

УДК 004.942

**РЕГРЕССИОННО-ДИФФЕРЕНЦИАЛЬНАЯ МОДЕЛЬ РАЗВИТИЯ
КОЛИЧЕСТВА ПОЛЬЗОВАТЕЛЕЙ СЕТИ ИНТЕРНЕТ В РОССИИ**

А. Р. Срульдинов

*Магистрант,
Пермский национальный исследовательский
политехнический университет,
Березниковский филиал,
г. Березники, Пермский край, Россия*

**REGRESSION-DIFFERENTIAL MODEL
OF INTERNET USERS NUMBER DYNAMICS IN RUSSIA**

A. R. Srul'dinov

*Undergraduate student,
Perm National Research Polytechnic University,
Berezniki branch,
Berezniki, Perm Region, Russia*

Abstract. The object of the research is the Internet services in Russia. The work is about of building of multivariate linear and regression-differential models for the prediction of the object for a decision support. Four factors were choose for the model: the number of organizations with sites, expenses for communication and Informatics, number of graduates, population of Russia. These models were built according to the presented data: multivariate linear model and regression-differential models of 1st, 2nd and 3rd orders. As a result of the research multivariate linear model impossible to use for forecasting. Regression-differential model of the 1st order has a minimum error of prediction. The forecasting horizon for this model is 3 years. Dependences of the dynamics of the object from the factors in the model are described.

Keywords: Internet; regression-differential model; forecasting.

Прогноз – это научно обоснованное суждение о возможных состояниях объекта в будущем и (или) об альтернативных путях и сроках их осуществления [1].

Необходимость прогноза обусловлена желанием знать события будущего, что невозможно на 100 % в принципе, исходя из статистических, вероятностных, эмпирических, философских принципов. Точность любого прогноза обусловлена:

- объемом истинных (верифицированных) исходных данных и периодом их сбора;
- объемом неверифицированных исходных данных, периодом их сбора;
- свойствами системы, объекта, подвергающихся прогнозированию;
- методиками и подходами прогнозирования.

Одним из основных методов прогнозирования является метод моделирования. В настоящее время существует большое

количество видов моделирования, например [2]:

- Информационное
- Компьютерное
- Имитационное
- Статистическое

Для выполнения нашей задачи нам понадобится математическое моделирование. При математическом моделировании объект исследования заменяется его математической моделью. Математические модели бывают линейные (Линейная система – любая система, для которой отклик системы на сумму воздействий равен сумме откликов на каждое воздействие) и нелинейные [3].

Нашей задачей является построение линейной многофакторной и регрессионно – дифференциальной моделей для прогнозирования и поддержки принятия решений по управлению социально-

экономическими системами, а именно развитием сети Интернет в России.

За последние годы количество пользователей интернета в России резко увеличилось и продолжает неуклонно расти (рис. 1). Так, по данным ФОМ, на конец 2009 года, полугодовая аудитория интернета составляет 36,6 % российского населения в возрасте 18 лет и старше. За 2009 год проникновение интернета в России увеличилось более чем на 20 %. Как и в прошлом году, основной источник увеличения российской аудитории интернета – регионы. Как и раньше в крупных городах проникновение интернета заметно выше,

чем в среднем по России. По данным ФОМ, в городах-миллионниках интернетом пользуется 51 % населения старше 18 лет, а в селах не больше 20 % [4].

Так же интернет оказывает влияние на экономику страны. В 2009 году доля интернет-экономики составила 1,6 % ВВП или 19,3 миллиарда долларов [4].

Всем выше перечисленным мы можем обусловить выбор критерия, это количество пользователей интернета России (табл. 1). Данный критерий является одним из показателей развития сети Интернет в России в целом.

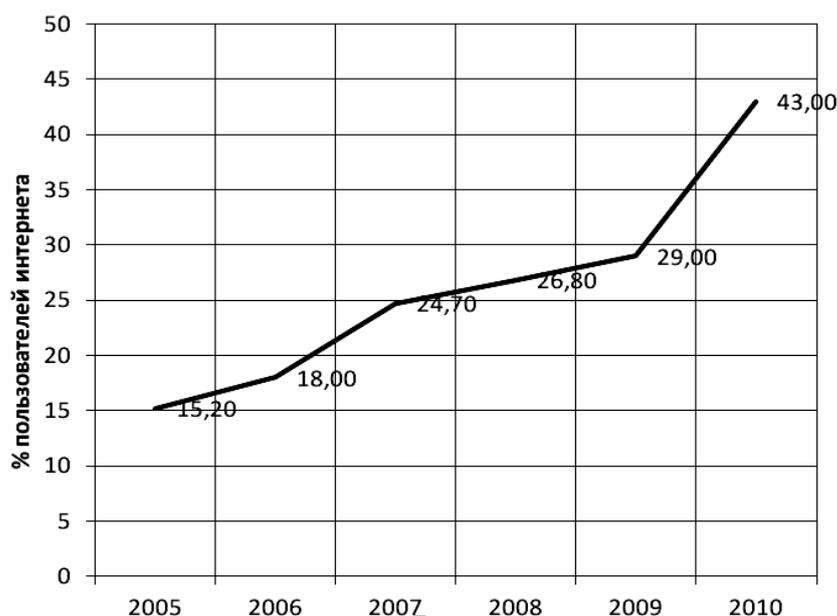


Рис. 1 – Доля интернет пользователей среди населения России

Таблица 4

Количество пользователей сети Интернет в России, млн. чел.

Период	2005	2006	2007	2008	2009	2010
Количество пользователей интернета, млн человек	21,75	25,65	35,10	38,06	41,15	61,23

Факторы могут быть управляемыми, которые поддаются изменению по решению системы управления объекта и неуправляемыми. Управляемые факто-

ры (x_1, x_2) и неуправляемые факторы модели (x_3, x_4) и их веса представлены в табл. 2.

На основании вышеизложенного построим линейную многофакторную модель социально-экономической системы и покажем что для прогноза данная модель не подходит.

$$y_{x,t} = a_0 + \sum_i a_i x_i t, \quad (1)$$

где a_0, a_i – константы, описывающие влияние одной реакции при построении тренда [5].

Таблица 5

Веса факторов

	2005	2006	2007	2008	2009	2010
Число организаций, имеющих WEB-сайты, единиц, x_1	22 338	34 104	33 626	38 733	41 967	50 355
Расходы: Связь и информатика - Федеральный бюджет, млрд руб., x_2	5,2	13,6	28,7	11,7	30,6	40,1
Выпуск учащихся средними (полными) общеобразовательными учреждениями, тыс. чел., x_3	1 312,10	1 213,80	1 105,30	959	810,4	719,6
Население России, млн чел., x_4	143,1	142,5	142,1	142	141,9	142,4

Построение линейной многофакторной модели производится по формуле (3) с применением методов градиентного или покоординатного спуска. Из табл. 3 видно, какое влияние оказывают факторы на

критерий. Например, на построение прогнозов на 1 и 2 года меньше всех влияет фактор x_3 (выпуск учащихся), больше всех влияет фактор x_4 (население России).

Таблица 3

Коэффициенты линейной модели

Прогноз	a	c_i			
		x_1	x_2	x_3	x_4
Исходный	0,1891	0,4451	0,2138	-0,3762	0,154
Прогноз 1 год	0,812	-0,3537	-0,0198	-0,3779	-0,4355
Прогноз 2 года	0,7645	-0,21	0,0885	-0,5661	-0,1997

Из рис. 2 видно, что для данного объекта линейная модель плохо приближает известные данные (погрешность аппроксимации 0,479), а также линейная модель

очень плохо прогнозирует данные, что видно на рис. 3. Отсюда делаем вывод что, для данного случая линейная – многофакторная модель плохо применима.

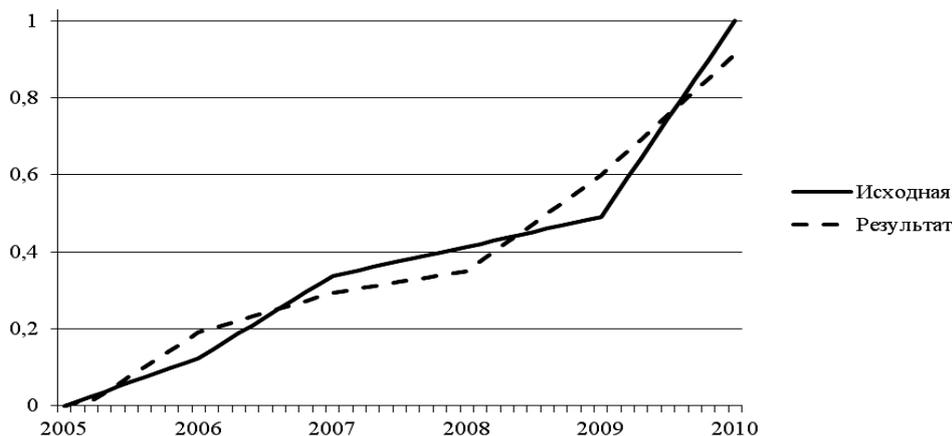


Рис. 2 – Линейное приближение

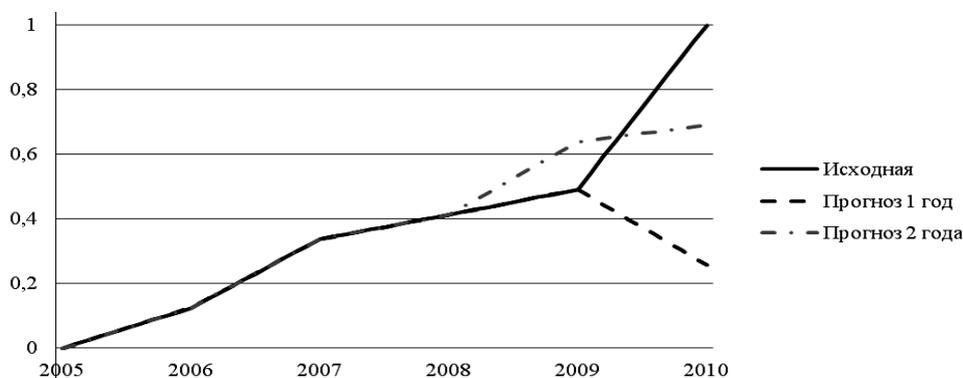


Рис. 3 – Постпрогноз линейной модели

Как мы видим, линейная – многофакторная модель плохо применима для прогнозирования данного объекта. В таких

случаях для прогнозирования используют регрессионно-дифференциальные модели (РДМ) вида:

$$\frac{d^n y}{dt^n} + \sum_{i=1}^{n-1} g_i \frac{d^i y}{dt^i} = a + b \cdot y(t - \tau_0) + \sum_{i=1}^m c_i \cdot x_i(t - \tau_i) + \sum_{i=1}^m \sum_{j=1}^m d_{ij} \cdot x_i(t - \tau_i) \cdot x_j(t - \tau_j) + \sum_{i=1}^m f_i \cdot [x_i(t - \tau_i)]^2, \quad (2)$$

где g_i – коэффициенты влияния младших производных, a – константа, описывающая влияние одной n -й производной реакции при построении тренда, b – коэффициент «обратной связи», описывающий влияние значения реакции на ее же n -ю производную, c_i – коэффициен-

ты влияния факторов, $d_{ij} : i \neq j$ – коэффициенты взаимного влияния факторов, $f_i \equiv d_{ii}$ – коэффициенты влияния квадратов факторов, τ_0 – запаздывание в обратной связи, τ_i – запаздывание воздействия i -го фактора. РДМ дополняется началь-

ными условиями $\frac{dy}{dt} \Big|_0 = y'_0, \frac{d^2y}{dt^2} \Big|_0 = y''_0, \dots$
 $\frac{d^{n-1}y}{dt^{n-1}} \Big|_0 = y_0^{(n-1)}$.

Неизвестными в данном случае являются все начальные условия, а также начальные условия $y'_0, \dots, y_0^{(n-1)}, a, b, c_i, d_{ij}, f_i, \tau_0, \tau_i$. Их поиск производится минимизацией квадратичного отклонения расчетного значения $y(t_k)$ от известных в отчетах значений критерия $y_{исх}(t_k)$

$$S = \sum_{k=0}^{K-1} (y(t_k) - y_{исх}(t_k))^2, \quad (3)$$

то есть решением задачи минимизации $y'_0, a, b, c_i, d_{ij}, f_i, \tau_0, \tau_i$:

$$:S(y'_0, a, b, c_i, d_{ij}, f_i, \tau_0, \tau_i) \rightarrow \min \quad (4)$$

Для начала выберем порядок ОДУ. Для первого приближения выбираем массивы коэффициентов ОДУ – b, c. Выбираем градиентный метод и рассчитаем РДМ 1, 2 и 3 порядков ОДУ. Результат показан на рис. 3.

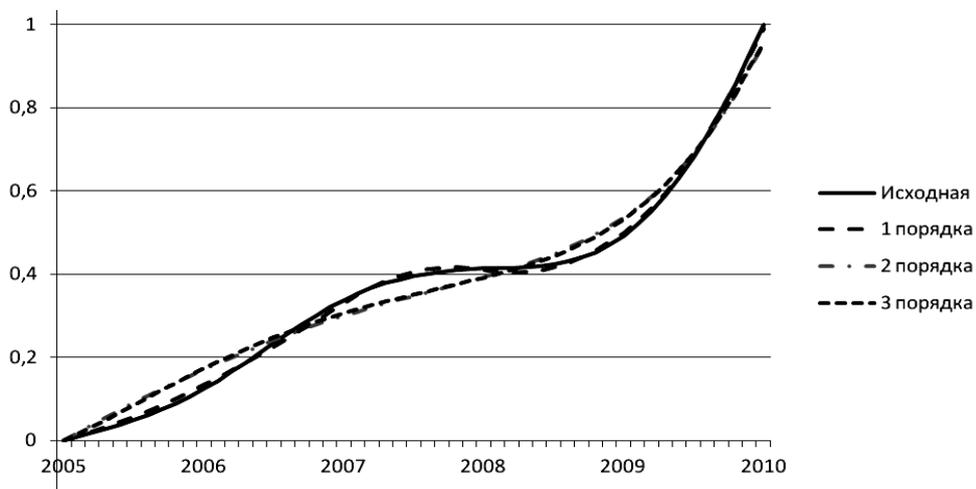


Рис. 1 – Результаты прогнозирования РДМ 1, 2, 3-го порядков

Погрешности аппроксимации (5) составили 0.0119, 0.2303, 0.2000 для 1, 2 и 3 порядков ОДУ, соответственно. Минимальную погрешность приближения для данного объекта показывает ОДУ 1-го порядка.

После того, как подобран порядок для РДМ модели, необходимо подобрать для всех факторов вид интерполяции.

- «левый» (указано значение в начале года, действующее до конца года; далее обозначен «0»);
- «правый» (указано значение в конце года, действовавшее весь год; далее обозначен «1»);
- линейный (указано значение в начале года, которое линейно изменяется до конца года; далее обозначен «Л»);
- фактор выключен (далее обозначен «X»).

Подбор интерполяции может быть осуществлен методами численного перебора. Полным перебором определен результат – «ЛЛЛЛ», что означает линейную интерполяцию для каждого из факторов.

Дальше будем по очереди отключать каждый фактор и повторять процедуру расчета для определения его значимости (табл. 4). Если погрешность прогнозирования изменяется существенно по сравнению с включенным коэффициентом, возможно, фактор можно оставить выключенным. После расчета отключенных факторов можно сделать вывод, что влияние факторов x_1, x_2 и x_4 является значительным. Но все-таки фактор x_3 оставим линейным, так как при таком положении достигается самая маленькая погрешность прогнозирования.

Таблица 4

Отключение факторов по одному

Вид интерполяции факторов	Погрешность прогнозирования
ХЛЛЛ	0,05005
ЛХЛЛ	0,25489
ЛЛХЛ	0,02726
ЛЛЛХ	0,30829

Так же проведем анализ с точки зрения взаимной корреляции, в котором вычисляются коэффициенты парной корреляции между факторами по формуле (7)

$$R_{xy} = \frac{\sum_i (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \cdot \sum_i (y_i - \bar{y})^2}} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sqrt{\sigma_x^2 \cdot \sigma_y^2}} \quad (7)$$

где средние значения $\bar{x}_i = \frac{1}{K} \sum_{k=1}^K x_i t_k$ и $\bar{x}_j = \frac{1}{K} \sum_{k=1}^K x_j t_k$, i, j – номера факторов. R_{xy}^2 – коэффициент детерминации. Если $R_{xy}^2 = 0.95$, то в 95% случаев изменения X приводят к изменению Y.

Проверка статистической значимости коэффициента корреляции производится по формуле (8):

$$t = |R_{xy}| \frac{\sqrt{n-2}}{\sqrt{1-R_{xy}^2}}, \quad (8)$$

где n – количество степеней свободы (в нашем случае - точек).

В случае, если $t > t_{\text{табл}}(n-2, \alpha)$, где α – уровень значимости, коэффициент корреляции статистически значим, и его можно использовать. Высокие значения коэффициентов парной корреляции выделяются на экране розовой заливкой, значения с недостаточным уровнем значимости – серой.

Мы видим, что наибольшую корреляцию со значением критерия имеют факторы x_1 и x_3 , а наименьшую – x_2 и x_4 (табл 5). Исходя из корреляции факторов, попробуем выключить одновременно факторы x_2, x_3 и x_1, x_4 (табл. б).

Таблица 5

Корреляция факторов линейной модели со значением критерия

x_1	x_2	x_3	x_4
0,93803	0,85781	-	-
		0,93062	0,44472

Таблица 6

Отключение факторов попарно

Положение факторов	Погрешность прогнозирования
ХЛХЛ	0,04941
ЛХЛХ	0,64589

Табл. 4 показывает, что влияние фактора x_3 незначительно. Из табл. 6 видно, что парное отключение факторов с наибольшей корреляцией в линейной модели также не значительно влияет на погрешность приближения.

Далее определим допустимый горизонт прогноза. Для этого возьмем известные данные и будем уменьшать количество лет до тех пор, пока погрешность прогнозирования будет отличаться боль-

ше, чем на 3 % от исходных данных. При прогнозировании 2 лет, по координатный спуск начинает расходиться, поэтому будем использовать градиентный спуск.

При прогнозировании 3 лет погрешность прогноза составляет 3,2 %. Следовательно, возможный горизонт прогнозирования при использовании этой модели составляет 2 года. Результат прогнозирования представлен на рис 4.

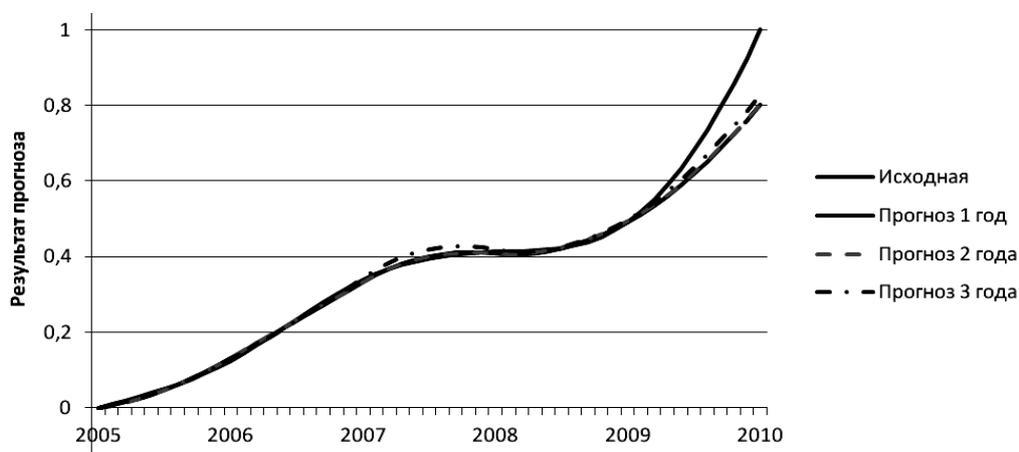


Рис. 2 – Определение горизонта прогнозирования

Следующим шагом для каждого фактора модели необходимо произвести изменение на $\pm 5\%$, при этом важно проконтролировать относительное изменение реакции системы. В результате, получится

тот или иной прогнозный сценарий направления развития объекта.

Попробуем спрогнозировать развитие объекта при малых отклонениях управляемых факторов x_1, x_2 в пределах $\pm 5\%$ (табл. 7).

Изменение управляемых факторов в прогнозе

	x_1		
x_2	-5%	0	+5%
-5%	118,76	119,51	120,26
0	120,79	121,54	122,29
+5%	122,82	123,57	124,32

Как видно из табл. 7, уменьшение факторов x_1 и x_2 ведет к снижению критерия и, наоборот, их повышение ведет к росту критерия, то есть критерий полностью зависит от данных факторов.

Из этого следует, что увеличение расходов на связь и информатику приводит к увеличению количества пользователей сети интернет, это логично, так как при этом развивается сетевая инфраструктура и расширяется зона обслуживания компаний-провайдеров, поэтому наблюдается прирост количества пользователей.

То же самое происходит и со вторым фактором, увеличивая количество компаний с WEB сайтами, мы получаем рост количества пользователей интернета, это так же не противоречит реальности, ведь компании сами являются пользователями интернета, а для поддержки собственного сайта необходимо интернет соединение.

Исходя из полученных данных, можно сказать, что динамика объекта положительная. Для управления данным объектом, лицу принимающему решение, достаточно повлиять хотя бы на один из управляемых факторов, чтобы добиться роста количества пользователей или наоборот его снижения. Так, например сокращение расходов на связь и информатику может сначала замедлить рост количества пользователей или совсем его остановить, а в дальнейшем может привести и к сокращению числа пользователей сети.

Библиографический список

1. Прогноз // Википедия. URL: <https://ru.wikipedia.org/wiki/Прогноз>
2. Моделирование // Википедия. URL: <https://ru.wikipedia.org/wiki/Моделирование>
3. Линейная система // Википедия. URL: https://ru.wikipedia.org/wiki/Линейная_система
4. Доля интернет пользователей среди населения – Google URL: https://www.google.ru/publicdata/explore?ds=d5bncppjof8f9_&met_y=it_net_user_p2&idim=country:RUS&hl=ru&dl=ru#!ctype=l&strail=false&bcs=d&nselm=h&met_y=it_net_user_p2&scale_y=lin&ind_y=false&rdim=region&idim=country:RUS&ifdim=region&tstart=1117479600000&tend=1275246000000&hl=ru&dl=ru&ind=false
5. Янченко Т. В., Сиротина Н. А., Затонский А. В. Об аппроксимации факторов дифференциальной модели социально – экономической системы // Современные исследования социальных проблем. – 2012. – № 11(19). – С. 6.
6. Янченко Т. В. Обоснование порядка регрессионно-дифференциальной модели краевого социального ресурса // Системы управления и информационные технологии. – 2014. – № 3.1(57). – С. 187–191.

Bibliograficheskiy spisok

1. Prognoz // Vikipedija. URL: <https://ru.wikipedia.org/wiki/Prognoz>
2. Modelirovanie // Vikipedija. URL: <https://ru.wikipedia.org/wiki/Modelirovanie>
3. Linejnaja sistema // Vikipedija. URL: https://ru.wikipedia.org/wiki/Linejnaja_sistema
4. Dolja internet pol'zovatelej sredi naselenija – Google URL: https://www.google.ru/publicdata/explore?ds=d5bncppjof8f9_&met_y=it_net_user_p2&idim=country:RUS&hl=ru&dl=ru#!ctype=l&strail=false&bcs=d&nselm=h&met_y=it_net_user_p2&scale_y=lin&ind_y=false&rdim=region&idim=country:RUS&ifdim=region&tstart=1117479600000&tend=1275246000000&hl=ru&dl=ru&ind=false
5. Janchenko T. V., Siroтина N. A., Zaton'skij A. V. Ob approksimacii faktorov



diferencial'noj modeli social'no – jekonomich-
eskoj sistemy // Sovremennye issledovanija so-
cial'nyh problem. – 2012. – № 11(19). – S. 6.

5. Janchenko T. V. Obosnovanie porjadka regres-
sionno-diferencial'noj modeli kraevogo so-
cial'nogo resursa // Sistemy upravlenija i infor-
macionnye tehnologii. – 2014. – № 3.1(57). – S.
187–191.

© Срульдинов А. Р., 2016

