ЛАБОРАТОРНАЯ РАБОТА 4. ОЦЕНКА АДЕКВАТНОСТИ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ

При рассмотрении основных предположений, положенных в основу МНК, мы приняли, что ошибки e_t представляют собой случайные независимые (некоррелируемые) величины. При работе с фактическими данными, особенно с динамическими рядами, такое допущение далеко не всегда имеет под собой реальную почву. В самом деле, если вид функции (закон развития) выбран неудачно, то вряд ли можно говорить о том, что отклонения от регрессии являются независимыми. В этом случае обычно наблюдается заметная концентрация положительных и отрицательных отклонений от регрессии и можно сомневаться в их случайном характере.

Одним из основных методов проверки адекватности разработанной модели является анализ остаточной (случайной) компоненты et. Значения случайной компоненты получаются после исключения систематической компоненты:

$$e_t = y_t - \hat{y}_t.$$

Принято считать, что модель адекватна процессу, если e_t является случайной независимой последовательностью чисел, подчиняющейся нормальному закону распределения.

Для того, чтобы определить, обладают ли значения остаточной компоненты e_t перечисленными свойствами, необходимо выполнить следующий анализ:

Проверка ***случайности*** последовательности e_t предполагает установление того факта, что в последовательности e_t отсутствует тенденция, что и можно определить одним из известных методов. Одним из методов проверки наличия тенденции является критерий поворотных точек, или пиков. Каждое значение e_t сравнивается с последующим и предыдущим значениями. Точка считается поворотной, если e_t либо больше и предыдущего, и последующего значений, либо меньше и того, и другого. В случайной последовательности должно выполняться строгое неравенство:

$$p > \left\lfloor 2(n-2)/3 - 2\sqrt{(16n-29)/90} \right\rfloor,$$

где p - количество поворотных точек;

[ ] – целая часть результата вычислений.

1.Если последовательные значения e_t коррелируют между собой, то говорят, что имеет место ***автокорреляция ошибок.*** Метод наименьших квадратов и в случае автокорреляции дает несмещенные и состоятельные оценки. Однако получаемая при наличии высокой автокорреляции стандартная ошибка и соответственно доверительный интервал имеют мало смысла в силу своей ненадежности. Во всяком случае, значительная автокорреляция говорит о том, что спецификация регрессии неправильна. При анализе соответствующей регрессии необходимо определить, присутствует или нет автокорреляция в остаточных членах e_t (особенно если исследуется временной ряд). Существует ряд приемов обнаружения автокорреляции. Наиболее простым и достаточно обоснованные из них является метод Дарбина — Уотсона, связанный с гипотезой о существовании автокорреляции первого порядка. Соответствующий критерий имеет вид:

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}.$$

Для d-статистики найдены критические границы (d_1—верхняя и d_2—нижняя), позволяющие принять или отклонить гипотезу об отсутствии автокорреляции при 1; 2,5 и 5%-ном уровне существенности. Сокращенная таблица значений этого критерия приведена в [22] для n от 15 до 100. Если расчетное значение критерия d больше двух, то его необходимо преобразовать:

$$d' = 4 - d.$$

Если эмпирическое значение $d(d')$ находится в пределах от нуля до d_1, то гипотеза об отсутствии автокорреляции не отклоняется. Если значение $d(d')$ попадает в интервал от d_2 до 2, то автокорреляция отсутствует. Если же эта статистика находится в пределах $d_1 < d(d') < d_2$, то нет статистических оснований ни принять, ни отклонить эту гипотезу (области неопределенности).

В этом случае используют другие критерии, например, коэффициент автокорреляции первого порядка:

$$r(1) = \frac{\sum_{t=2}^{n} e_t \cdot e_{t-1}}{\sum_{t=1}^{n} e_t^2}.$$

Если $|r(1)|$ окажется меньше табличного значения, то нулевая гипотеза о присутствии автокорреляции отвергается.

Если на обнаружена существенная автокорреляция отклонений от регрессии, то логично признать наличие ошибки в спецификации уравнения. Следовательно, надо вернуться к этой проблеме, пересмотреть набор включаемых в уравнение переменных и уточнить форму уравнения.

2.Соответствие остаточной компонеты e_t нормальному распределению также свидетельствует о том, что спецификация уравнения регрессии проведена правильно.

Наиболее простым методом проверки является так называемый RS метод. Согласно этому методу, для исследуемой совокупности вычисляется отношение R/s, где

R – размах выборки: $R=e_{max}-e_{min}$,

s – стандартное отклонение:

$$s=\sqrt{\frac{\sum e_t^2}{n-2}}.$$

Если рассчитанное значение попадает между табулированными границами с заданным уровнем доверительной вероятности, то гипотеза о нормальном распределении принимается.

При анализе адекватности существенной характеристикой модели является её точность, для определения которой вычисляют среднюю относительную ошибку:

$$\bar{E}_{отн}=\frac{1}{n}\sum_{t=1}^{n}\frac{|e_t|}{y_t}\cdot 100\%.$$

Задание.

1) Оценить адекватность построенных моделей (линейной и адаптивной) на основе исследования[1]:

   а) случайности остаточной компоненты по критерию пиков;

   б) независимости ряда остатков по d–критерию (в качестве критических используйте уровни d_1=1,08 и d_2=1,36) или по коэффициенту корреляции первого порядка $r(1)$=0,36;

   в) нормальности распределения остаточной компоненты по R/S–критерию с критическими уровнями 2,7-3,7;

   г) для оценки точности модели используйте среднее квадратическое отклонение и среднюю по модулю ошибку.

2) Построить точечный и интервальный прогнозы на два шага вперед (для вероятности p=0,7 используйте коэффициент K_p=1,05)

Отобразить на графиках фактические данные, результаты расчетов и прогнозирования (1,3,5,6).

Вычисления провести с одним знаком в дробной части.

---

[1] Основные и промежуточные результаты вычислений представить в **табл. 1**

Таблица 1.

t	y_t	$\hat{y}_t = a_0 + a_1 \cdot t$	$e_t = y_t - \hat{y}_t$	e_t^2	$e_t - e_{t-1}$	$[e_t - e_{t-1}]^2$	$e_t \cdot e_{t-1}$	$\left(\lvert e_t\rvert / y_t\right) \cdot 100\%$	$t - \bar{t}$	$(t - \bar{t})^2$
					-		-			
$\sum$				$\sum$		$\sum$	$\sum$	$\sum$		