

ПРИМЕНЕНИЕ КОРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА ПРИ ДОЛГОСРОЧНОМ ПРОГНОЗИРОВАНИИ СПРОСА НА ГАЗОХИМИЧЕСКУЮ ПРОДУКЦИЮ

*А.В. Хорев, С.В. Семенова, И.В. Кропотова, А.В. Оскирко
(ООО «Газпром ВНИИГАЗ»)*

В настоящее время для производственных объединений в России маркетинговое управление деятельностью становится одним из условий успешного функционирования бизнеса. Обеспечение эффективности такого управления требует умения предвидеть вероятное будущее состояние предприятия и среды, в которой оно существует.

Перспективное освоение новых запасов углеводородного сырья связано с организацией на их базе новых газоперерабатывающих и газохимических производств, с перспективами сбыта готовой продукции этих предприятий. Для российских газодобывающих актуально стоит вопрос реализации ресурсного потенциала Дальнего Востока и Восточной Сибири, где располагаются крупные углеводородные запасы. Решение этой задачи должно быть увязано с долгосрочными прогнозами спроса на продукцию.

Согласно фундаментальному закону рыночной экономики, спрос и предложение какого-либо вида продукции зависят друг от друга. Другими словами, спрос рождает предложение. Прогнозирование спроса на определенный вид продукции имеет важное значение при планировании и организации нового производственного объекта, на котором будет производиться эта продукция. Прогнозирование объемов потребления продукции помогает определить объемы производственных мощностей, которые будут введены в строй при организации того или иного объекта, а также направления реализации продукции.

В настоящее время, по оценкам ученых, насчитывается свыше 150 различных методов прогнозирования, однако на практике используются в качестве основных всего 15–20.

Методы прогнозирования можно классифицировать как эвристические, при применении которых преобладают субъективные начала, и как экономико-математические, при применении которых преобладают объективные начала [1, 2]. В свою очередь экономико-математические методы включают: прикладную статистику, математическую экономику и эконометрику, методы принятия оптимальных решений, экономико-математическое моделирование для централизованно планируемой экономики, экономико-математическое моделирование для конкурентной экономики, экономическую кибернетику, экспериментальную экономику.

Эвристические методы предполагают, что подходы, используемые для формирования прогноза, не изложены в явной форме и неотделимы от лица, делающего прогноз, при разработке которого доминируют интуиция, прежний опыт, логика, творчество и воображение. К данной категории методов относятся методы социологических исследований и экспертные методы.

Экономико-математические методы основаны на комплексе экономических и математических научных дисциплин, которые позволяют определить взаимосвязи факторных и результативных показателей. Факторным называют независимый признак, результативным – зависимый. На этой базе создается математическая модель, с помощью которой с определенной степенью вероятности можно спрогнозировать результативные показатели на основе прогнозных значений факторных показателей [1, 2].

Среди всего множества экономико-математических методов корреляционно-регрессионный анализ является основным методом в изучении взаимосвязей явлений и построения математических уравнений, которые описывают эти взаимосвязи. Корреляционно-регрессионный анализ относится к методам прикладной статистики и состоит из двух частей – корреляционный анализ и регрессионный анализ. Корреляционный анализ – это количественный метод определения тесноты и направления взаимосвязи между выборочными переменными величинами. Регрессионный анализ – это количественный метод определения вида математической функции в причинно-следственной зависимости между переменными величинами.

Применение корреляционно-регрессионного анализа при прогнозировании спроса на газохимическую продукцию представлено ниже на примере прогноза потребления полиэтилена высокой плотности (ПЭВП) на российском рынке.

В ходе применения данного метода прогнозирования были решены следующие задачи:

- выделены и проанализированы наиболее существенные факторы, влияющие на уровень потребления полиэтилена высокой плотности на российском рынке;
- определена форма зависимости между факторами и результативными показателями (потребление ПЭВП);
- смоделирована взаимосвязь между результативными и факторными показателями;
- проведена проверка значимости разработанной модели.

Сложный характер и взаимосвязь процессов, протекающих в экономике страны, ставят задачу отбора наиболее существенных факторов, оказывающих влияние на уровень потребления газохимической продукции. Таких факторов достаточно много ввиду усложнения и неоднозначности экономической динамики. Наиболее значимые из них представлены в табл. 1 [3, 4].

Таблица 1

Основные факторы, влияющие на потребление продукции газопереработки

№ п.п.	Наименование	Характеристика
<i>1</i>	<i>Экономические факторы</i>	
1.1	Внутренний валовый продукт	ВВП является основным показателем экономического развития страны, который отражает масштабы национального производства товаров и услуг в данный период
1.2	Цена товара	Цена является одной из основных характеристик товара и по закону спроса регулирует объемы спроса и предложения этого товара на рынке
1.3	Цены на товары-заместители	Цены на взаимозаменяемые товары оказывают влияние на потребление других взаимозаменяемых товаров посредством перехода части потребителей на более дешевый аналог
1.4	Доходы покупателей	Уровень доходов покупателей является рыночным фактором, оказывающим прямое воздействие на объем потребления товара
1.5	Кредитно-денежная политика государства (изменение учетной ставки рефинансирования ЦБ РФ)	Кредитно-денежная политика является одним из главных факторов, влияющих на уровень спроса на товар посредством регулирования учетной ставки Центрального банка Российской Федерации, согласно которой ЦБ предоставляет кредиты другим коммерческим банкам, которые в свою очередь выдают кредиты основным потребителям
1.6	Инфляционные ожидания	На базе инфляционных ожиданий производители и потребители строят свою кредитно-финансовую и ценовую политику, что в свою очередь оказывает влияние на производство и потребление товаров
<i>2</i>	<i>Неэкономические факторы</i>	
2.1	Вкусы и предпочтения покупателей	Связь вкусов и предпочтений потребителей заключается в желании приобрести тот или иной товар
2.2	Политические действия	Изменение налогового законодательства, введение акцизов, пошлин, заградительных мер по ввозу либо вывозу товара из страны может оказать влияние на уровень спроса на товар
2.3	Форс-мажорные обстоятельства	Природные катастрофы, войны, разрушая экономику, влияют на уровень спроса и предложения товаров

В табл. 1 рассмотрены экономические, социальные, политические и форс-мажорные факторы, оказывающие влияние как на экономику страны в целом, так и на потребление газохимической продукции в частности. Экономические факторы характеризуются определенными функциональными зависимостями, согласно которым они изменяются. Таким образом, расчет их точных значений производится путем выполнения математических операций в рамках законов математики, экономики и других наук. В свою очередь изменение неэкономических факторов не подчиняется общепризнанным фундаментальным законам, а лежит в области психологии и социальных взаимоотношений людей, что делает трудновыполнимым точный расчет вероятности возникновения этих событий в долговременной перспективе. Принимая во внимание вышеизложенное, при разработке методики учитывались только экономические факторы, значения которых за 2000–2009 гг. представлены в табл. 2 [5].

Таблица 2

Значения основных факторов, влияющих на потребление продукции газопереработки в РФ за 2000–2009 гг.

Год	Потребление ПЭВП*, тыс. т, Y	ВВП, млрд руб., X1	Цена ПЭВП, руб./т, X2	Цена ПЭНП**, руб./т, X3	Цена ПП***, руб./т, X4	Учетная ставка ЦБ РФ, %, X6	Инфляция, %, X7	Среднедушевые денежные доходы руб., X8
2000	300	7306	18031	19474	17195	37,3	20,2	2281
2001	330	7678	18437	20129	17850	25	18,6	3062
2002	305	8039	17869	19468	17885	23	15,1	3947
2003	377	8626	19503	20835	22041	18,3	12	5170
2004	452	9247	26914	25226	23922	14,3	11,7	6410
2005	548	9839	33561	40990	31098	12,5	10,9	8112
2006	646	10596	38992	36062	34934	11,5	9	10196
2007	810	11455	40560	41538	33669	10,5	11,9	12603
2008	781	12096	45060	42582	39655	26,2	13,3	14943
2009	745	11068	42997	31391	30254	10,7	8,1	16887

* Полиэтилен высокой плотности.

** Полиэтилен низкой плотности.

*** Полипропилен.

Источники: открытые материалы конференции «Полиэтилен 2011 г.», «Полипропилен 2011 г.», данные Минпромэнерго России, Министерства экономического развития и торговли.

На основании представленных в табл. 2 основных факторов, влияющих на потребление продукции газопереработки в РФ, была составлена матрица парных коэффициентов корреляции (табл. 3) [6].

Таблица 3

Матрица парных коэффициентов корреляции

Наименование	Потребление ПЭВП	ВВП	Цена ПЭВП	Цена ПЭНП	Цена ПП	Учетная ставка ЦБ РФ	Инфляция	Среднедушевые денежные доходы
Потребление ПЭВП	1							
ВВП	0,98	1						
Цена ПЭВП	0,97	0,97	1					
Цена ПЭНП	0,87	0,88	0,94	1				
Цена ПП	0,90	0,93	0,97	0,97	1			
Учетная ставка ЦБ РФ	-0,57	-0,57	-0,46	-0,41	-0,40	1		
Инфляция	-0,67	-0,72	-0,59	-0,48	-0,54	0,90	1	
Среднедушевые денежные доходы	0,95	0,95	0,90	0,73	0,79	-0,54	-0,72	1

Анализ табл. 3 показывает, что наиболее существенными факторами, влияющим на изменение уровня потребления ПЭВП в РФ, являются ВВП, цена ПЭВП, цена ПЭНП, цена ПП, инфляция и среднедушевые денежные доходы. Остальные факторы в меньшей степени влияют на потребление ПЭВП. Также следует отметить, что между наиболее существенными факторами существует взаимосвязь, то есть наблюдается мультиколлинеарность, которая характеризуется сильной корреляционной связью [6]. Мультиколлинеарность вызывает неустойчивость работы модели прогноза и существенно затрудняет оценку ее статистической значимости. Так, между ВВП и остальными существенными факторами наблюдается мультиколлинеарность (коэффициент корреляции колеблется в пределах от 0,72 до 0,97). Для борьбы с мультиколлинеарностью используется метод исключения сильно коррелированных переменных. В связи с этим исключаем из дальнейшего рассмотрения факторы, наименее коррелированные с потреблением ПЭВП. Таким образом, дальнейшее построение модели прогнозирования уровня потребления ПЭВП будет основываться на изменении ВВП как основного влияющего факторного признака.

Математическую модель процесса представим в виде уравнения регрессии, которое в общем виде выглядит так:

$$y = f(b_0, x_1). \tag{1}$$

Общий вид уравнения регрессии обычно представляется в форме линейной зависимости:

$$y = b_0 + b_1 x, \quad (2)$$

где y – результирующий признак (потребление ПЭВП); x – влияющий фактор (динамика ВВП); b_0, b_1 – коэффициенты регрессии, которые подлежат вычислению. Свободный член b_0 характеризует сдвиг и равен тому значению y , которое получается при $x = 0$, а коэффициент b_1 определяет наклон линии.

Для определения коэффициентов уравнения регрессии (b_0, b_1) применяют разные методы (графический, метод средних), однако наибольшее распространение получил метод наименьших квадратов (МНК).

Имеется n экспериментальных точек (9 пар значений по годам потребления ПЭВП в РФ и объема ВВП в период за 2000–2009 гг.): $(x_1, y_1); (x_2, y_2); \dots (x_9, y_9)$. Введем следующие обозначения: y_i – имеющиеся экспериментальные значения изучаемого параметра, а \hat{y}_i – его теоретические (рассчитанные по уравнению) показатели.

Предположим, что экспериментальные точки на графике укладываются так, что по ним вполне возможно провести прямую линию (рис. 1). Значения функции \hat{y}_i в этом случае можно записать в виде линейного уравнения: $\hat{y}_i = b_0 + b_1 \cdot x_i$. Расстояние по ординате от точки y_i до прямой составит: $b_0 + b_1 \cdot x_i - y_i = \varepsilon_i$, где $b_0 + b_1 \cdot x_i = \hat{y}_i$ – рассчитанное (теоретическое) значение функции; y_i – ее измеренное (опытное) значение и ε_i – разница (расстояние) между \hat{y}_i и y_i .

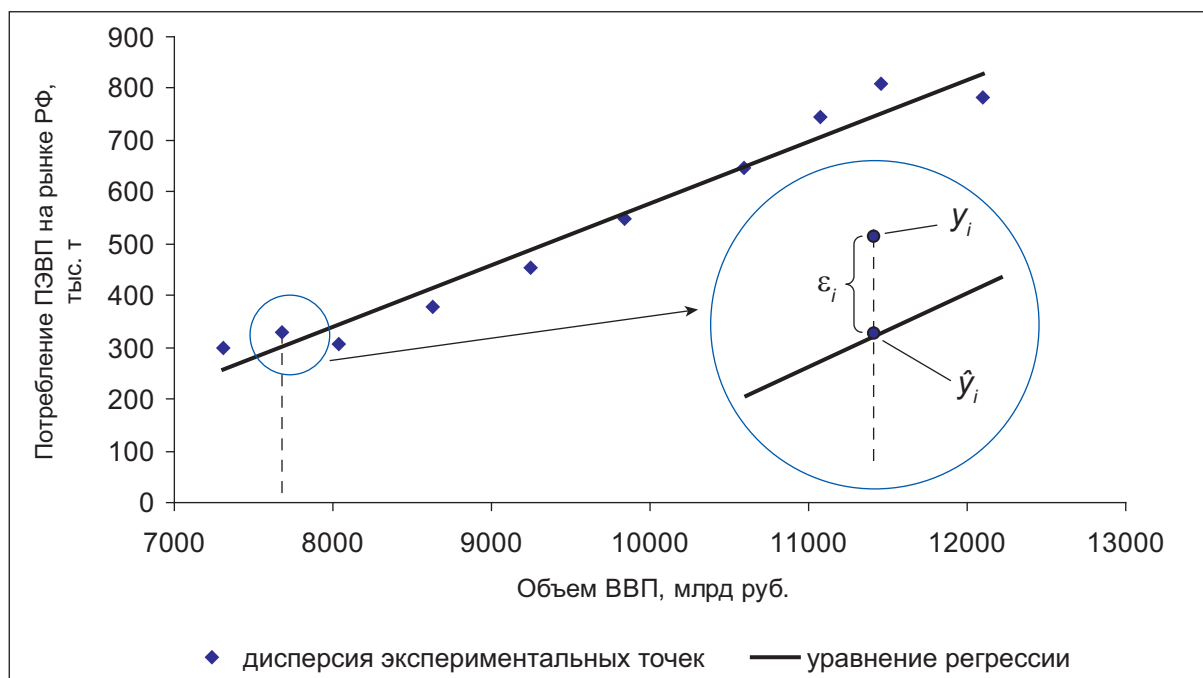


Рис. 1. Схематическое содержание метода наименьших квадратов

В соответствии с МНК предполагается, что искомая прямая будет наилучшей, если сумма квадратов всех расстояний $(b_0 + b_1 \cdot x_i - y_i)^2 = \varepsilon_i^2$ окажется наименьшей [6].

Коэффициенты b_0 и b_1 уравнения для случая линейной регрессии могут быть определены с использованием следующих уравнений:

$$b_0 = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}; \quad (3)$$

$$b_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}. \quad (4)$$

Наиболее просто расчет коэффициентов b_0 и b_1 может быть произведен посредством пакета MS Excel.

Полученное уравнение регрессии имеет вид

$$y = -614,44 + 0,117 \cdot x. \quad (5)$$

Для оценки значимости (пригодности) полученного уравнения регрессии была проведена проверка адекватности модели [6]. Для количественной проверки гипотезы об адекватности модели был использован F -критерий (критерий Фишера):

$$F = \frac{S_{ад}^2}{S_{общ}^2}, \quad (6)$$

где $S_{ад}^2$ – остаточная дисперсия, или дисперсия адекватности, которая характеризует величину среднего разброса экспериментальных точек $\Delta y = y_i - \hat{y}_i$ (Δy – ошибка в предсказании экспериментального результата на основании математической модели); остаточная дисперсия позволяет оценить ошибку, с которой уравнение регрессии предсказывает фактический результат; $S_{общ}^2$ – усредненная, или общая дисперсия, в качестве которой принят квадрат стандартной ошибки. Этот показатель характеризует случайную ошибку для всей выборки, то есть оценивает несоответствие между конкретными (текущими) значениями результата эксперимента и средним арифметическим.

Применение критерия минимальной остаточной дисперсии является вполне надежным способом отбора адекватных экономико-математических моделей. Для определения величины ошибки в предсказании эмпирических результатов ее нужно сопоставить с некоторой статистической величиной (эталоном), принимаемой в качестве критической. Поэтому используется расчетный F -критерий, который сравнивается с критическим F -критерием, определяемым по таблицам значений критерия Фишера. Если $F_{расч} < F_{кр}$, то модель признается адекватной, то есть с заданной степенью достоверности (надежности) она верно предсказывает реальный результат.

Остаточная дисперсия $S_{ад}^2$ рассчитана по формуле

$$S_{ад}^2 = \frac{\sum_{i=1}^n \Delta y^2}{f}. \quad (7)$$

Здесь число степеней свободы $f = n - (k + 1)$, где n – число опытов в выборке, а k – число изучаемых факторов.

Общая дисперсия $S_{общ}^2$ рассчитана по формуле

$$S_{общ}^2 = \frac{\sum_{i=1}^n (\bar{y} - y_i)^2}{f}, \quad (8)$$

где \bar{y} – среднее значение всей выборки y_i .

Расчет коэффициентов $S_{ад}^2$, $S_{общ}^2$ и F -критерия произведен посредством пакета MS Excel. Расчетное значение $F_{расч} = 0,014$, а $F_{кр} = 0,31$ для уровня значимости $\alpha = 0,05$, $n - 1$ степеней свободы для $S_{ад}^2$ и $n - 2$ степеней свободы для $S_{общ}^2$. Поскольку $0,014 < 0,31$, то с вероятностью 95 % можно утверждать, что рассматриваемое уравнение адекватно и способно с указанной достоверностью предсказывать прогнозные значения потребления ПЭВП в РФ.

Для расчета потребления ПЭВП в прогнозном периоде необходимо принять численные значения основного влияющего фактора (ВВП) в этом же периоде (табл. 4).

Таблица 4

Прогноз ВВП Российской Федерации на 2010–2020 гг.

Год	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Динамика изменения, %	3,8	4,2	3,9	4,5	5,7	5,9	5,6	5,3	4,7	4,4	4,4
ВВП, млрд руб., X	11511	12029	12522	13135	13987	14951	15970	17046	18129	19264	20425

Источник: [7, 8].

Подставляя в (5) значения ВВП из табл. 4, производим расчет потребления ПЭВП в прогнозном периоде на 2010–2020 гг. Фактические и прогнозные (расчетные) значения потребления ПЭВП в РФ за 2000–2020 гг. представлены на рис. 2.

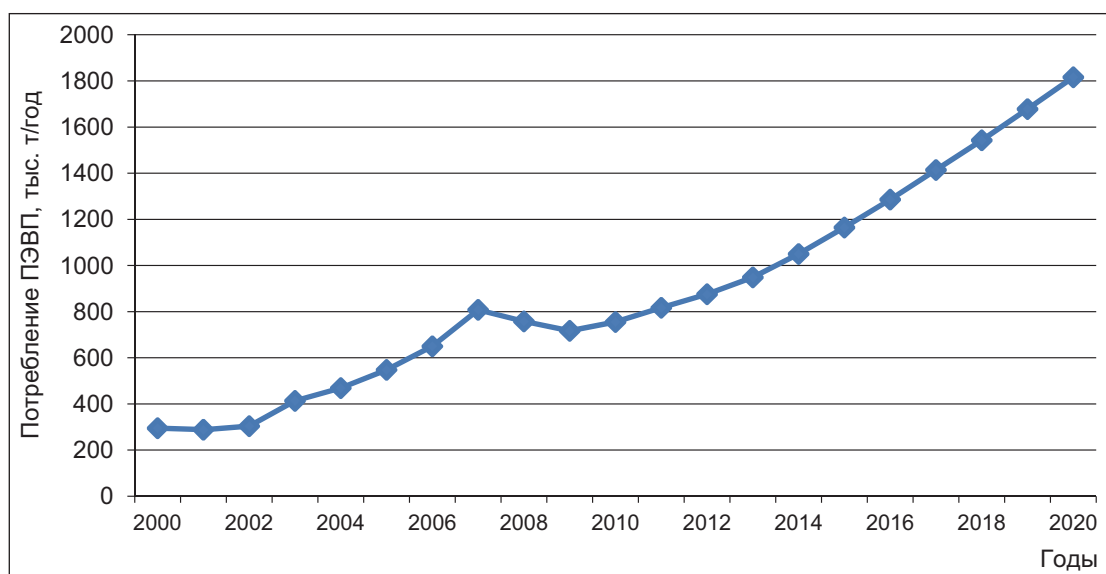


Рис. 2. Фактическое и прогнозируемое потребление ПЭВП в РФ за 2000–2020 гг.

Список литературы

1. Минюк С.А. Математические методы и модели в экономике / С.А. Минюк, Е.А. Ровба, К.К. Кузьмич. – Минск: Тетра Системс, 2002. – 432 с.
2. Дыброва Т.А. Статистические методы прогнозирования в экономике / Т.А. Дыброва, М.Ю. Архипова. – М.: МЭСИ, 2004. – 136 с.
3. Методические рекомендации к разработке прогнозов социально-экономического развития субъектов Российской Федерации. – М.: Мин-во экономического развития и торговли Российской Федерации, 2007. – 134 с.
4. Методические рекомендации и материалы по разработке прогноза научно-технологического и социально-экономического развития России до 2030 года // Матер. секций Координационного совета Российской академии наук по прогнозированию). – М.: ИНЭС, 2010. – 545 с.
5. Ивашиковский С.Н. Макроэкономика / С.Н. Ивашиковский. – М.: Дело, 2002. – 472 с.
6. Бараз В.Р. Корреляционно-регрессионный анализ связи показателей коммерческой деятельности с использованием программы Excel: учеб. пособие / В.Р. Бараз. – Екатеринбург : ГОУ ВПО «УГТУ–УПИ», 2005. – 102 с.
7. Концепция долгосрочного социально-экономического развития Российской Федерации до 2020 года: Распоряжение Правительства РФ от 17 ноября 2008 г. № 1662-р. // Собрание законодательства РФ. – 2008. – № 47. – Ст. 5489.
8. Сценарные условия для формирования вариантов социально-экономического развития в 2012–2014 гг. – М.: Мин-во экономического развития и торговли Российской Федерации, 2011. – 18 с.