

УДК 658.051.012

М.С. МАЗОРЧУК, Т.Ю. ПАПАЗОВА, Е.В. КОНОВАЛОВА

Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Украина

ПРОГНОЗИРОВАНИЕ ПОКАЗАТЕЛЕЙ РАБОТЫ ПРЕДПРИЯТИЯ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ РЕГРЕССИОННОГО АНАЛИЗА

Предлагается алгоритмическая модель прогнозирования показателей работы предприятия (прибыли, совокупного дохода, резервного фонда и т.д.) на основе методов статистического анализа данных, а именно, методов регрессионного анализа. Данная модель позволяет в комплексе решить задачу прогнозирования показателей работы предприятия, поскольку дает возможность системного учета различных факторов, которые могут оказать непосредственное влияние на деятельность предприятия в целом.

факторы, оказывающие влияние на работу предприятия, прогноз, регрессионная модель, корреляционная связь

Постановка проблемы

Задача прогнозирования состояния предприятия, т.е. определение будущей прибыли, оценка совокупных затрат, прогнозирование возможных убытков, является актуальной для любого предприятия. Однако, как правило, прогноз не всегда соответствует действительности, поскольку определить, от каких факторов зависят те или иные показатели, является достаточно сложным, что обусловлено высокой трудоемкостью расчетов и, зачастую, нежеланием проведения дополнительного анализа. С появлением современных информационных универсальных средств статистического анализа данных (Пакет анализа MS Excel, пакеты STATISTICA, SPSS и др.) задача прогнозирования стала решаться значительно проще, однако не всегда полученные результаты правильно анализируют и делают верные выводы. Поэтому, в данной статье предлагается алгоритмическая модель проведения анализа факторов, оказывающих влияние на работу предприятия, выявления главных из них, которые оказывают наиболее существенное влияние, и построение регрессионной модели, которая позволит получать точный прогноз показателей в будущем.

Анализ существующих методов прогнозирования

Прогнозирование можно разделить на четыре основных вида: *качественное, анализ временных*

рядов, анализ причинных (каузальных) связей и моделирование [1].

Качественные методы основаны на субъективных оценках и мнениях. В основе анализа временных рядов лежит идея использования информации о прошлых периодах и прогнозирования ее на будущее. При этом данные могут включать в себя несколько компонент, таких как тренды, сезонные или циклические колебания и т.д. Причинное прогнозирование использует метод линейной регрессии, в основе которого лежит предположение о том, что каждый показатель работы предприятия связан с некоторыми факторами, как внутренними, так и внешними. Основная цель регрессионного анализа – получить математическую модель (уравнение регрессии, линию предсказания), на основе которой возможно спрогнозировать будущие значения показателей. Моделирование позволяет определить влияние внешних и внутренних факторов при ряде допущений на прогноз показателей и оценить меру этого влияния.

Каждый из методов прогнозирования применяется в той или иной ситуации, в зависимости от наличия информации об объекте прогноза, числа наблюдений, качества информации, горизонта прогнозирования, наличия необходимых инструментальных средств. От этих факторов также будет зависеть

и точность прогноза. Например, качественные методы прогнозирования в основном применяются на ранних этапах анализа для получения предварительной оценки. В рамках данных методов широко используются модели экспертного оценивания и аналого-сопоставительные. Анализ временных рядов используется при построении краткосрочного или среднесрочного прогноза, когда модель данных представлена в виде тренда или имеются сезонные колебания. В этом случае не требуется высокой математической подготовки от лица, проводящего прогноз. Моделирование может проводиться на базе одной из моделей и используется тогда, когда практически невозможно получить строгую математическую модель в аналитическом виде.

В большинстве случаев на деятельность любого предприятия в современных рыночных условиях оказывает влияние большое количество факторов. При этом зависимость между ними может быть как функциональной, так и корреляционной. Поэтому, для выявления причинно-следственных связей в долгосрочном периоде целесообразно использовать методы регрессионного анализа, которые базируются на построении моделей линейной и многомерной регрессии.

Алгоритмическая модель прогнозирования показателей работы предприятия

Процедура регрессионного анализа описана в работах [2 – 4] и является достаточно известной. На первом этапе определяются оцениваемые показатели (зависимые переменные – отклики) и факторы, от которых предположительно зависит значение каждого отклика (независимые переменные – факторы). На втором этапе оценивается корреляционная связь между переменными и определяется вид этой связи. Если связь существует, то по методу наименьших квадратов строится уравнение регрессии (линейное или нелинейное). Затем анализируется качество полученной модели и проводится прогноз.

Однако в процессе проведения данного анализа часто возникают вопросы, связанные с оценкой значимости коэффициента связи, анализа ошибок, последовательности проведения этапов анализа и др. Поэтому в данной работе предлагается следующая алгоритмическая модель прогнозирования показателей работы предприятия на основе анализа влияния множества различных факторов (рис. 1):

Шаг 1. Выявляется множество факторов (переменных), от которых, предполагается, будут зависеть показатели работы предприятия. В качестве показателей, как отмечалось ранее, могут выступать величина спроса на продукцию, величина резервного фонда предприятия, прибыль, затраты. В общем случае можно выделить n различных внутренних и внешних факторов, которые могут оказать влияние на работу предприятия. К внутренним факторам относятся номенклатура и ассортимент выпускаемой продукции, качество изделий, риски предприятия и т.д. К внешним рыночным можно отнести такие факторы, как уровень инфляции, процентная ставка по безрисковой ценной бумаге, величина налога на прибыль, объемы иностранных инвестиций и другие.

Для расчета модели многомерной регрессии необходимо определиться с данными, которые будут использоваться для анализа. Это могут быть данные по работе предприятия за прошедший период, т.е. должны быть данные по величинам спроса, резервному фонду предприятия, прибыли или затратам в каждый отчетный период (например, за последние 10-15 лет), и за тот же период значения анализируемых факторов. Если предприятие новое и отсутствуют необходимые статистические данные, то возможно использовать отчетные данные по другим предприятиям, аналогичным по сфере деятельности и объему производства.

Шаг 2. Осуществляется проверка наличия связи между переменными. Наличие связи между переменными характеризует коэффициент корреляции, которые рассчитывается по формуле [5]:

$$r_{xy} = \frac{n \sum_{i=1}^n X_i Y_i - \left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right)}{\sqrt{\left[n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2 \right] \left[n \sum_{i=1}^n Y_i^2 - \left(\sum_{i=1}^n Y_i \right)^2 \right]}}, \quad (1)$$

где (X, Y) – двумерная случайная величина, которая в n наблюдениях может принимать случайные значения $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$.

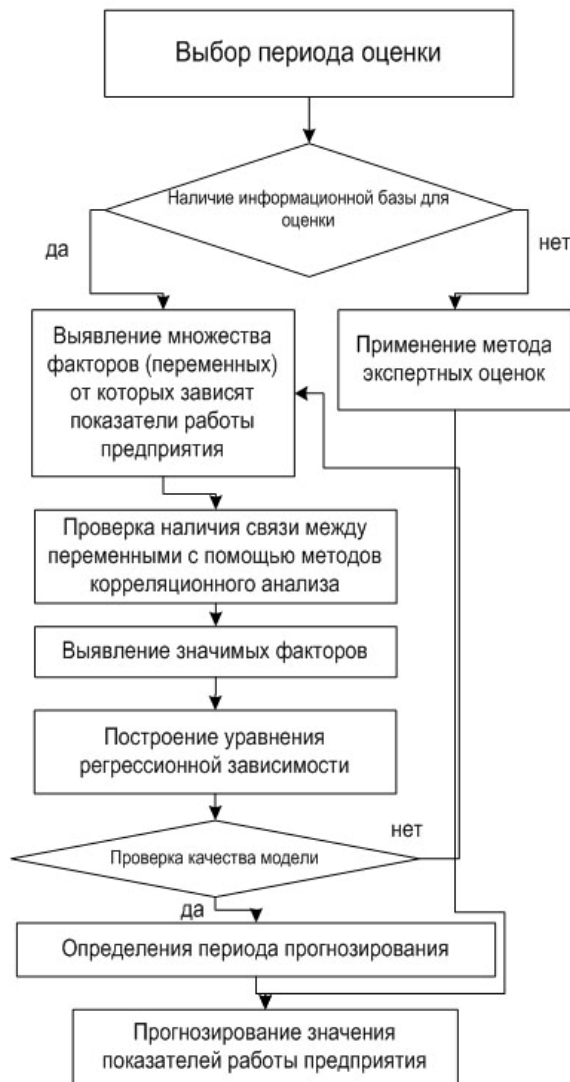


Рис. 1. Алгоритмическая модель прогнозирования показателей состояния предприятия

При определении коэффициента корреляции необходимо учитывать, что все наблюдения взаимонезависимы и имеют нормальный закон распределения [4, 5].

Коэффициент корреляции (1) отражает тесноту линейной связи между переменными и принимает значения от -1 до $+1$. Однако может иметь место и

криволинейная зависимость. В этом случае для проверки наличия связи можно использовать меру, называемую корреляционным отношением. Корреляционное отношение определяется так (5):

$$\eta_{y,x}^2 = 1 - \frac{S_w^2}{S_v^2}, \quad (2)$$

где $S_v^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2$ – сумма квадратов отклонений

каждого значения Y от среднего всех n значений

\bar{Y} ; $S_w^2 = \sum_{j=1}^J \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2$ – сумма квадратов от-

клонений соответствующих значений относительно их среднего (внутригрупповая сумма); n_j – количество переменных в группе; J – число групп. Корреляционное отношение принимает значения от 0 до 1.

Шаг 3. Оценка значимости коэффициента корреляции. Поскольку при оценке показателей работы предприятия в основном анализируются случайные параметры, то одной величины коэффициента парной корреляции для вывода о статистической значимости связи недостаточно [4]. Необходимо проверить, значимо ли оно отличается от нуля. Для этого можно воспользоваться критерием Стьюдента, т.е. проверить гипотезу о равенстве коэффициента корреляции нулю. Для этого необходимо рассчитать критериальное значение по формуле

$$t_{pac} = r \sqrt{(N-2)} / \sqrt{(1-r^2)}, \quad (3)$$

где r – значение коэффициента корреляции; N – количество наблюдений.

Если расчетное значение t_{pac} больше табличного, взятого с $N-2$ степенями свободы, то нулевая гипотеза отвергается. Это означает, что коэффициент корреляции значимо отличается от нуля (с выбранным уровнем значимости, например 5%). Полуширина доверительного интервала для коэффициента корреляции определяется по формуле:

$$\Delta = t_{N-2,\alpha} (1-r^2) / \sqrt{N}, \quad (4)$$

где N – число пар наблюдений, по которым рассчитывается коэффициент корреляции; r – значение коэффициента корреляции; $t_{N-2,\alpha}$ – табличное значение критерия Стьюдента, взятого с $N-2$ степенями свободы для заданного уровня значимости α .

Таким образом, оценив критериальное значение (3) и доверительный интервал (4), можно выявить наиболее значимые факторы, от которых будет зависеть каждый из показателей работы предприятия.

Следует отметить, что корреляционная зависимость еще не говорит о причинно-следственной связи между параметрами. Зависимость между переменными может быть значительно более сложной, чем парная [2, 4]. Однако в рамках данного исследования задачи оценки однородности коэффициентов корреляции, анализа аномальных наблюдений и другие не решаются.

Шаг 4. Для оценки прогноза показателей необходимо получить математическую модель, в которой бы была отображена зависимость оцениваемых величин от множества рыночных факторов. Получение по экспериментальным данным математических моделей, описывающих поведение некоторой характеристики в зависимости от изменения множества факторов – есть построение эмпирической зависимости, т.е. уравнения регрессии, которое имеет вид:

$$Y = F(X_1, X_2, \dots, X_m) + \varepsilon,$$

где Y – зависимая переменная (отклик); X_1, X_2, \dots, X_m – независимые переменные (факторы); ε – случайная ошибка.

Уравнение линейной регрессии имеет вид:

$$y_i = b_0 + \sum_{i=1}^k b_i f_i(x_i), \quad (5)$$

где $f_i(x_i)$ – некоторая функция от исходных переменных.

Для нахождения неизвестных коэффициентов b_0 и b_i уравнения (5) используется метод наименьших квадратов (МНК) при ряде допущений и предпосылок [2, 4, 5].

Поскольку данный метод довольно трудоемкий, особенно при анализе большого числа факторов, то для нахождения коэффициентов уравнения регрессии целесообразно воспользоваться пакетом анализа в MS Excel или известным пакетом статистической обработки данных STATISTICA.

Построение уравнения (5) выполняется тогда, когда между переменными имеет место линейная связь, которая определяется на основе линейного коэффициента корреляции (1). В случае наличия нелинейной связи ($\eta_{y,x}^2 \rightarrow 1$) строится нелинейная регрессионная модель (полиномиальная, экспоненциальная и другие) [2, 4]. Для определения структуры уравнения регрессии могут использоваться методы полного или частичного перебора, или пошаговой регрессии. До настоящего момента теоретически данная задача полностью не решена, поэтому часто можно получить ошибочные результаты.

Шаг 5. Анализ качества модели. Построенная модель линейной или нелинейной регрессии оценивается на информативность, адекватность, устойчивость, описывающие и предсказывающие свойства модели, отражение структуры связи между факторами и откликами. В табл. 1 приведены основные коэффициенты для оценки этих характеристик. Если по каким-либо показателям построенная модель не соответствует требуемому качеству, то производится возврат на шаг 1, т.е. необходимо выявить другие факторы, оказывающие влияние на каждый из показателей предприятия.

Шаг 6. На данном этапе определяется период, на который будут прогнозироваться показатели. Этот период, как правило, задается руководством предприятия. Если на предприятии планируется новые проекты к реализации, то период прогноза может определяться суммарным временем реализации проектов.

Обычно определяют максимально возможное время реализации проектов, т.е. при условии их последовательной реализации.

Анализ качества регрессионной модели

Свойства модели	Критерий	Характеристика
Информативность – наиболее полное отображение структуры связей	Множественный коэффициент корреляции $R^2 = \frac{\sum_{i=1}^N (\bar{Y} - \hat{Y}_i)^2}{\sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2} \rightarrow 1$	Величина R^2 представляет собой долю общей суммы квадратов, объясняемой моделью. Эмпирически установлено, что значение коэффициента должно быть не менее 0,96–0,97. Необходимое условие информативности
	Критерий Фишера $F_R = \frac{S_R^2}{S_{ocm}^2} > F_{\alpha, \nu_R, \nu_{ocm}}$	Достаточное условие информативности. Качественная характеристика информативности
	Критерий Бокса-Веца $\gamma \geq 2 \div 3$, который определяется по формуле: $F_0 \cong (1 + \gamma) F_{\alpha, \nu_0, \nu_{ocm}}$, где $\nu_0 = \nu_R (1 + \gamma)^2 / (1 + 2\gamma^2)$.	Количественная характеристика информативности. Если задать приближенно значение $\gamma = 3$, то $F_{расч} \approx 10 F_{табл}$
Адекватность – соответствие модели описываемому процессу	Критерий Фишера $F_R = \frac{S_R^2}{S_{ocm}^2} > \frac{(N - k) F_{\alpha, k-1, n-k}}{(N - 1)(1 + (k - 1)(N - k))}$	Применяется при отсутствии повторных опытов
	Критерий Фишера $F_R = \frac{S_R^2}{S_{ocm}^2} < F_{\alpha, (n-1)N, N-k}$	Применяется при наличии повторных опытов
Устойчивость – – устойчивость структуры уравнения регрессии; – устойчивость оценок коэффициентов регрессии	Коэффициент парной корреляции между регрессорами $ r_{x_i x_j} \leq 0,3 \div 0,4$ Коэффициент парной корреляции с откликом $ r_{x_i y} \gg r_{x_i x_j} $	Выполнение этих условий позволяет сделать выводы об устойчивости структуры уравнения регрессии
	Число обусловленности $cond = 1$	Означает, во сколько раз может увеличиться относительная ошибка в коэффициентах при наличии ошибки в исходной матрице
Описывающие и предсказывающие свойства	Дисперсия по контрольной выборке должна стремиться к дисперсии по экспериментальным данным $S_{ocm}^2 = \sum_{j=1}^N (\bar{Y}_i - \hat{Y}_i)^2 / (N - k) \approx$ $\approx S_R^2 = \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2 / (k - 1)$	Проверка адекватности по обучающей и контрольной выборке
Анализ структуры связей – позволяет получить графическую иллюстрацию распределения силы влияния регрессоров на отклик	Сила регрессора определяется как доля участия регрессора, рассчитываемая через отношение суммы квадратов, объясняемой данным регрессором, к общей сумме квадратов	Анализируется графически, на основе столбиковых диаграмм. Наличие «обрыва» в распределении свидетельствует о наличии в модели статистически незначимых регрессоров

Свойства модели	Критерий	Характеристика
<i>Анализ остатков</i> – позволяют оценить нарушения в структуре модели или проверить заданные предпосылки	Остатки определяются как n разностей $\varepsilon_i = \hat{Y}_i - Y_i, i = 1, 2, \dots, n$ Графики остатков могут иметь вид: – нормальный вид зависимости – горизонтальная полоса; – остатки при непостоянной дисперсии – расширяющаяся полоса; – наличие линейной функциональной связи 0 полоса под наклоном; – наличие нелинейной функциональной связи – полоса с изгибом	Остатки анализируются с помощью графиков остатков. Если в графиках остатков обнаружены нарушения, то это означает, что имеет место нарушение допущений регрессионного анализа, либо частная структура уравнения выбрана неправильно. Вид полосы определяет правильность полученной модели

Обозначения: N – число опытов; k – число членов модели; n – число дублирующих опытов в каждом эксперименте (в случае проведения эксперимента); \bar{Y}_i – среднее значение по повторным опытам в i -м эксперименте; \hat{Y}_i – значение отклика, рассчитанное для i -го эксперимента по модели; \bar{Y} – общее среднее; $\nu_R, \nu_{ост}$ – степени свободы для дисперсии, объясняемой моделью и остаточной дисперсии соответственно; α – уровень значимости.

Шаг 7. На основе построенного уравнения регрессии (5) производится прогнозирование оцениваемых величин. Для этого необходим прогноз рыночных факторов, которые являются значимыми для каждого из показателей. Такой прогноз может быть получен и по более простым моделям (анализ временных рядов, экспертные оценки). Однако в рамках данной статьи они не рассматриваются.

Выводы

Таким образом, с использованием данной алгоритмической модели можно провести достаточно полный и обоснованный анализ влияния множества различных факторов на показатели работы предприятия, получить математические модели для расчета каждого из показателей и спрогнозировать на основе этих моделей значения оцениваемых величин на будущее.

Данная модель может быть использована не только для проведения аналитических расчетов на предприятии, но и в других сферах деятельности, например при прогнозировании показателей социально-экономических и бизнес-процессов.

Литература

1. Чейз Р.Б., Эквилайн Н.Дж., Якобс Р.Ф. Производственный и операционный менеджмент, 8-е изд.: Пер. с англ. – М.: Издательский дом «Вильямс», 2003. – 704 с.
2. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высш. шк., 2004. – 479 с.
3. Шамша Б.В., Гуржий А.М., Дудар З.В., Левикин В.М. Математичне забезпечення інформаційно-управляючих систем. – Х.: ТОВ Компанія «СМІТ», 2005. – 448 с.
4. Лапач С.Н., Чубенко А.В., Бабич П.Н. Статистика в науке и бизнесе. – К.: МОРИОН, 2002. – 640 с.
5. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. – М.: Прогресс, 1976. – 494 с.

Поступила в редакцию 13.07.2007

Рецензент: д-р техн. наук, проф. А.Н. Баранов, Украинская инженерно-педагогическая академия, Харьков.