

УДК 519.233.5, 519.254

О.И. Синельникова<sup>1</sup>, Л.В. Перепелкина<sup>2</sup>ХНУРЭ, г.Харьков, Украина, <sup>1</sup>ol.sinelnikova@gmail.com, <sup>2</sup>ludapv@mail.ru

## МОДЕЛИ И МЕТОДЫ ПРИ МНОГОФАКТОРНОМ МОДЕЛИРОВАНИИ ЦЕНЫ НА НЕФТЬ

В статье исследуются модели и методы многофакторного моделирования цены на нефть. В ходе численных экспериментов на двух классах моделей подтвердилась целесообразность включения в модель факторов, отражающих резонанс в СМИ. Во всех моделях была учтена упреждающая зависимость, рассчитанная с помощью кросскорреляции. Также в ходе экспериментов был проведен сравнительный анализ приемлемости этих моделей для поставленной задачи, в результате которого, квадратичная модель оказалась адекватнее линейной.

**МНОГОФАКТОРНОЕ МОДЕЛИРОВАНИЕ, МЕТОД ГРУППОВОГО УЧЕТА АРГУМЕНТОВ, КРОССКОРРЕЛЯЦИЯ, МАКРОЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ, РЕЗОНАНС В СМИ.**

### Введение

Нефть — это сырье, одно из наиболее стратегически важных для подавляющего числа государств. Модель ценообразования нефти интересует не только участников торгов на фондовых рынках или инвесторов и акционеров в нефтедобывающем секторе [1],[2]. Планирование бюджета государства невозможно представить без учета цен на нефть так же, как и составление долгосрочных бизнес-проектов в любой отрасли: будь то энергетика или сельское хозяйство, металлургия или легкая промышленность, не говоря уже о химической, в которой полимеры давно завоевали лидирующие позиции.

И во всех этих сферах деятельности менеджерам крупного звена необходимо принимать решения с учетом котировок на нефть, что делает область многофакторного моделирования цены на нефть актуальной для исследований.

Ранее уже были предложены модели подобного рода [2], однако они включали в себя данные, собранные до 2008 г., когда наблюдался относительно устойчивый положительный тренд, и не включали данные стремительного падения цены на нефть во второй половине 2008 г. Также период, в котором были построены модели зависимости, был поделен на две части, что может говорить о неуниверсальности модели.

В данной работе предполагается, что для построения моделей цены на нефть следует учитывать такие показатели, как: индексы производства, задействованных мощностей производства, деловой активности, экономической уверенности крупных производителей, уверенности потребителей; объемы экспорта, импорта, промышленного производства, розничных продаж; внешнеторговый оборот и оборот розничной торговли.

Также предполагается получить модель, которая отображает динамику роста цены на нефть с учетом анализа резонанса в прессе, телевидении и на интернет ресурсах о нефти. В качестве факторов было взято количество появившихся в интернете статей, проиндексированных в поисковой системе Google

за каждый месяц по запросам: «нефть», «цена на нефть», «Ближний восток», «Россия», «oil price», «middle east», «кризис», «цена на нефть» (по России). Такой подход к построению модели позволит учесть влияние «шумихи» в прессе на цену нефти [3].

Для более детального анализа проблемы предлагается получить модель, которая учитывала бы обе группы факторов — макроэкономические показатели и резонанс в СМИ.

Выбранные факторы могут влиять на цену нефти не прямо, а с некоторым упреждением, поэтому при построении многофакторной модели будет учтен временной лаг [4].

### 1. Формальная постановка

Целью данной работы является проведение сравнительного анализа моделей, полученных на различных классах, для выявления модели наиболее адекватно отображающей зависимость цены на нефть от независимых факторов — макроэкономических показателей, а также от факторов, отображающих резонанс в СМИ.

Приведем формальную постановку задачи исследования. В данной работе предлагается рассматривать задачу многофакторного моделирования макроэкономических показателей государства на примере модели цены на нефть, а также сравнительный анализ линейных и нелинейных моделей в экономическом моделировании.

Пусть исходная величина, курс гривны по отношению к доллару, задана в виде временного ряда:  $Y = \{Y(t), Y(t-1), \dots, Y(t-T)\}$ , длиной  $T$ .

Исходя из вышесказанного, приведем формальную постановку задачи определения функциональной зависимости между независимыми факторами и ценой на нефть:

$$\hat{Y}(t+1) = F(x_1(t), \dots, x_1(t-\tau_1), x_2(t), \dots, x_2(t-\tau_2), \dots, x_n(t), \dots, x_n(t-\tau_n), S, \bar{a}) + \varepsilon(t), \quad (1)$$

где:

—  $F_S(\cdot)$  — модель, описывающая связь между значением ВВП в момент времени  $t$  и значениями

независимых факторов, определенная методами многофакторного моделирования;

- $S$  – структура модели;
- $\bar{a}$  – вектор параметров модели;
- $\tau_1, \tau_2, \dots, \tau_N$  – временные лаги для соответствующих переменных;
- $x_1, x_2, \dots, x_N$  – статистические значения макроэкономических показателей, которые являются независимыми переменными.

К независимым факторам, которые целесообразно рассматривать, можно отнести следующие:

- $x_1$  – индекс производства;
- $x_2$  – объем экспорта;
- $x_3$  – внешнеторговый оборот;
- $x_4$  – оборот розничной торговли;
- $x_5$  – задействованные мощности производства;
- $x_6$  – индекс деловой активности;
- $x_7$  – объем розничных продаж;
- $x_8$  – объем промышленного производства;
- $x_9$  – объем импорта;
- $x_{10}$  – индекс уверенности крупных производителей;
- $x_{11}$  – индекс уверенности потребителя;
- $x_{12}$  – количество новостей в интернете по запросу «нефть»;

–  $x_{13}$  – количество новостей в интернете по запросу «цена на нефть»;

–  $x_{14}$  – количество новостей в интернете по запросу «Ближний восток»;

–  $x_{15}$  – количество новостей в интернете по запросу «Россия»;

–  $x_{16}$  – количество новостей в интернете по запросу «oil price»;

–  $x_{17}$  – количество новостей в интернете по запросу «middle east»;

–  $x_{18}$  – количество новостей в интернете по запросу «Кризис»;

–  $x_{19}$  – количество новостей в интернете по запросу «Цена на нефть» по России.

Далее введем обозначение  $f_S(\bar{a})$ , как функции некоторой модели  $S$  с параметрами  $\bar{a}$ , то есть

$$f_S(\bar{a}) = F(x_1(t), \dots, x_1(t - \tau_1), x_2(t), \dots, \dots, x_2(t - \tau_2), \dots, x_N(t), \dots, x_N(t - \tau_N), S, \bar{a}).$$

В данной работе зависимость  $F_S(\cdot)$  в модели (1) будем рассматривать на классе моделей:

– линейные модели, полученные методом группового учета аргументов [5] в виде обобщенных полиномов Колмогорова-Габора:

$$\hat{Y}(t+1) = a_0 + \sum_{i=1}^N a_i x_i(t) + \sum_{i=1}^N \sum_{j=1}^N a_{ij} x_i(t) x_j(t) + \dots + \varepsilon(t)$$

с частичными описаниями вида:

$$f_{i,k}(t) = a_0 + a_i x_i(t) + a_k x_k(t);$$

– полиномиальные модели, полученные методом группового учета аргументов в виде

обобщенных полиномов Колмогорова-Габора:

$$\hat{Y}(t+1) = a_0 + \sum_{i=1}^N a_i x_i(t) + \sum_{i=1}^N \sum_{j=1}^N a_{ij} x_i(t) x_j(t) + \dots + \varepsilon(t)$$

с частичными описаниями вида:

$$f_{i,k}(t) = a_0 + a_i x_i(t) + a_k x_k(t) + a_{i,k} x_i(t) x_k(t) + a_{i,i} x_i^2(t) + a_{k,k} x_k^2(t).$$

Структура и параметры модели (1) должны удовлетворять следующему векторному критерию:

$$\begin{aligned} \sum_{t=1}^T (Y(t) - \hat{Y}(t))^2 \rightarrow \min \min_S a \\ R^2(f_S(\bar{a})) \rightarrow \max \max_S a \\ S(f_S(\bar{a})) \rightarrow \min \min_S a \end{aligned} \quad (2)$$

где  $R^2(f_S(\bar{a}))$  – коэффициент множественной детерминации для модели  $f_S(\bar{a})$ ;  $S(f_S(\bar{a}))$  – оценка среднеквадратического отклонения для ряда ошибок  $e(t)$ , полученных по модели  $f_S(\bar{a})$ , то есть:

$$e(t) = Y(t) - f_S(\bar{a}).$$

Для определения  $\tau_1, \tau_2, \dots, \tau_N$  – временных лагов для соответствующих переменных, используется кросс-корреляционная функция вида:

$$r_{x,y}[k] = \frac{\sum_{i=1}^{n-k} y_i x_{i+k} - \frac{\sum_{i=1}^{n-k} y_i \sum_{i=1}^{n-k} x_{i+k}}{n-k}}{\sqrt{\left( \sum_{i=1}^{n-k} y_i^2 - \frac{\left( \sum_{i=1}^{n-k} y_i \right)^2}{(n-k)} \right) \left( \sum_{i=k+1}^n x_i^2 - \frac{\left( \sum_{i=1}^{n-k} x_i \right)^2}{(n-k)} \right)}}, \quad (3)$$

где  $k$  – лаг, на который смещаются ряды;  $n$  – общее количество элементов;  $x, y$  два ряда, для которых рассчитывается кросс-корреляция.

Лаг  $\tau$  выбирается как:

$$\tau_i = k_0^i; k_0^i \in K; r_{k_0}^i = \max_k |r_k^i| \vee |r_{k_0}^i| > |\sigma(r_k)|;$$

$$\sigma(r_k) = \sqrt{\frac{1 + 2 \sum_{j=1}^{k-1} r_j^2}{K}}.$$

При построении нелинейной модели с учетом дисперсионного анализа подсчитывается коэффициент детерминации по формуле:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\bar{Y}_i - Y_i)^2}{\sum_{i=1}^n (Y_S - Y_i)^2}, \quad (4)$$

где  $\bar{Y}_i$  – полученные значения регрессии;  $Y_i$  – исходные значения;  $n$  – количество элементов в выборке;  $Y_S$  – среднее значение исходных данных. Величина  $R^2$  – коэффициент детерминации, который пока-

зывает, какая доля дисперсии функции объясняется изменениями входящих в уравнение регрессии независимых переменных при полученных значениях коэффициентов модели.

Решение задачи (2) по сути распадается на два отдельных этапа: идентификация структуры модели; параметрическая идентификация модели.

### 2. Выбор и обоснование метода

Метод группового учета аргументов (МГУА) использует идеи самоорганизации и механизмы живой природы – скрещивание (гибридизацию) и селекцию (отбор).

Наиболее полная зависимость между входами  $X_i$  и выходами  $Y_i$  может быть представлена с помощью обобщенного полинома Колмогорова-Габоара.

Пусть есть  $X = (x_1, \dots, x_N)$ , тогда такой полином имеет вид:

$$\hat{Y} = a_0 + \sum_{i=1}^N a_i x_i + \sum_{i=1}^N \sum_{j \leq i} a_{i,j} x_i x_j + \dots + \sum_{i=1}^N \sum_{j \leq i} \sum_{k \leq j} a_{i,j,k} x_i x_j x_k + \dots$$

с частичными описаниями вида:

$$f_{i,k}(t) = a_0 + a_i x_i(t) + a_k x_k(t),$$

$$f_{i,k}(t) = a_0 + a_i x_i(t) + a_k x_k(t) + a_{i,k} x_i(t) x_k(t) + a_{i,i} x_i^2(t) + a_{k,k} x_k^2(t),$$

где все коэффициенты  $a$  неизвестны.

При построении модели (при определении значений коэффициентов) в качестве критерия используется критерий регулярности (точности):

$$\bar{\epsilon}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - F(x_i))^2;$$

$$\bar{\epsilon}^2 \rightarrow \min .$$

### 3. Анализ численного эксперимента

#### 3.1 Получение линейной модели, учитывающей макроэкономические показатели по России

В результате аппроксимации линейными полиномами на третьем слое была получена модель цены на нефть, в которую вошли такие факторы:

- $x_1$  – индекс производства;
- $x_2$  – объем экспорта;
- $x_3$  – внешнеторговый оборот;
- $x_4$  – оборот розничной торговли;
- $x_5$  – задействованные мощности производства;
- $x_6$  – индекс деловой активности;
- $x_7$  – объем розничных продаж;
- $x_8$  – объем промышленного производства;
- $x_9$  – объем импорта.

Сравнительные графики аппроксимации представлены на рис. 1.



Рис. 1. График фактического значения цены на нефть и ее линейной модели зависимости от макроэкономических показателей

#### 3.2 Получение линейной модели, учитывающей резонанс в СМИ

В результате аппроксимации линейными полиномами на третьем слое была получена модель цены на нефть, учитывающая резонанс в СМИ, в которую вошли такие факторы:

- $x_2$  – объем экспорта;
- $x_3$  – внешнеторговый оборот;
- $x_4$  – оборот розничной торговли;
- $x_5$  – задействованные мощности производства;
- $x_6$  – индекс деловой активности;
- $x_7$  – объем розничных продаж;
- $x_8$  – объем промышленного производства;
- $x_9$  – объем импорта;
- $x_{12}$  – количество новостей в интернете по запросу «нефть»;
- $x_{16}$  – количество новостей в интернете по запросу «oil price».

Сравнительные графики аппроксимации представлены на рис. 2.



Рис. 2. График фактического значения цены на нефть и ее линейной модели зависимости от макроэкономических показателей с учетом резонанса в СМИ

#### 3.3 Получение нелинейной модели, учитывающей макроэкономические показатели

В результате аппроксимации квадратичными полиномами была получена на пятом слое модель цены на нефть, в которую вошли такие факторы:

- $x_1$  – индекс производства;
- $x_2$  – объем экспорта;
- $x_3$  – внешнеторговый оборот;

- $x_4$  – оборот розничной торговли;
- $x_5$  – задействованные мощности производства;
- $x_6$  – индекс деловой активности;
- $x_7$  – объем розничных продаж;
- $x_8$  – объем промышленного производства;
- $x_9$  – объем импорта.

Сравнительные графики аппроксимации представлены на рис. 3.



Рис. 3. График фактического значения цены на нефть и ее нелинейной модели зависимости от макроэкономических показателей

### 3.4 Получение нелинейной модели, учитывающей резонанс в СМИ

В результате аппроксимации квадратичными полиномами была получена на пятом слое модель цены на нефть, учитывающая резонанс в СМИ, в которую вошли такие факторы:

- $x_1$  – индекс производства;
- $x_2$  – объем экспорта;
- $x_3$  – внешнеторговый оборот;
- $x_5$  – задействованные мощности производства;
- $x_6$  – индекс деловой активности;
- $x_7$  – объем розничных продаж;
- $x_8$  – объем промышленного производства;
- $x_9$  – объем импорта;
- $x_{11}$  – индекс уверенности потребителя;
- $x_{13}$  – количество новостей в интернете по запросу «цена на нефть».

Сравнительные графики аппроксимации представлены на рис. 4.



Рис. 4. График фактического значения цены на нефть и ее нелинейной модели зависимости от макроэкономических показателей с учетом резонанса в СМИ

### 3.5 Сравнительный анализ полученных моделей

Статистические свойства ошибок линейных моделей представлены в табл. 1.

Таблица 1

Статистические свойства ошибок линейных моделей

Модель	Линейная модель (факторы)	Линейная модель (новости)
Среднее значение	9,3131	8,2266
Медиана	8,2798	7,8994
СК отклонение	16,0991	14,9307
Min значение	-47,9401	-54,2372
Max значение	53,3807	46,3760

Статистические свойства ошибок квадратичных моделей представлены в табл. 2.

Таблица 2

Статистические свойства ошибок нелинейных моделей

Модель	Нелинейная модель (факторы)	Нелинейная модель (новости)
Среднее значение	5,6797	5,2497
Медиана	-0,5123	-0,6260
СК отклонение	8,2798	7,8994
Min значение	-27,8756	-28,3662
Max значение	19,8046	19,2112

### Выводы

Как видно из статистических свойств ошибок линейных моделей, несмотря на ее адекватность, модель менее приемлема для моделирования цен на нефть в реальном времени, так как дает довольно большую погрешность.

В то же время, квадратичная модель имеет приблизительно в два раза меньшую погрешность, то есть является более адекватной. Следовательно, ее использование для моделирования цены на нефть в реальном времени целесообразнее.

Данная работа была проведена в качестве одного из этапов комплексного исследования по разработке метода, который по сути является надстройкой над существующими подходами многофакторного моделирования. Разрабатываемый метод призван учитывать как можно более полную зависимость при многофакторном моделировании (сезонные составные, тренды, линейные и нелинейные зависимости, упреждающие зависимости и т.д.). Рассмотренный пример показал высокую значимость факторов, отражающих резонанс в масс-медиа: эксперименты на обоих классах моделей доказали преимущества модели, которая учитывает не только макроэкономические показатели, но и резонанс в СМИ.

Следующим этапом исследования будет разработка метода многофакторного моделирования отдельных компонентов зависимой переменной на основе метода «Гусеница».

**Список литературы:** 1. Брагинский, О. Б. Цены на нефть: история, прогноз, влияние на экономику [Текст] / О. Б. Брагинский // Журнал. Рос. хим. об-ва им. Д. И. Менделеева. – 2008. – Т. LII, № 6. 2. Варшавский, Л. Е. Моделирование динамики цены на нефть при разных режимах развития рынка нефти [Текст] / Л. Е. Варшавский // Прикладная эконометрия. – 2009. – № 1(13). – С. 70–88. 3. Boardman, J. Systems Thinking: Coping with 21st Century Problems (Industrial Innovation) [Текст] / J. Boardman // CRC Press. – 2008. – 240 p. 4. Елисеева, И. И. Общая теория статистики [Текст] / И. И. Елисеева, М. М. Юзбашев // М: «Финансы и статистика». – 2004. – 657 с. 5. Ивахненко, А. Г. Моделирование сложных систем по экспериментальным данным [Текст] / А. Г. Ивахненко, Ю. П. Юрачковский. – М.: «Радио и связь». – 1987. – 119 с.

*Поступила в редколлегию 28.08.2012*

УДК 519.233.5, 519.254

**Моделі та методи при багатофакторному моделюванні ціни на нафту** / О.І. Синельнікова, Л.В. Перепьолкіна // Біоніка інтелекту: наук.-техн. журнал. – 2012. – № 2 (79). – С. 28–32.

У роботі розглянуто результати аналізу застосування лінійних та нелінійних моделей при багатофакторному моделюванні ціни на нафту. Також розглянуто доцільність врахування не тільки макроекономічних факторів, але й факторів, що відображають резонанс у ЗМІ. Для відображення упереджувальної залежності ціни на нафту від незалежних факторів було враховано часовий лаг, визначений за допомогою кроскореляції.

Табл .2. Іл .4. Бібліогр.: 5 найм .

UDC 519.233.5, 519.254

**Models and methods in multivariate modeling oil price** / O.I. Sinelnikova, L.V. Perepolkina // Bionics of Intelligence: Sci. Mag. – 2012. – № 2 (79). – P. 28–32.

The article deals with the analysis of linear and nonlinear models for multivariate modeling of the oil price. Expediency of taking into model not only the macro-economic factors, but also the factors that reflect media coverage has also been examined. To reflect the proactive depending on oil prices from independent factors the calculated by cross-correlation time lag has been taken into account.

Tab. 2. Fig. 4. Ref.: 5 items.