параметров между собой.

В большинстве задач по исследованию сложных систем информационные потоки формируют на основе последовательностей измерений параметров. Как правило, каждый из параметров образует

самостоятельную случайную последовательность $X_{i,1}, \dots, X_{i,t}$, i - номер исследуемого параметра, $t = 1, \dots, T$ - дискретное время. В момент времени t параметры образуют вектор параметров $x_t = (X_{1,t}, \dots, X_{n,t})'$, / - символ транспонирования. Для примера выбраны случайные последовательности измерений трех параметров, характеризующих работу карбюраторного двигателя внутреннего сгорания в ходе контрольных испытаний (рис.1). Обозначение и размерность параметров представлены в табл. 1. По исследуемой методике [4,7,8] разработаны алгоритмы и вычислительные процедуры с использованием прикладного программного обеспечения RATS / CATS [9,10].

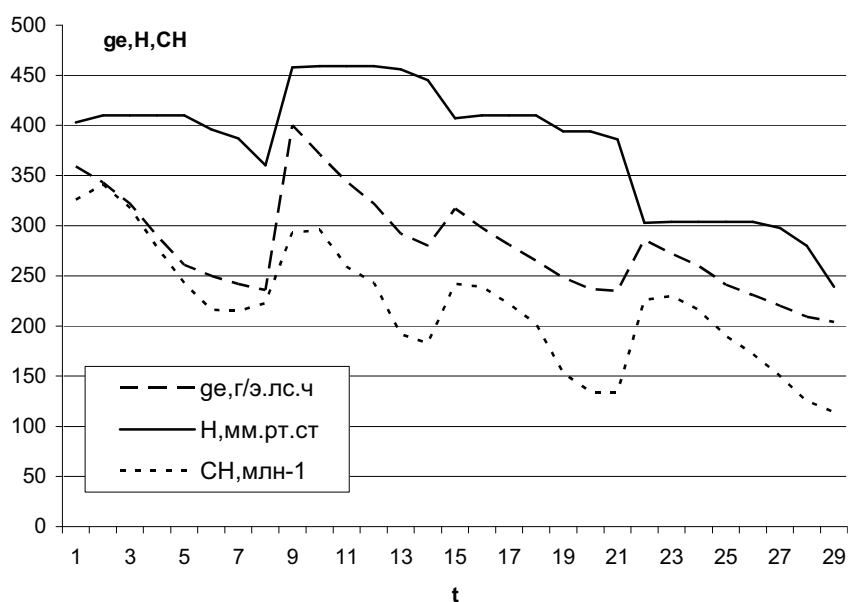


Рисунок 1 – Исследуемые случайные последовательности измерений параметров карбюраторного двигателя внутреннего сгорания

Таблица 1

Исследуемые параметры

Обозначение	Параметр	Размерность	Обозначение в работе
ge	Удельный расход топлива	г/э.лс.ч	X_1
H	Разрежение давления впускном коллекторе	мм.рт.ст.	X_2
CH	Содержание отработавших газов	млн ⁻¹	X_3

Рис. 1 демонстрирует нестационарность и схожие тенденции в изменении значений заданных последовательностей. На первом этапе аналитической проверки этого предположения выполнен корреляционный анализ. Парный коэффициент корреляции характеризует величину статистической связи между каждыми двумя параметрами [7]:

$$\rho_{ik} = \frac{E[(X_i - EX_i)(X_k - EX_k)]}{\sqrt{E[(X_i - EX_i)^2]E[(X_k - EX_k)^2]}}. \quad (1)$$

Здесь ρ_{ik} - обозначение парного коэффициента корреляции между i -м и k -м параметрами, $E[\cdot]$ - символ математического ожидания. Выборочные значения коэффициентов (1) для всех параметров представлены в табл. 2.

Таблица 2

Корреляционные зависимости между параметрами

	ge, г/эл.с.ч	H, мм.рт.ст.	СН, млн ⁻¹
ge, г/эл.с.ч	1	0.699344	0.849649
H, мм.рт.ст.	0.699344	1	0.531973
СН, млн ⁻¹	0.849649	0.531973	1

Как видно из табл. 2, все параметры попарно обладают сильной корреляционной зависимостью. Однако причиной этой зависимости может быть некоторый дополнительный параметр или общая тенденция параметров к уменьшению. Для исключения данной причины, или наоборот, подтверждения гипотезы о существовании общей тенденции к изменению случайных последовательностей измерений параметров $X_1 - X_3$, выполнен, так называемый, коинтеграционный анализ [7,8,11,12].

Случайные нестационарные последовательности измерений исследуемых параметров сведены к стационарным с помощью оператора разностей [7]. Так, разности первого порядка определены как $Y_{i,t} = X_{i,t} - X_{i,t-1}$, и исходные процессы $X_{i,t}$ тогда называют интегрированными процессами первого порядка. В общем случае элементы вектора $y_t = (Y_{1,t}, \dots, Y_{n,t})'$ называют коинтегрированными порядка d, b и обозначают $y_t \sim CI(d, b)$, если они представляют интегрированные процессы порядка d и существует отличный от нуля

вектор β , такой что линейная комбинация $y_t\beta$ есть интегрированный процесс порядка $(d-b)$. Вектор β называют коинтегрирующим вектором.

Методики Энгла-Грейнджера и Йохансена [11,12] для проверки выдвинутой гипотезы предполагают последовательное выполнение нескольких этапов. На первом этапе визуально оценивают графики последовательностей измерений (рис.1). На втором этапе определяют ранг коинтеграции r по числу линейно независимых коинтегрирующих векторов, не превосходящему число исследуемых параметров. Далее проверяют коинтегрирующие ограничения и строят модель исправления ошибки для выбора окончательной структуры аналитического описания динамики исследуемых параметров. Построение модели исправления ошибок связано с выдвинутой гипотезой общей тенденции изменения параметров, или некоторой общей траектории, от которой они могут отклоняться, но к которой возвращаются при нормальном режиме функционирования объекта.

Выборочные значения автокорреляционных функций для каждой последовательности определяют с использованием выражения, аналогичного парному коэффициенту корреляции (1):

$$\rho_k = \frac{E[(Y_t - \mu)(Y_{t+k} - \mu)]}{\sqrt{E[(Y_t - \mu)^2]E[(Y_{t+k} - \mu)^2]}}, \quad \mu = E[Y_t],$$

как зависимости коэффициента ρ_k от k

j -й коэффициент автокорреляционной функции можно представить в виде

$$\rho_j = \phi_{k1}\rho_{j-1} + \dots + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k}, \quad j=1,2,\dots,k,$$

где ϕ_{kk} определяет частную автокорреляционную функцию от k . Графики автокорреляционной и частной автокорреляционной функций последовательностей измерений всех параметров $X_1 - X_3$ (рис.2) практически не убывают с ростом t после нескольких первых значений, что подтверждает нестационарность последовательностей [7]. Анализ графиков дополнен тестом Дики-Фуллера по проверке гипотезы о существовании единичных корней и расширенным тестом Дики-Фуллера, когда дополнительно исследуют необходимость включения в уравнение

$$\Delta X_{i,t} = \alpha + \mu t + \Psi D_t + \gamma_i X_{i,t-1} + \sum_{j=1}^{p-1} \delta_j \Delta X_{i,t-j} + \varepsilon_{i,t}; i=1,2,3 \quad (2)$$

свободного члена α и тренда t , а также детерминации длины лага. В основу тестов положена регрессия с коэффициентами $\alpha, \mu, \gamma, \delta_j; \varepsilon_t$ – остаточный член регрессии (ошибка оценивания), p – количество лагов. В ходе тестирования последовательно проверены гипотезы о равенстве нулю (отдельно или совместно) коэффициентов α, μ, γ . Величина лага выбрана таким образом, чтобы устранить автокорреляцию ошибок оценивания коэффициентов в уравнении (2).

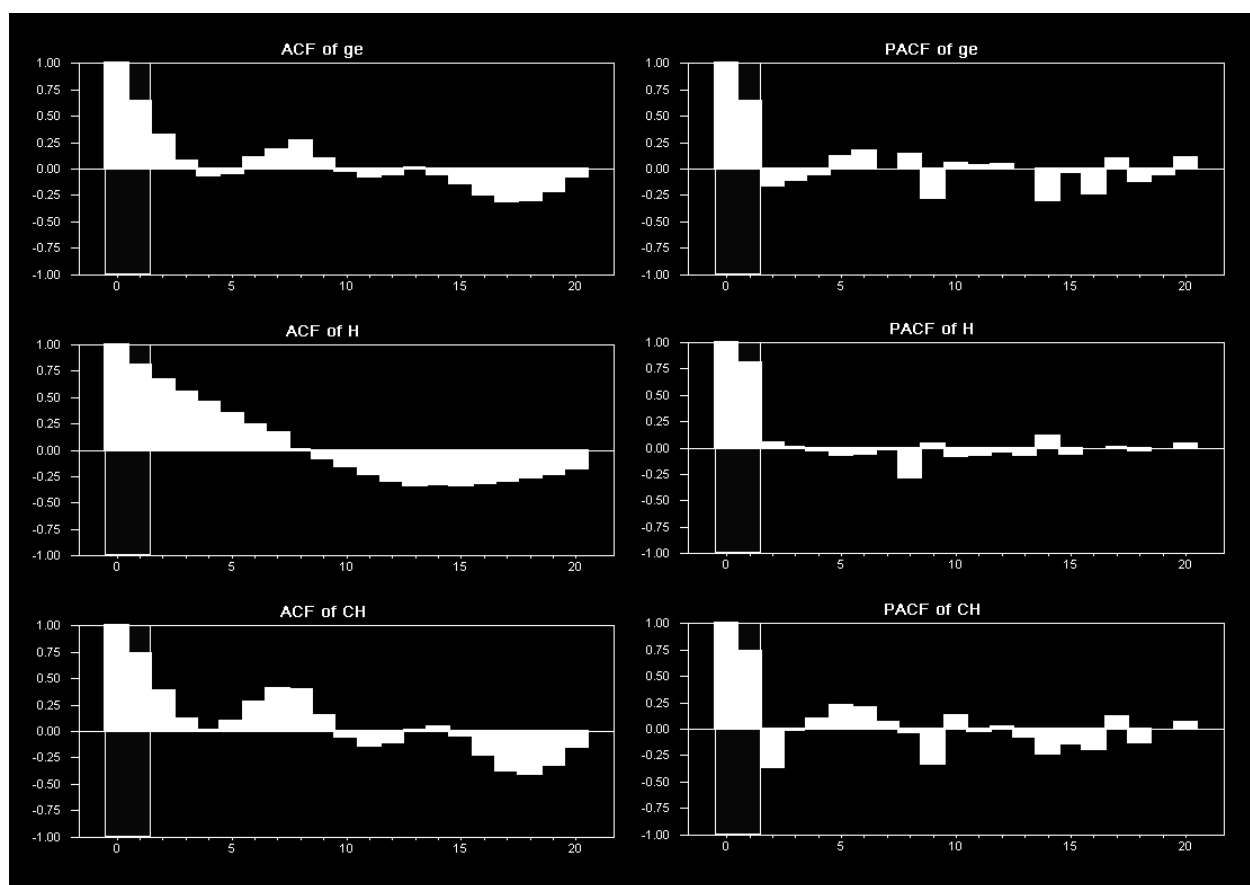


Рисунок 2 – Автокорреляционная (ACF) и частная автокорреляционная (PACF) функции последовательностей измерений

Результаты расширенного теста Дики-Фуллера для параметров $X_1 - X_3$, а также их первых разностей представлены в табл. 3, количество лагов равно шести.

Результаты табл. 3, как и графики на рис. 1,2, свидетельствуют о нестационарности последовательностей измерений параметров $X_1 - X_3$, поскольку гипотеза о наличии единичного корня не отвергнута на всех рассмотренных уровнях значимости. При

тестировании первых разностей случайных процессов на стационарность гипотеза о существовании единичного корня отвергнута на всех уровнях значимости, и, следовательно, трансформированные последовательности являются стационарными. В результате исходные последовательности $X_1 - X_3$ приняты интегрированными первого порядка. Стационарность первых разностей Dge, DH, DCH последовательностей ge, H, CH дополнительно иллюстрирует рис. 3.

Таблица 3

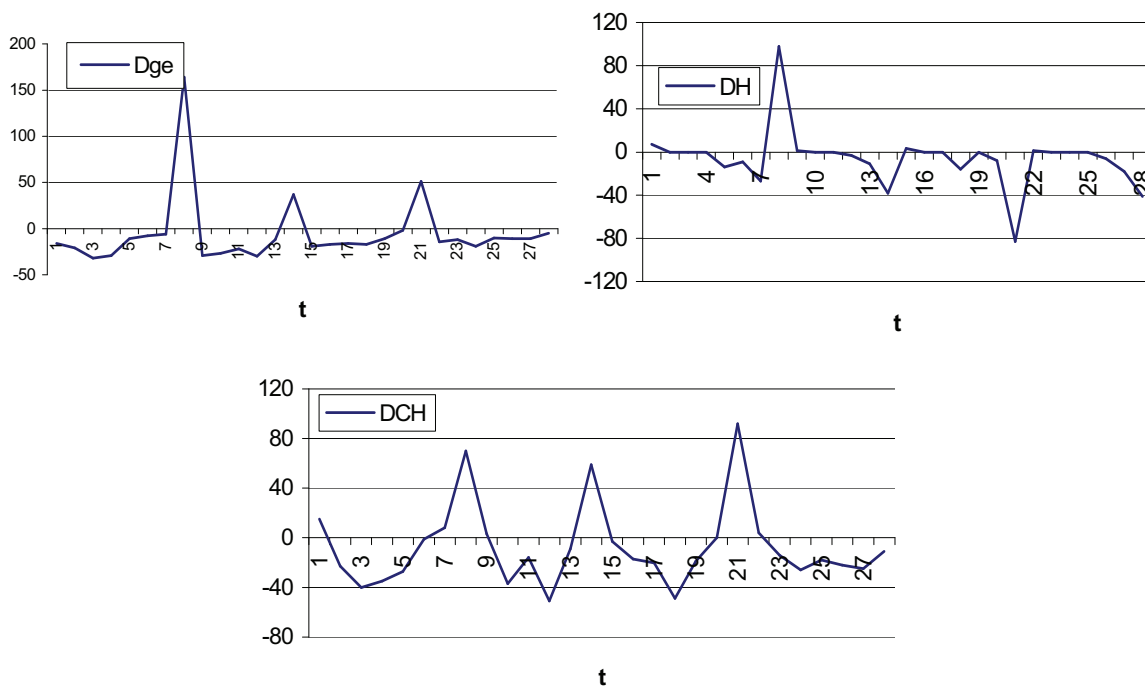
Результаты расширенного теста Дики-Фуллера

Параметры	t-adf	Критические значения (5%)	Единичный корень	Константа	Тренд
ge	-2.151	-2.97	есть	есть	нет
H	0.005	-2.97	есть	есть	нет
CH	-1.518	-2.97	есть	есть	нет
Dge	-5.240**	-2.97	нет	есть	нет
DH	-5.173**	-2.97	нет	есть	нет
DCH	-3.870**	-2.97	нет	есть	нет

Для проверки коинтегрированности последовательностей $X_1 - X_3$ проведен тест Йохансена [12]. На первом шаге тестирования в векторную авторегрессионную модель размерности p включают все возможные параметры [12]:

$$x_t = \pi_0 + \sum_{j=1}^p A_j x_{t-j} + \Psi \cdot D_t + \varepsilon_t, t = 1, \dots, T \quad (3)$$

В выражении (3) анализируемый вектор $x_t = (X_{1,t}, \dots, X_{n,t})'$, вектор констант $\pi = (\pi_1, \dots, \pi_n)'$, вектор ошибок оценивания (остатков) $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{nt})'$ имеют размерность $(n \times 1)$, n - число исследуемых параметров, $A_j = (a_{ik}(j))$; $i, k = 1, \dots, n$, $j = 1, \dots, p$. Для проверки гипотезы о длине лага использован тест отношения правдоподобия, имеющего распределение χ^2 . Детерминированный член D_t в общем случае состоит из линейного члена, периодической переменной и других регрессоров.

Рисунок 3 - Стационарность первых разностей рядов $X_1 - X_3$

После оценки векторной авторегрессионной модели (3) в тесте Йохансена рассчитывается оценка матрицы $\pi = I - A_1 - A_2 - \dots - A_p$, соответствующая представлению модели исправления ошибки:

$$\Delta x_t = \pi_0 + \pi x_{t-1} + \sum_{j=1}^{p-1} \pi_j \Delta x_{t-j} + \Psi \cdot D_t + \varepsilon_t, t=1, \dots, T \quad (4)$$

Нулевой гипотезой является то, что ранг матрицы π не превышает некоторого числа k : $H_0: r < k$. В качестве альтернативной гипотезы используется $H_1: \text{rank} \pi = k$ или $H_1: \text{rank} \pi = r + 1$. Статистика для проверки нулевой гипотезы против первой из приведенных альтернативных называется статистикой следа и имеет вид [12]:

$$LR^{\text{trace}} = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i) \quad (5)$$

Статистика (5) пропорциональна сумме логарифмов собственных чисел матрицы (остальные $(k - r)$ собственных чисел считаются нулевыми), т.е. следу матрицы. Для проверки против второй альтернативной гипотезы тестовая статистика принимает вид:

$$LR^{\lambda-\max} = -T \ln(1 - \hat{\lambda}_{r+1}) \quad (6)$$

и называется статистикой максимального собственного числа. Статистика (6) пропорциональна логарифму максимального из собственных чисел матрицы, считающихся нулевыми. В приведенных

выражениях (5) и (6) T - число наблюдений, а через $\hat{\lambda}_i$ обозначена оценка максимального правдоподобия i -го корня полученного Йохансенем уравнения. Корни упорядочены в порядке убывания.

Результаты вычислений статистик (5) и (6) для рассматриваемого примера приведены в табл. 4.

Таблица 4

Определение ранга коинтеграции

Собственные числа	$LR^{\lambda-\max}$	LR^{trace}	$H_0 :$	$LR^{\lambda-\max}$ (90%)	LR^{trace} (90%)
0.5598	22.16	29.54	0	13.39	26.70
0.1618	4.77	7.39	1	10.60	13.31
0.0925	2.62	2.62	2	2.71	2.71

По результатам анализа табл. 4 ранг коинтеграции принят равным 1 ($r=1$).

В выражении (4) матрица $\pi = \alpha\beta^T$, где α и β - ($n \times r$) матрицы полного ранга r . Матрицы α и β^T и матрица π представлены в табл. 5. Каждая строка матрицы β^T содержит компоненты одного из возможных коинтегрирующих векторов, причем эти векторы линейно независимы, а элементы матрицы α интерпретируются как скорость приведения процесса к равновесию. Первая из строк матрицы β^T содержит коэффициенты линейной комбинации последовательностей, ближайшей к стационарной. Вторая строка соответствует линейной комбинации, занимающей в этом отношении второе место, и т.д.

Таблица 5

Результаты вычисления матриц α и β^T

BETA (transposed)				ALPHA			PI			
GE	H	CH	CONST	9.827	9.647		GE	H	CH	CONST
-0.028	0.014	-0.010	5.243	-9.372	3.758	9.456	-1.041	0.416	0.182	91.016
-0.013	0.014	0.002	-3.198	17.001	12.330	0.998	-0.238	-0.025	0.424	-24.882
-0.048	0.006	0.034	3.836				-0.678	0.412	-0.112	53.533

В рассматриваемом примере ранг коинтеграции равен 1 ($r = 1$), значит, в качестве оценки коинтегрирующего вектора выбран вектор с элементами первой строки, т.е.

$$(-0.028, 0.014, -0.010, 5.243)' \quad (7)$$

Коинтегрирующий вектор получают из вектора (7) делением всех элементов на первый элемент:

$$\beta_{(1)} = (1.000, -0.497, 0.365, -187.514)' \quad (8)$$

Полученный коинтегрирующий вектор характеризует параметры статистически значимой долгосрочной зависимости между рассмотренными параметрами:

$$(est)_t = ge_t - 0.497 * H_t + 0.365 * CH_t - 187.514 \quad (9)$$

Для тестирования коинтеграции случайных последовательностей по методу Энгла-Грейнджера оценивают регрессию:

$$X_{3,t} = \gamma_0 + \gamma_2 X_{2,t} + \gamma_3 X_{1,t} + e_t. \quad (10)$$

В результате оценивания параметров модели получили соотношение между параметрами в виде:

$$CH_t = -50.90 - 0.119 * H_t + 1.129 * ge_t + e_t \quad (11)$$

Ошибки оценивания (остатки регрессии) в полученной модели (11) должны быть независимыми, нормально распределенными случайными величинами с нулевым средним. Поэтому проверяют гипотезу о возможной нестационарности остатков, т.е. гипотезу о наличии единичного корня в последовательности $e(t)$. График остатков регрессии изображен на рис. 4, а результаты расширенного теста Дики-Фуллера представлены в табл. 6.

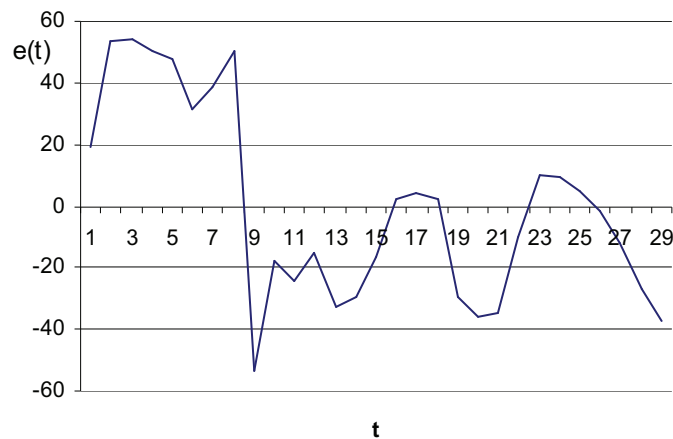


Рисунок 4- График остатков регрессии

Таблица 6

Результаты расширенного теста Дики – Фуллера (ошибок оценивания)

Параметр	t-статистика ADF-теста	Вероят.	Критические значения по Мак-Киннону	
e(t)	-3.006	0.0057	1%	-2.650145
			5%	-1.953381
			10%	-1.609798

Гипотеза о наличии единичного корня отвергнута на всех рассмотренных уровнях значимости, и, следовательно, последовательность остатков является стационарной.

Таким образом, выявлены причинно-следственные зависимости между рассматриваемыми параметрами ge , H , CH , что подтверждается физическими законами функционирования исследуемого карбюраторного двигателя внутреннего сгорания. Параметры находятся в динамическом взаимодействии, как следствие его нарушение позволит выявить отклонение режима функционирования объекта от номинального. Эмпирическая модель количественных отношений между параметрами является базовой для диагностирования состояния объекта по выделенным параметрам в реальном режиме времени с использованием известных алгоритмов оценивания.

ЛИТЕРАТУРА

1. Скурихин В.И. Информационные технологии в испытаниях сложных объектов / Скурихин В.И., Квачев В.Г., Валькман Ю.Р., Яковенко Л.П. // – К.: Наук. думка, 1990. – 320 с.
2. Alwan L.C., Roberts H.V. Time Series Modeling for Statistical Process Control / L.C. Alwan, H. V. Roberts. - Journal of Business and Economics Statistics, 1988, vol.6, N1. - P. 87-95.
3. Радченко С.Г. Устойчивые методы оценивания статистических моделей / С.Г. Радченко. – К.: ПП «Санспарель», 2005. – 504 с.
4. W. Enders. Applied Econometric Time Series / W. Enders. - New York: John Wiley & Sons, 1995 – 433 p.
5. Emmenegger J.-F. Cointegration Analysis of Ukrainian Time Series: Consumer Price Index, Wages and Food Prices in the Decade 1992-2001 / J.-F. Emmenegger, T.A. Bardadym // International Journal of Applied Mathematics, 2003. - Vol. 12. - No 2. - P.141-158.
6. Рыбалко В.В. Параметрическое диагностирование энергетических объектов на основе факторного анализа в среде Statistica / В.В. Рыбалко // Exponenta Pro, 2004. – С. 78 – 83.
7. Box G.E. Time Series Analysis, Forecasting and Control / G.E. Box., G.M. Jenkins, G.C. Reinsel. - New Jersey: Prentice Hall, 1994. – 598 p.
8. Hamilton J.D. Time Series Analysis / J.D. Hamilton. – Princeton: Princeton University Press, 1994. - 799 p.
9. Doan T.A. RATS Software Package, User’s Manual, Version 6.0. / T.A. Doan. - Illinois: ESTIMA, 2005. – 364 p.
10. Hansen H. CATS in RATS: Cointegration Analysis of Time Series, Handbook for the software package CATS / H. Hansen, K. Juselius. - Illinois: ESTIMA, 1995. – 155 p.

11. Engle R.E. Cointegration and Error Correction: Representation, Estimation and Testing / R.E. Engle, C.W.J. Granger // *Econometrica*, 1987. – Vol. 55. - P.251-276.
12. Johansen S. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models / S. Johansen // *Econometrica*, 1989 - Vol. 59. - N6. – P. 1551-1580.

Получено __.__.2006 г.