



МЕТОДИКА АНАЛИЗА ДАННЫХ МОНИТОРИНГА ЗАКУПОЧНЫХ ЦЕН НА ПРИМЕРЕ САРАТОВСКОЙ ОБЛАСТИ

Воротников И.Л., д.э.н., Розанов А.В., к.ф.-м.н,
Сидельникова М.В., к.э.н., ФГБОУ ВО СГАУ им. Н.И. Вавилова

Мониторинг цен является важным этапом анализа и прогнозирования конъюнктуры на зерновом рынке. Значительное влияние на результаты прогнозирования динамики цен оказывают предельно большие или аномально малые значения цен, резко выделяющиеся из общего ряда других значений. Целью статьи является обоснование выбора эффективной методики предпрогнозной обработки данных мониторинга, обладающей свойством устойчивости к наличию аномальных значений в исходных данных и к нарушению стандартных предположений относительно вида распределения вероятностей, на основе которой может быть организована автоматизированная обработка данных. Предлагается методика оперативного анализа данных мониторинга цен на зерновом рынке, которая предполагает проведение идентификации законов распределения вероятности временных рядов цен на зерно и применение робастных (помехоустойчивых) статистических оценок в случае отклонения распределений от нормального (гауссовского). На примере анализа закупочных цен на пшеницу в Саратовской области показано, что формальное применение типовых процедур идентификации и статистической обработки данных без учета их вероятностных характеристик может приводить к серьезным ошибкам прогнозирования.

Ключевые слова: мониторинг цен на зерновом рынке, идентификация распределений вероятности, прогнозирование, робастные М-оценки.

THE TECHNIQUE OF ANALYSIS OF MONITORING DATA OF PURCHASE PRICES ON THE EXAMPLE OF THE SARATOV REGION

Vorotnikov I.L., doctor of economic sciences,
Rosanov A.V., candidate of physical and mathematical sciences,
Sidelnikova M.V., candidate of economic sciences,
Saratov State Agrarian University named after N.I. Vavilov

Price monitoring is an important step in the analysis and forecasting of the conjuncture on the grain market. Extremely large or abnormally low price values, which stand out sharply from the general range of other values, have a significant impact on the results of forecasting price movements. The purpose of the article is to substantiate the choice of an effective method for predictive processing of monitoring data that has the property of resistance to the presence of anomalous values in the source data and to a violation of standard assumptions regarding the type of probability distribution, on the basis of which automated data processing can be organized. It is proposed a technique for the on-line analysis of price monitoring data in the grain market, which involves conducting a primary analysis of the laws of probability distribution of time series and applying robust statistical estimates in the case of identifying the deviation from normal (Gaussian) distributions. Using the example of the analysis of procurement prices in the Saratov region, it is shown that the formal application of standard identification procedures and statistical data processing without taking into account their probabilistic characteristics can lead to serious forecasting errors.

Keywords: monitoring of grain market prices, identification of probability distributions, forecasting, robust M-estimates.

Введение.

Мониторинг цен является важным этапом анализа и прогнозирования конъюнктуры на зерновом рынке. Закупочные цены на этом рынке (см. рис.1), как и любые другие реальные процессы в агросфере, подвержены случайным влияниям погрешностей, обусловленных неконтролируемым воздействием внешних или внутренних факторов различной природы. Зна-



чительное влияние на результаты прогнозирования динамики цен оказывают предельно большие или аномально малые значения цен, резко выделяющиеся из общего ряда других значений. Статистическая обработка данных мониторинга цен призвана, во-первых, уменьшить влияние чисто случайных воздействий, и во-вторых, выявить факторы, значительное отклонение которых от статистически средних объясняется собственными скрытыми свойствами сложных систем.

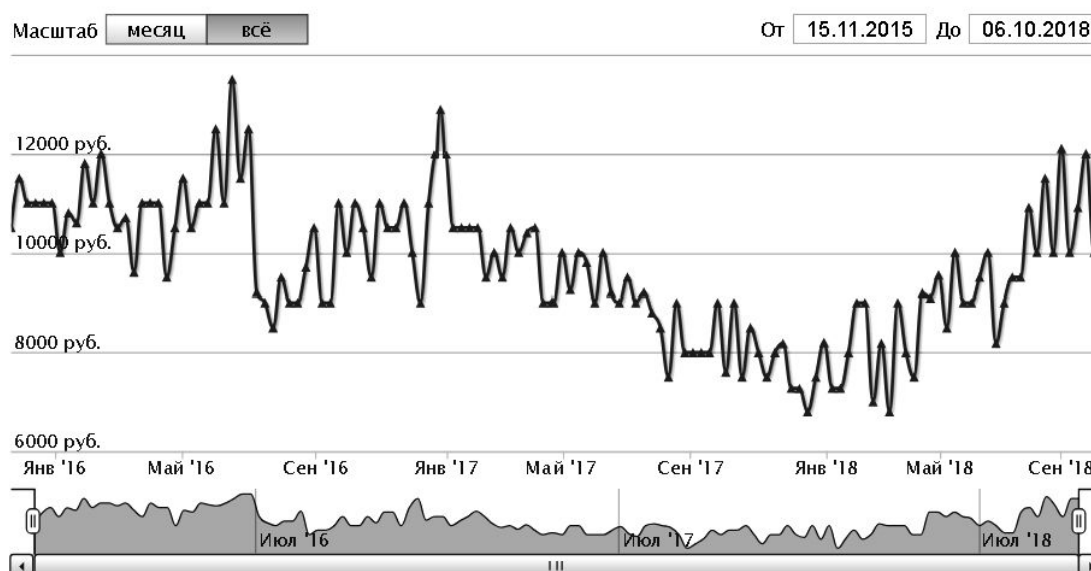


Рисунок 1 – Закупочные цены (за тонну) на пшеницу 3 класса по Саратовской области [1]

Проблема выявления и коррекции резко выделяющихся или пропущенных значений осложняется тем, что заранее неизвестно, выполняются ли для рассматриваемых рядов стандартные предположения относительно нормальности распределения вероятностей и устойчивости к действию аномальных значений. Поэтому первым этапом анализа «сырых» данных должна быть процедура идентификации распределения вероятностей.

Цель исследования.

Целью настоящей работы является обоснование выбора эффективной методики прогнозной обработки данных мониторинга, обладающей свойством устойчивости к наличию аномальных значений в исходных данных и к нарушению стандартных предположений относительно вида распределения вероятностей, на основе которой может быть организована автоматизированная обработка данных.

Методика исследования.

Оперативную оценку нормальности целесообразно проводить на основе следующей процедуры, легко поддающейся автоматизации [2, 3]. Закон распределения можно с 95% уверенностью считать близким к нормальному, если выполняются условия:

$$\left| \frac{CAO}{\sigma} - 0,7979 \right| < \frac{0,4}{\sqrt{n}}, \quad \sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - x_{cp})^2}, \quad CAO = \frac{\sum_{i=1}^n |x_i - x_{cp}|}{n}, \quad i = 1, 2, \dots, n, \quad (1)$$

где n – число членов ряда, σ – выборочное среднее квадратическое отклонение, CAO – среднее абсолютное отклонение, x_i и x_{cp} – текущее и среднее значения элементов выборки. В особо критических случаях следует выполнять дополнительную проверку нормальности, используя более продвинутые критерии нормальности исследуемой выборки [4].

Если нормальность распределения подтверждена, можно проводить идентификацию «подозрительных» значений ряда. Для каждого члена ряда следует выполнить проверку ус-



ловия (K1): $|x_i - x_{cp}| > t_q \cdot \sigma$, $i = 1, 2, \dots, n$, где t_q – критическое значение квантиля распределения Стьюдента, определяемое по таблицам для заданных величин параметров n и q . Величину q выбирают обычно равной 0,05 или 0,01, исходя из заданного уровня доверительной вероятности (95% или 99%). Если это условие выполняется, то соответствующее значение выборки должно быть признано случайным и его следует исключить из набора данных.

На практике реальные данные почти никогда не имеют чисто гауссовского распределения, что заметно усложняет использование параметрических оценок, алгоритмы которых базируются на свойствах нормального распределения вероятностей и быстро теряют свойства эффективности и несмещённости даже при небольших отклонениях от гауссовской статистики. В подобных условиях действенным способом повышения качества оценок случайных процессов с негауссовскими распределениями вероятности является применение так называемых робастных оценок. Под робастностью понимается слабая чувствительность к отклонениям от стандартных условий и высокая эффективность для широкого класса вероятностных распределений [5].

В настоящее время наибольшее распространение получили робастные методы на основе так называемых М-оценок, в которых идентификация параметров опирается на принцип максимального правдоподобия. Критерием отнесения «подозрительных» значений ряда к величинам, подлежащим исключению из выборки, в данном случае является условие (K2): $|x_i - x_{cp}| > k \cdot S_{MED}$, $i = 1, 2, \dots, n$, где $S_{MED} = MED(|x_i - MED(x_i)| / 0,6745)$, MED – оператор вычисления медианы. Величина k выбирается равной 3, согласно полуэмпирическому правилу «трех сигм» [6]. Параметр нормировки 0,6745 выбирают из условия, что при $n \rightarrow \infty$ оценка параметра масштаба S_{MED} должна стремиться к обычному среднему квадратическому отклонению при нормальном распределении вероятностей.

Результаты исследования.

Эффективность представленной выше методики идентификации «подозрительных» значений проиллюстрируем на примерах из практики обработки данных мониторинга закупочных цен на зерновые на примере Саратовской области. На рисунке 2 представлен фрагмент графика изменения максимальной цены на пшеницу 3 класса по Саратовской области в период январь 2016 – сентябрь 2016 [1].

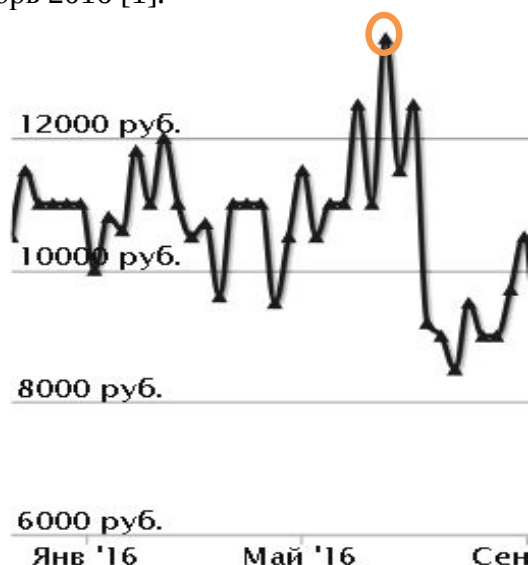


Рисунок 2 – Фрагмент графика закупочных цен на пшеницу 3 класса по Саратовской области [1]. Цветом выделено значение, идентифицированное как случайное отклонение по критерию (K1)

Проверка условия нормальности распределения вероятностей по формулам (1) показывает, что условие $|CAO / \sigma - 0,7979| = 0,02 < 0,4 / \sqrt{n} = 0,08$ выполняется и, следовательно,



можно применять критерий (K1) для выявления аномального значения рассматриваемого ряда.

На рисунке 3 представлен ещё один фрагмент графика изменения максимальных цен на пшеницу [1]. Распределение вероятностей на данном фрагменте не может быть отнесено к нормальному. На это указывает условие $|CAO/\sigma - 0,7979| = 0,132 > 0,4/\sqrt{n} = 0,08$. Следовательно, применение критерия нормальности распределения на основе критерия (K1) для выявления выделяющихся значений рассматриваемого ряда не является оправданным.

Резко выделяющееся значение, показанное на рисунке 3, было идентифицировано робастной процедурой (K2) на основе М-оценок параметров положения $m=MED$ и масштаба $S = S_{MED}$.

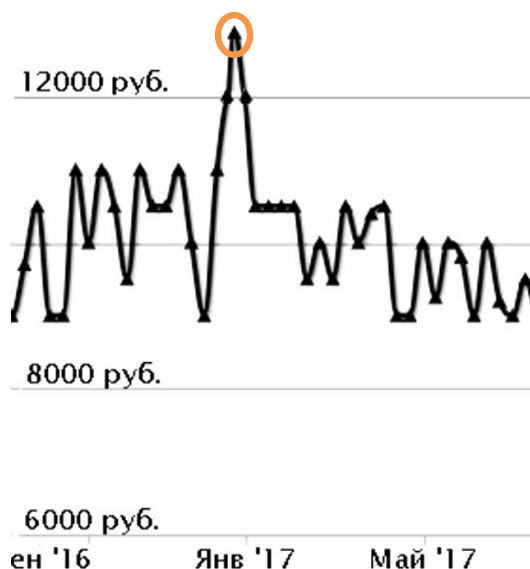


Рисунок 3 – Фрагмент графика закупочных цен на пшеницу 3 класса по Саратовской области [1]. Выделено аномальное значение, идентифицированное по робастному критерию (K2)

Подводя итог, важно подчеркнуть, что, хотя визуально фрагменты временных рядов на рисунках 2 и 3 мало отличаются друг от друга, их вероятностные характеристики принципиально различны и применение однотипных процедур идентификации для анализа и статистической обработки данных может приводить к серьезным ошибкам. Поэтому во всех случаях, когда не удастся однозначно установить нормальность распределения вероятностей, целесообразно использовать робастные оценки, малочувствительные к резко выделяющимся значениям, отклонению от нормального (гауссовского) характера распределения и наличию «тяжелых» хвостов распределения. Это обстоятельство становится особенно актуальным при мониторинге данных в зерновой и продовольственной сфере АПК, когда число отсчетов, поступающих в реальном масштабе времени, настолько велико, что визуальный контроль отдельных значений оказывается практически невозможным [7].

Выводы.

Автоматизация анализа и идентификации уровня цен на основе предлагаемой методики позволяет повысить оперативность обработки данных, дает возможность применять более эффективные экономико-математические модели данных и на их основе строить более точные прогнозы [8]. В частности, «очистка» данных мониторинга закупочных цен на основе представленной выше методики показала, что в Саратовской области в 2018 году период минимума закупочных цен на пшеницу и другие зерновые культуры оказался сдвинутым на период январь-февраль, в противоположность среднегодовым тенденциям, в которых минимальные закупочные цены традиционно формировались в августе-сентябре предшествующего года. Это во многом обусловлено заметным ростом урожайности зерновых в 2016-2017



годах, что связано не со случайными воздействиями, а с особо благоприятными погодными условиями. Учитывая сложившуюся в Саратовской области внутригодовую динамику цен, следует ожидать более быстрого роста уровня цен на зерновые ещё до начала нового 2019 года и соответствующим образом корректировать финансовый менеджмент на местном и экспортно-импортном рынках.

Список литературы:

1. АгроНовости. Электронный ресурс. Код доступа <https://agro-bursa.ru/prices/wheat/>
2. Гайдышев И. Анализ и обработка данных. Специальный справочник. — СПб.: Питер, 2001. — 752 с.: ил.
3. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. М.: Физматлит, 2006. — 816 с., ил.
4. Крянев А.В., Лукин Г.В. Метрический анализ и обработка данных. — М.: Физматлит, 2010. — 280 с., ил.
5. Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В. Робастность в статистике. Подход на основе функций влияния. - М.: Мир, 1989. — 512 с., ил.
6. Montgomery D.C. Applied Statistics and Probability for Engineers / D.C. Montgomery, G.C. Runger – 4th ed. – Wiley, 2006 – 784 p.
7. Пселтис Эндрю Дж. Потокковая обработка данных. Конвейер реального времени. Пер. с англ. – М.: ДМК-Пресс, 2018. – 218 с., ил.
8. Воротников И.Л., Розанов А.В., Котова М.В. Анализ и прогнозирование динамики цен на продукты питания (на примере Саратовской области) // Экономика сельскохозяйственных и перерабатывающих предприятий. 2016. № 5. с. 59-62.