

ПРОГНОЗИРОВАНИЕ ОБЪЕМОВ ПРОДАЖ КОМПАНИИ МЕТОДАМИ СТРУКТУРНОГО АНАЛИЗА ДАННЫХ

Лабунец Н. Л., Лабунец Л. В.

Рассмотрена методика моделирования и прогнозирования доходов от продаж компании методами структурного анализа данных. Разложение временного ряда объемов продаж на структурные составляющие реализовано методом «Гусеница». Обосновано применение метода главных компонент для формирования моделей нелинейной регрессии, устанавливающих взаимные зависимости между средними доходами от продаж и влияющими на них факторами. На основе выборочных оценок статистик показана эффективность применения SSA- прогноза сезонной составляющей для прогнозирования объемов продаж компании с заданной доверительной вероятностью.

THE COMPANY REVENUE FORECASTING BY THE METHODS OF THE STRUCTURAL DATA ANALYSIS

Labunets N. L., Labunets L. V.

The modeling and forecasting technique of the company revenue by methods of the structural data analysis is considered. Time series decomposition of the sales on structural components is carried out by the "Caterpillar" method. Application of a principal components analysis for formation of the nonlinear regression models establishing interdependences between average revenue and factors influencing them is proved. On the basis of the sample statistical data estimations the efficiency of application SSA- forecast of a seasonal component for forecasting sales with the set confidential probability is shown.

Рассмотрена методика моделирования и прогнозирования доходов от продаж компании методами структурного анализа данных. На основе современных эконометрических моделей получено решение задач:

- выделения структурных составляющих временного ряда (ВР) объемов продаж;
- определения зависимости средних объемов продаж от количества продавцов и партнеров компании (прямая задача);
- определения количества продавцов и партнеров, необходимых для обеспечения планового уровня средних объемов продаж (обратная задача);
- прогнозирования объемов продаж с заданной доверительной вероятностью.

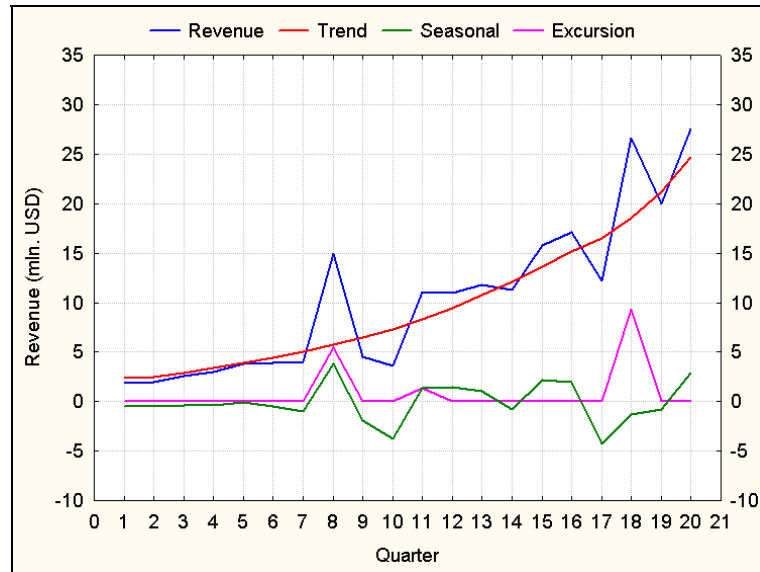
Исследована факторная модель доходов от продаж продукции одной из российских компаний в зависимости от количества продавцов первого и второго типа - X_1 и X_2 ; технических специалистов - X_3 ; бизнес партнеров компании первого и второго уровней - X_4 и X_5 , а также от тенденции изменения средних объемов продаж X_6 по рынку анализируемой продукции. Важно отметить, что особенностью анализа являлся малый объем выборки за 20 предшествующих кварталов.

Методом «Гусеница» [1] выделены структурные составляющие ВР объемов продаж в виде тренда и 4-х сезонных компонент (рис. 1). В качестве параметров метода применялись: временной интервал вложения ряда в траекторное пространство – 12 кварталов; ранг временного ряда - 9 сингулярных троек; количество структурных групп - 5 ($E1$ – тренд; $E2_3, E4_5, E6_7, E8_9$ – сезонные компоненты).

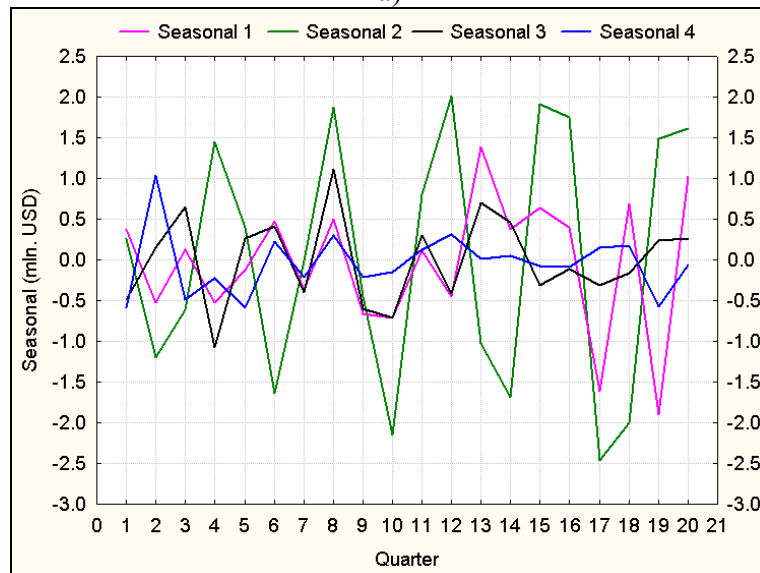
SSA- анализ показал, что коэффициент взаимной корреляции суммарной сезонной составляющей и тренда не превышал значения 0,272 по абсолютной величине. Тренд демонстрировал экспоненциальный темп роста $Trend(q) = \exp\{0.6852 + 0.1286*q\}$, где q – номер квартала (рис. 1). Поэтому дальнейшее исследование выполнялось в логарифмическом масштабе данных.

Обоснована эффективность применения метода главных компонент (МГК) для сокращения размерности пространства исходных предикторов $\ln(X_1), \dots, \ln(X_6)$. В частности показано, что выборочные оценки коэффициентов взаимной корреляции факторов не ниже значения 0,904. Первая главная компонента в разложении матрицы коэффициентов взаимной корреляции предикторов по собственным значениям и векторам является

доминирующей. На ее долю приходится 95,91% от обобщенной дисперсии логарифмов исходных факторов.



а)



б)

Рис. 1. Структурные составляющие доходов от продаж компании. Тренд, суммарная сезонная составляющая и выбросы - а); сезонные компоненты – б).

С помощью декоррелирующего преобразования получена зависимость главного фактора от предикторов

$$Factor(X_1(q), \dots, X_6(q)) = \sum_{n=1}^6 W_n \{ \ln(X_n(q)) - M_n \},$$

где $\vec{M} = (M_1, \dots, M_6)$ - оценка вектора математического ожидания для логарифмов исходных переменных; $\vec{W} = (W_1, \dots, W_6) = \vec{U}_1 / \sqrt{\lambda_1}$ - вектор весовых коэффициентов;

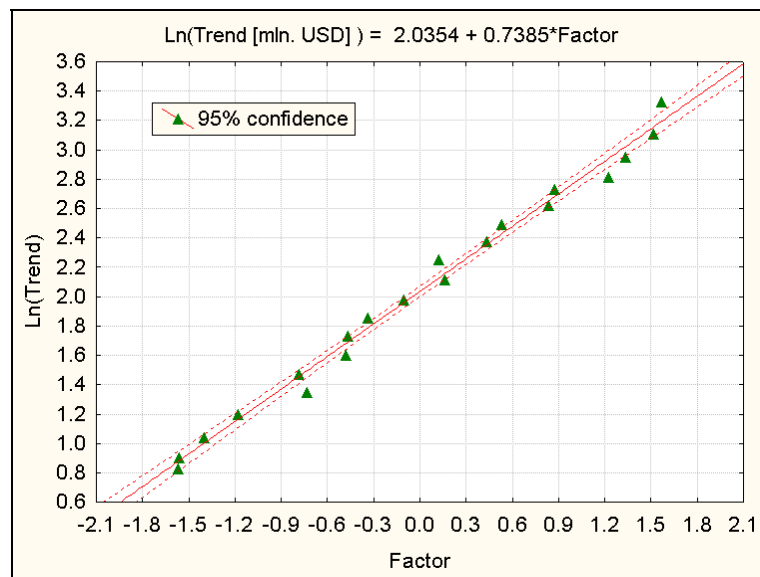
$$\vec{U}_1 = (0.451105, 0.500209, 0.488444, 0.349816, 0.403611, 0.149816) \text{ и } \lambda_1 = 2,2066$$

- ортонормированный собственный вектор и соответствующее ему доминирующее собственное значение.

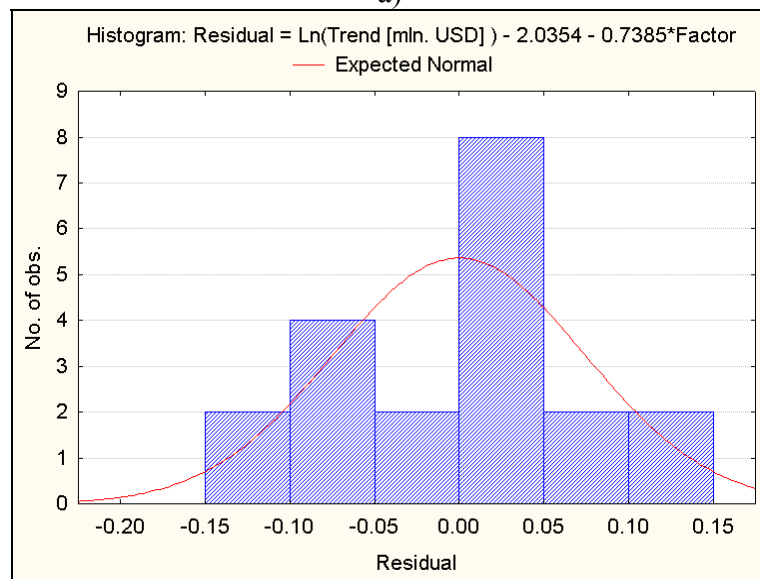
В рамках МГК получено решение прямой задачи в виде модели нелинейной регрессии

$$Trend(q) = \exp \{ 2.0354 + 0.7385 * Factor(X_1(q), \dots, X_6(q)) + Residual(q) \} \quad (1)$$

главного фактора на средние доходы от продаж продукции компании. Погрешности моделирования иллюстрируют диаграмма рассеяния и выборочная оценка гистограммы ошибок аппроксимации $Residual(q)$ (рис. 2).



а)



б)

Рис. 2. Погрешности модели регрессии главного фактора на средние доходы от продаж. Диаграмма рассеяния - а) и гистограмма ошибок аппроксимации – б).

МГК позволил также эффективно решить обратную задачу, т.е. оценить необходимые величины исходных факторов для достижения плановых значений средних объемов продаж. Соответствующая модель линейной регрессии в логарифмическом масштабе данных имеет вид коррелирующего преобразования

$$\ln(X_n(q)) = M_n + V_n Factor(q) + Error_n, \quad (n = \overline{1, 6}). \quad (2)$$

Здесь $\vec{V} = (V_1, \dots, V_6) = \vec{U}_1 \sqrt{\lambda_1} = \lambda_1 \vec{W}$ - вектор весовых коэффициентов. После подстановки в формулу (2) выражения для главного фактора, полученного из уравнения (1), и несложных преобразований получена модель *нелинейной* регрессии

$$X_n(q) \approx Y_n \{Trend(q)\}^{Z_n}, \quad (3)$$

$$Y_n = \exp(M_n - 2.0354 Z_n), \quad Z_n = V_n / 0.7385, \quad (n = \overline{1,6})$$

средних доходов от продаж на исходные переменные. Погрешности этих зависимостей удобно оценивать с помощью диаграмм рассеяния и выборочных оценок для гистограмм ошибок аппроксимации $Error_n$. Например, для технических специалистов X_3 указанные выше статистики представлены на рис. 3.

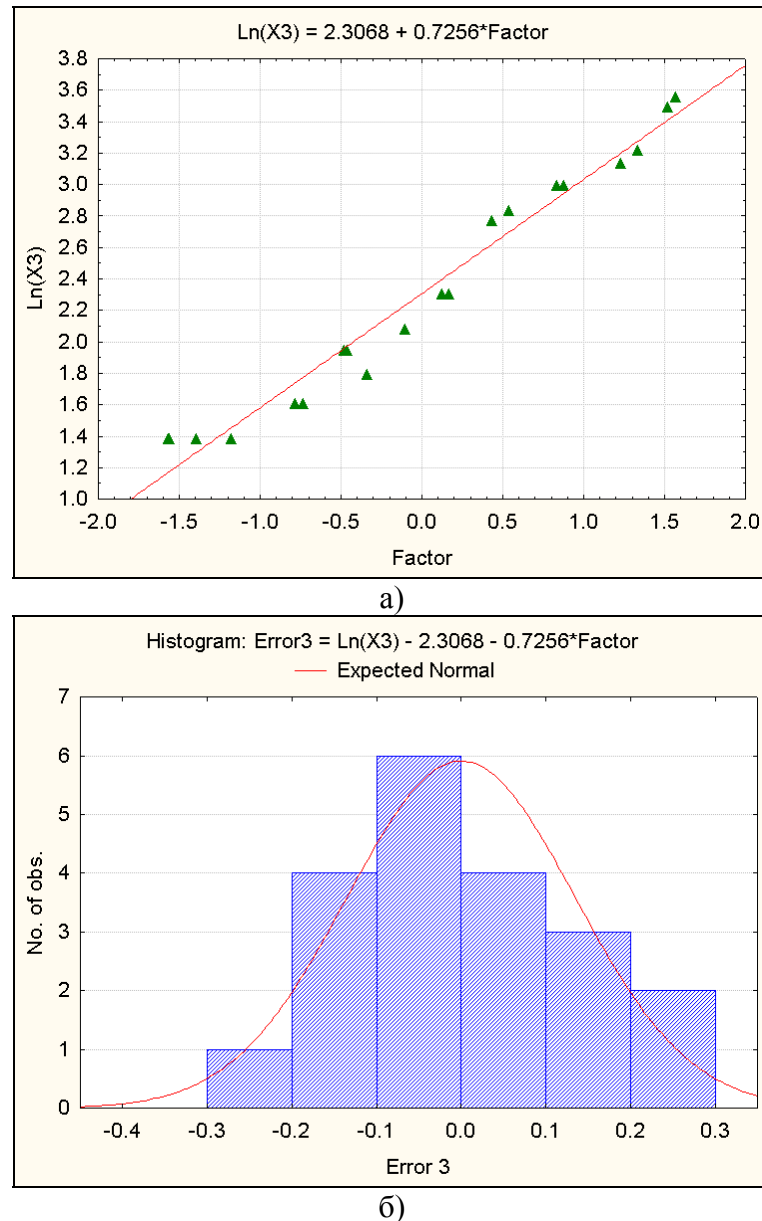
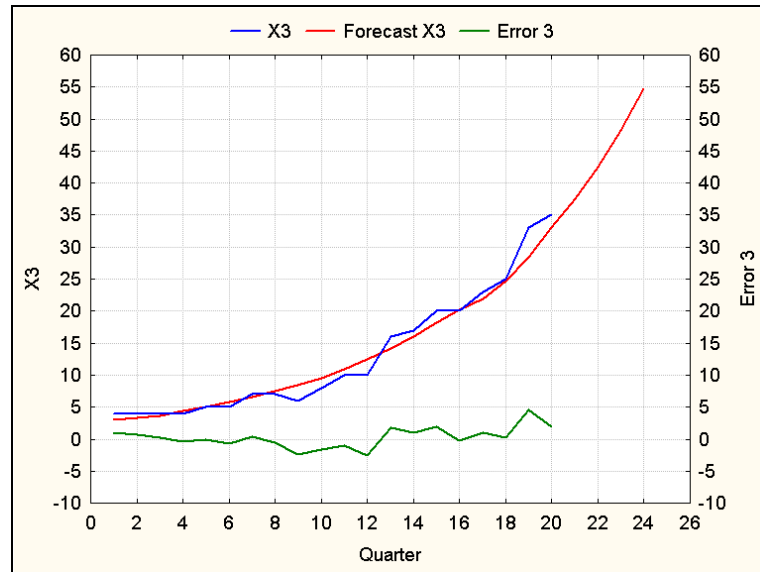


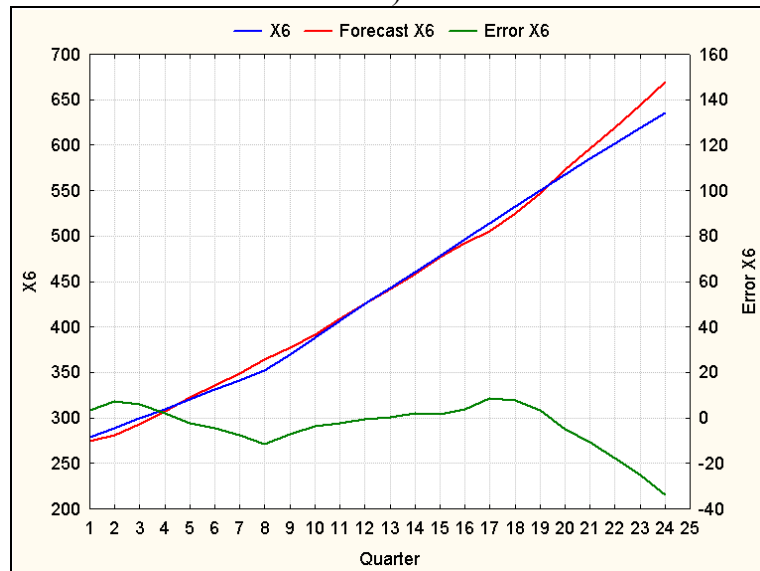
Рис. 3. Погрешности модели регрессии главного фактора на количество технических специалистов. Диаграмма рассеяния - а) и гистограмма ошибок аппроксимации – б).

Показательные зависимости (3) исходных переменных X_1, \dots, X_6 от плановых средних объемов продаж $Trend$ позволяют формировать необходимые решения по стимуляции бизнеса. Например, для того, чтобы сохранить экспоненциальный темп роста средних доходов компании от продаж, планируемые на год вперед по модели Хольта [2], необходимо увеличение технических специалистов X_3 в количестве, представленном на рис. 4.а. Кроме того, средние объемы продаж X_6 по рынку анализируемой продукции в

последующий год должны расти в среднем в соответствии с зависимостью, представленной на рис. 4.б.



а)



б)

Рис. 4. Планируемые значения факторов для сохранения роста объемов продаж компании. Количество технических специалистов - а) и средние по рынку объемы продаж – б).

Статистический анализ показал, что сезонная составляющая практически не коррелирована с предикторами X_1, \dots, X_6 . В частности, наибольшие абсолютные величины коэффициентов взаимной корреляции сезонной компоненты с исходными переменными, соответствующие временному лагу (-1), не превышают значения 0,23. Иными словами, построение регрессионной модели предикторов на сезонную составляющую не представляется возможным.

SSA- прогноз [3] сезонной компоненты для доходов от продаж компании реализован с помощью линейной рекуррентной формулы (ЛРФ) девятого порядка

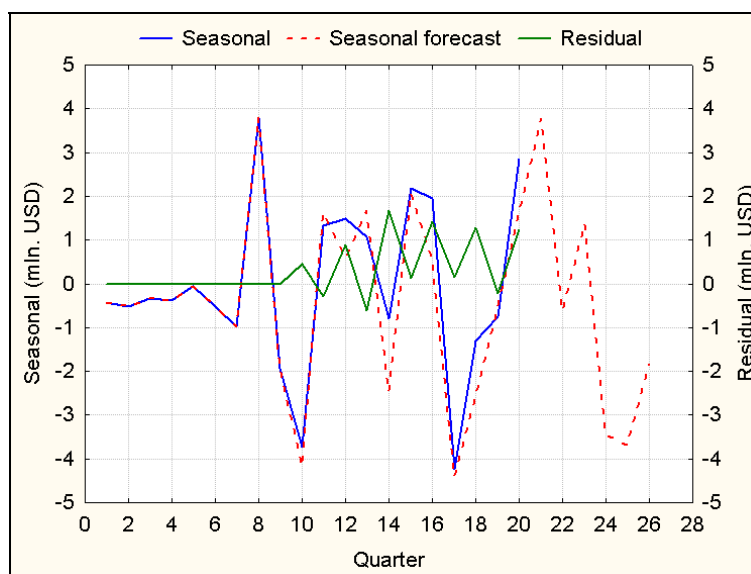
$$Seasonal(q) = \sum_{k=1}^9 C(k) Seasonal(q-k).$$

Оптимальные коэффициенты ЛРФ $C(k)$, обеспечившие наименьшую ошибку аппроксимации сезонной компоненты за 9 первых кварталов, представлены в таблице 1.

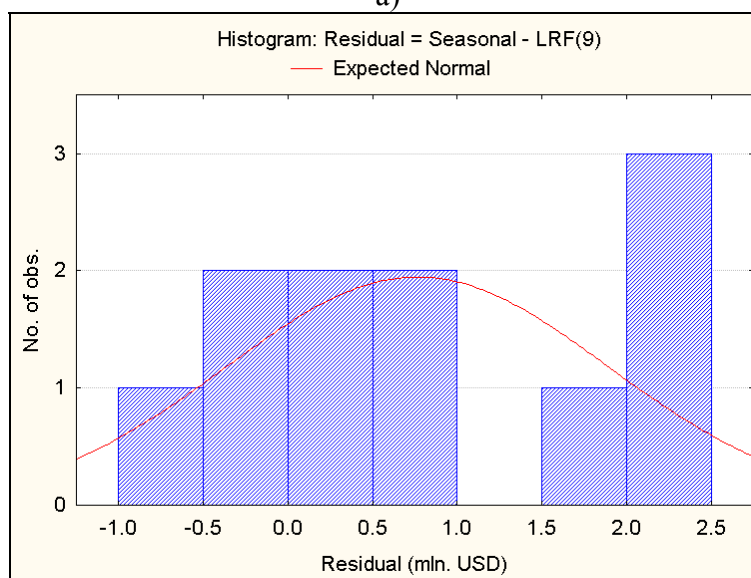
Таблица 1. Коэффициенты ЛРФ для сезонной компоненты.

C(1)	C(2)	C(3)	C(4)	C(5)
0.263991	-1.060703	0.052788	-0.991987	0.179521
C(6)	C(7)	C(8)	C(9)	
-0.997894	0.908010	-0.273092	0.638550	

Результаты структурного прогноза сезонной составляющей для объемов продаж и его погрешности иллюстрирует рис. 5.



а)

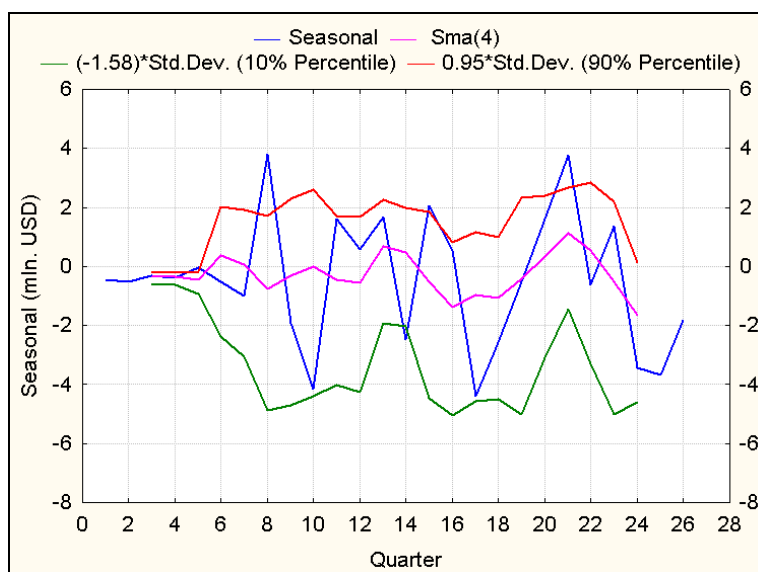


б)

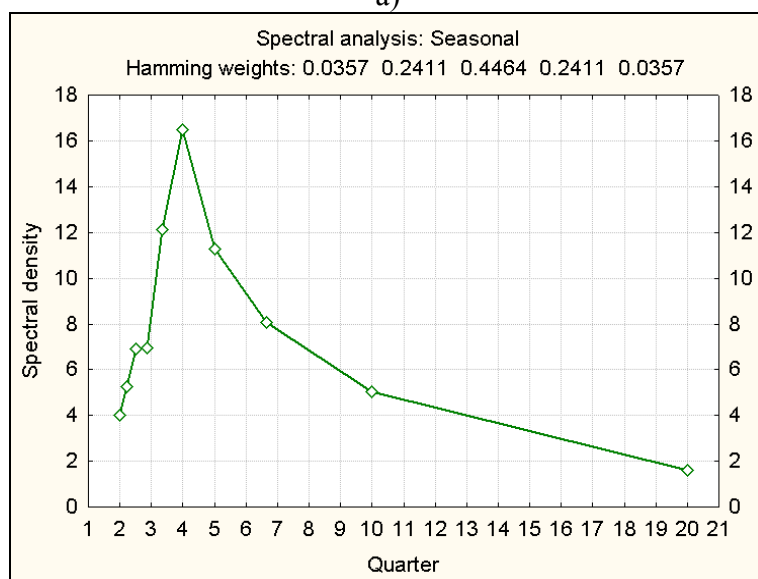
Рис. 5. Прогнозирование сезонной компоненты для доходов от продаж компании SSA- прогноз – а) и гистограмма ошибок прогноза – б).

Нижнюю и верхнюю огибающую структурного прогноза сезонной составляющей для объемов продаж удобно трактовать как меру разброса возможных доходов компании относительно их средних значений. Иными словами линии поддержки и сопротивления сезонной компоненты являются рациональной основой для формирования доверительных интервалов при прогнозировании объемов продаж. На рис. 6.а показаны 10%-ая нижняя и

90%-ая верхняя огибающие, соответствующие отклонению сезонной компоненты от ее текущего среднего значения (тренда) с доверительной вероятностью 0,8. В качестве модели тренда сезонной компоненты, в данном случае, применялась простая скользящая средняя SMA(4) с интервалом сглаживания 4 квартала. Такой временной интервал отвечал основному периоду торговых циклов, что подтверждается спектральным анализом (рис. 6.б).



а)

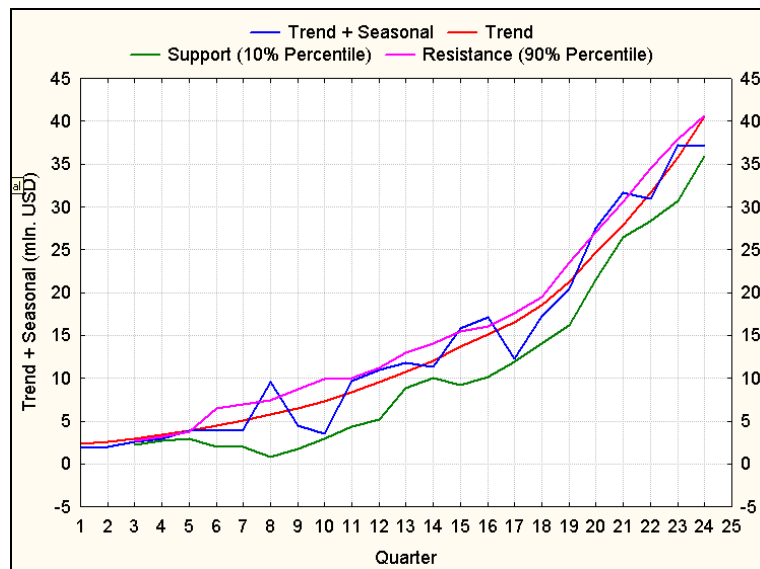


б)

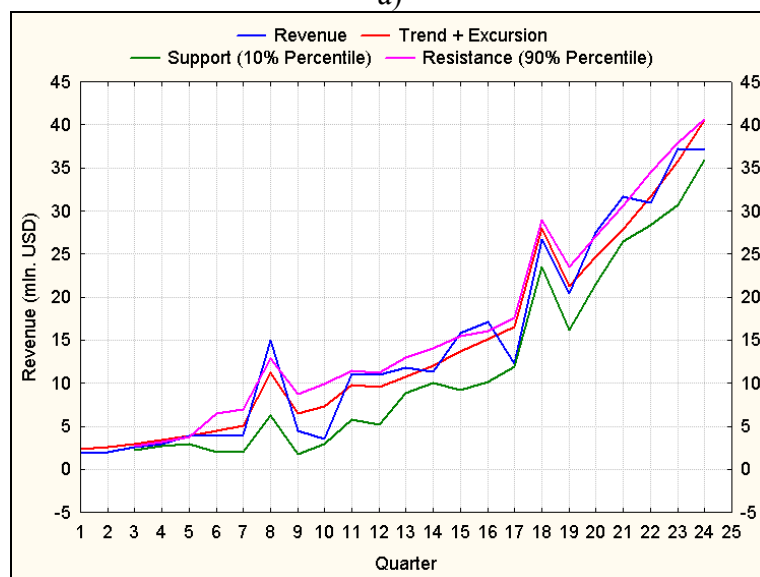
Рис. 6. Доверительный интервал для сезонной составляющей объемов продаж. Линии поддержки и сопротивления - а) и спектральная плотность мощности – б).

Доверительный интервал для прогноза объемов продаж компании на год вперед с доверительной вероятностью 0,8 представлен на рис. 7.

Важно отметить, что структурные методы прогнозирования, основанные на рациональном сочетании классических эконометрических моделей и сингулярного анализа данных относительно малого объема, позволяют получить адекватные оценки для доходов от продаж компании.



а)



б)

Рис. 7. Прогноз доходов от продаж компании на год вперед.
Без учета выбросов - а) и с учетом выбросов – б).

ЛИТЕРАТУРА

1. Голяндина Н. Э. Метод «Гусеница» - SSA: анализ временных рядов: Учебное пособие. – СПб.: С. Петербургский гос. университет, 2004. – 76 с.
2. Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов: Учебное пособие. – М.: Финансы и статистика, 2003. – 416 с.
3. Голяндина Н. Э. Метод «Гусеница» - SSA: прогноз временных рядов: Учебное пособие. – СПб.: С. Петербургский гос. университет, 2004. – 52 с.

Лабунец Н. Л. - ОАО Банк ВТБ, главный специалист департамента информационных технологий;
Лабунец Л. В. – д.т.н., профессор, зав. кафедрой ИТФР РосНОУ