**1 Описание лабораторной работы №3 «Прогнозирование временного ряда с помощью нейронных сетей»**

Цель работы заключается в выработке навыков прогнозирования временного ряда с помощью нейронных сетей.

Лабораторная работа включает следующие этапы:

* постановку задачи;
* ознакомление с порядком выполнения работы в ППП Statistica 10.0;
* выполнение индивидуальных заданий на компьютере и анализ результатов;
* сравнение с результатами, полученными в ходе лабораторной работы №1 8-го семестра обучения "Прогнозирование временного ряда на основе методов экспоненциального сглаживания"
* подготовку письменного отчета с выводами по работе;
* защиту лабораторной работы.

**2 Постановка задачи**

По данным Приложения А:

1. провести анализ компонентного состава временного ряда;
2. получить несколько моделей;
3. выбрать наилучшую модель с помощью нахождения относительной ошибки;
4. исследовать адекватность модели;
5. осуществить прогнозирование на 2 периода;
6. сравнить с результатами, полученными в ходе Лабораторной работы №1 "Прогнозирование временного ряда на основе методов экспоненциального сглаживания"

**3 Порядок выполнения работы в ППП Statistica**

Рассмотрим процедуру прогнозирования с помощью нейронных сетей, используя квартальную информацию о численности граждан, незанятых трудовой деятельностью, состоящих в службе занятости Оренбургской области () за период 1997-2006 гг.

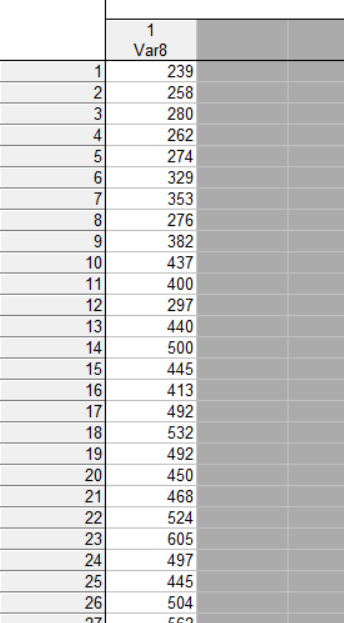


Рисунок 1. Исходные данные

Сначала необходимо посмотреть на динамику данного ряда.



Рисунок 2. Линейный график исходных данных

Видим, что график имеет явно возрастающий линейный тренд, а также имеет явную периодичность. Увеличим определенный отрезок графика чтобы точно определить периодичность

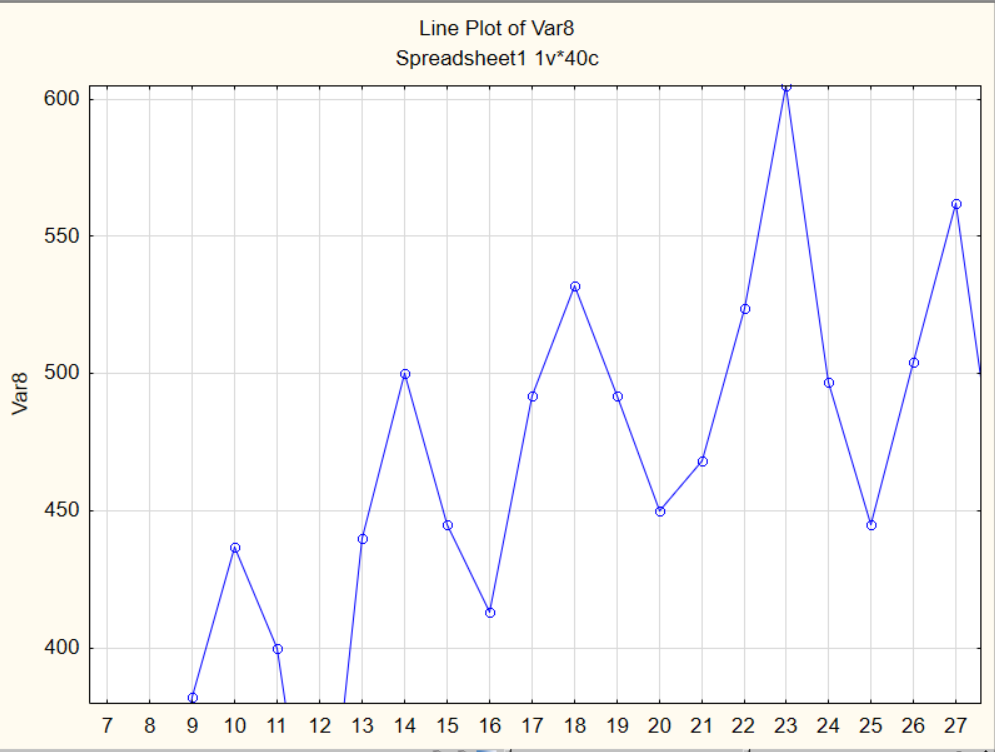


Рисунок 3. Увеличение графика

Из данного графика видно, что период равен 4-ем кварталам. Для подтверждения воспользуемся более аналитическим методом - спектральным анализом Фурье.

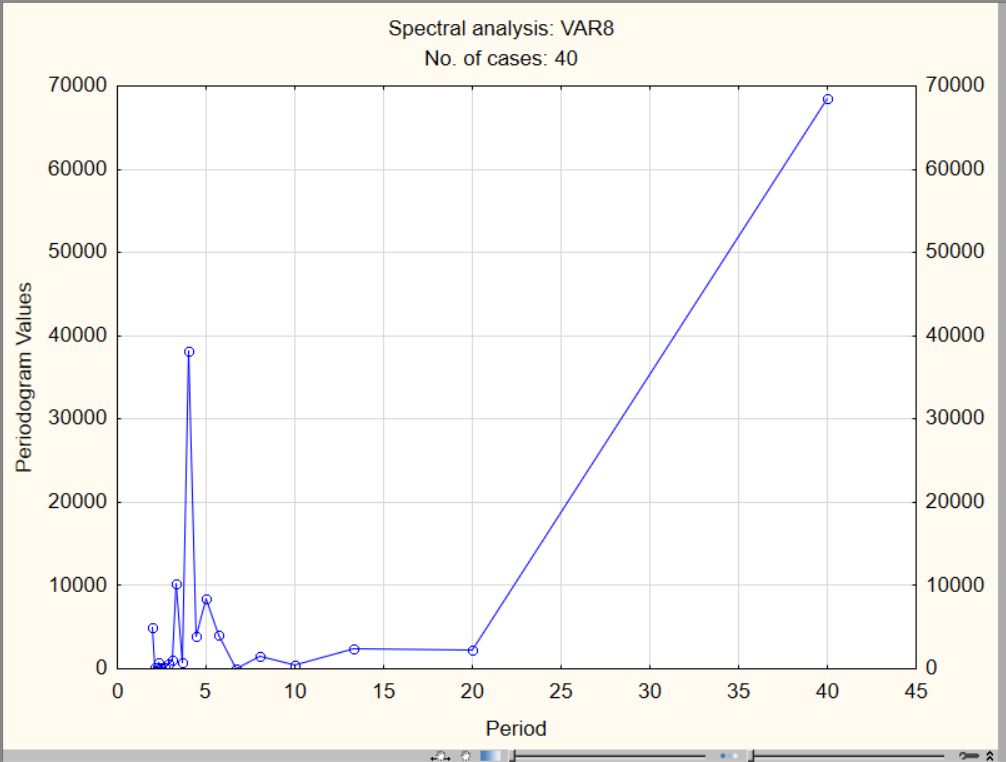


Рисунок 4. Периодограмма по спектральному одномерному анализу Фурье

На периодограмме наибольший пик будет соответствовать нашему периоду. Опять же с помощью увеличения графика можно убедиться, чему он равен.

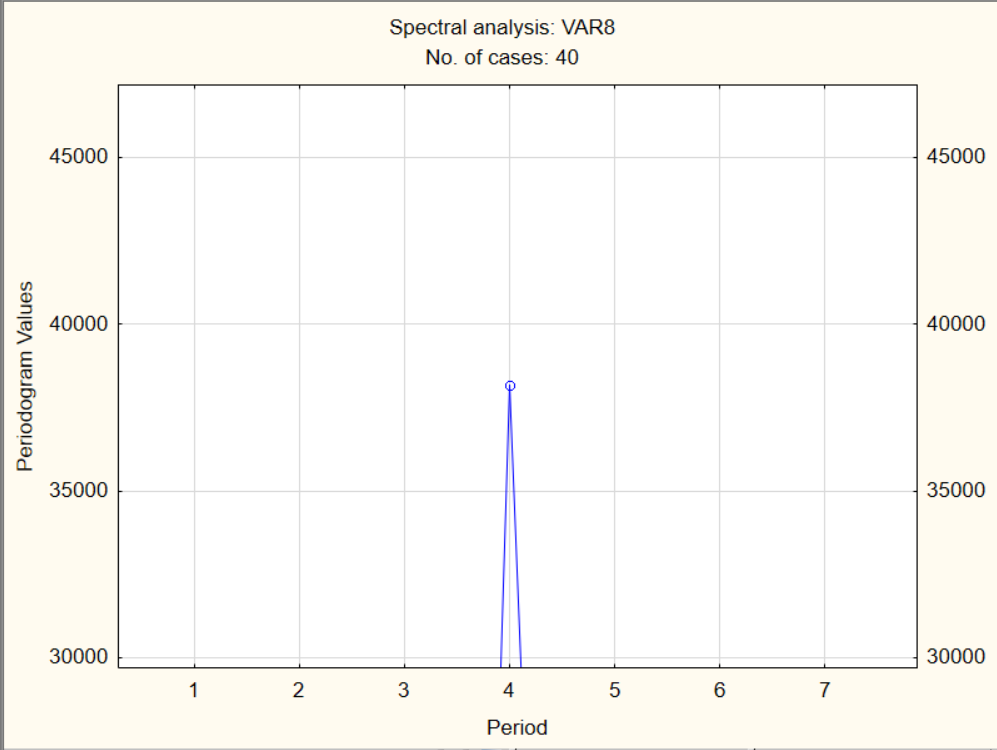


Рисунок 5. Определение периода

В данном случае видно, что период однозначно равен 4-ем кварталам.

Основная задача, которую мы хотим решить-построить нейросетевую модель, которая будет прогнозировать данный ряд, проверить качество модели за последний год и построить прогноз на более длительный срок.

Итак, спрогнозируем наш ряд с помощью нейронных сетей.

Рисунок 12. Выбор переменных

В качестве условия будет установлено что мы не будем использовать последние 4 наблюдения, то есть 4 квартала, оставив их для кросс-проверки. Заранее необходимо посмотреть количество наблюдений всего (их 40), поэтому мы будем использовать наблюдения с 1 по 36.

На контрольную (Test) зададим 30%. Контрольная выборка используется для остановки процесса обучения нейронной сети, то есть когда ошибка на контрольной выборке начинает возрастать- это служит признаком того, что наша модель уже переобучается и в этот момент необходимо прекращать обучение.

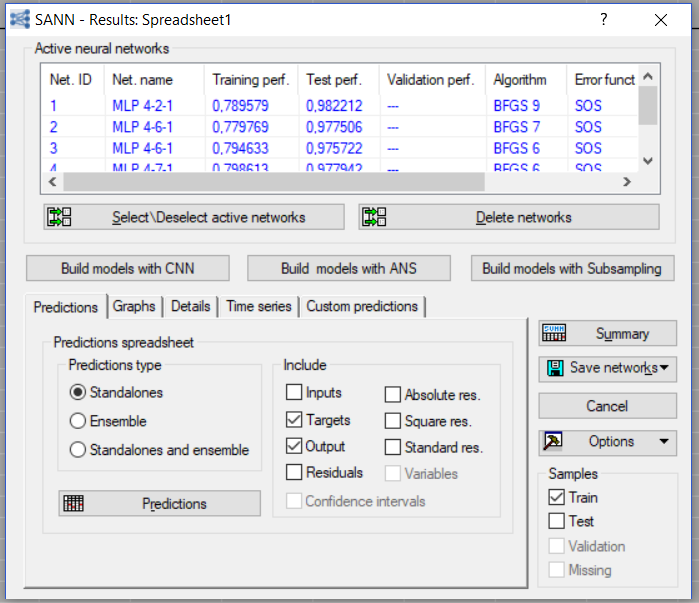


Рисунок 6. Результат обучения

Переходим к анализу результатов. П необходимо сделать-это выбрать наилучшую сеть. Сначала необходимо обратить внимание на описание сетей

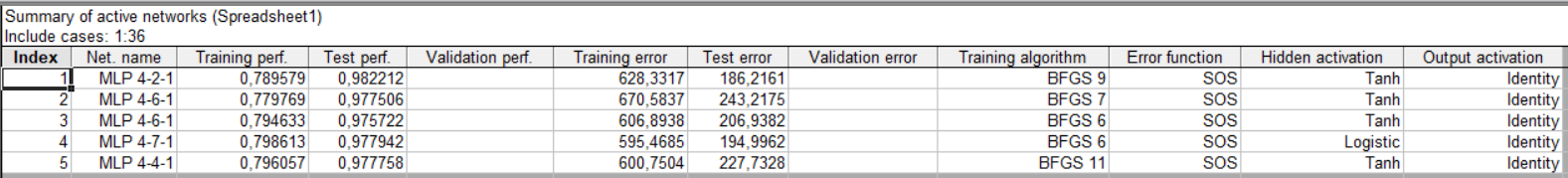


Рисунок 7. Описание сетей

Первый столбец означает архитектуру сети (Net. Name), то есть Многослойный персептрон - количество входов - количество скрытых нейронов - количество выходов.

Training perf. -Производительность - характеризуется величиной корреляции между исходным рядом и предсказанным на каждой подвыборке. Чем больше это значение, то есть чем ближе оно к 1, тем лучше.

Также необходимо обращать внимание, чтобы значения столбца Test perf. - Контрольная производительность, имели примерно одинаковые значения, чтобы не было резких отличий между ними. Тогда можно говорить о том, что модель равномерно работает, как на обучении, так и на контроле.

Также можно увидеть значение ошибки, как на обучении, так и на контроле. И функции активации.

Для оценки качества строим график рассеивания целевых (target) и выходных переменных (output). Получаем следующий результат:

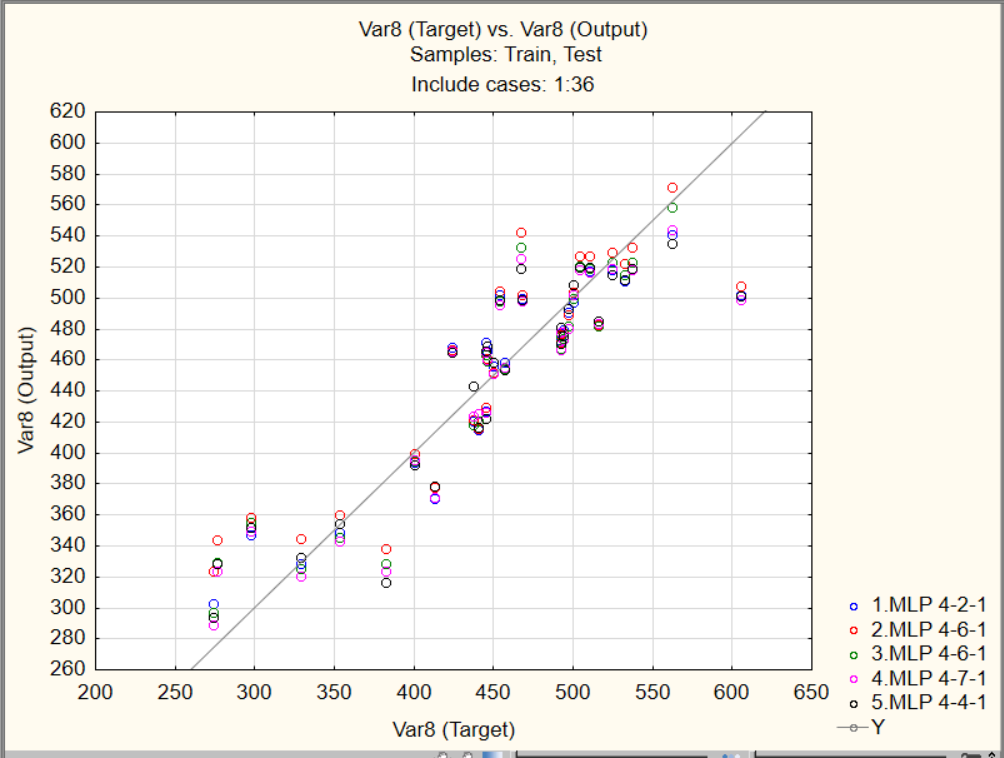


Рисунок 8. Диаграмма рассеяния

По данному графику можно оценить результаты обучения. Самой лучшей сетью является та, у которой точки лежат на прямой. В данном случае видно, что наилучший результат показывает 4-ая сеть (зеленые точки) и наихудший 5-яя сеть (черные точки), так как черные точки показывают наиболее яркие выбросы, то есть лежат удаленно от прямой.

Теперь посмотрим на сколько хорошо наши построенные модели прогнозируют исходный ряд.

Для этого построим следующий график:

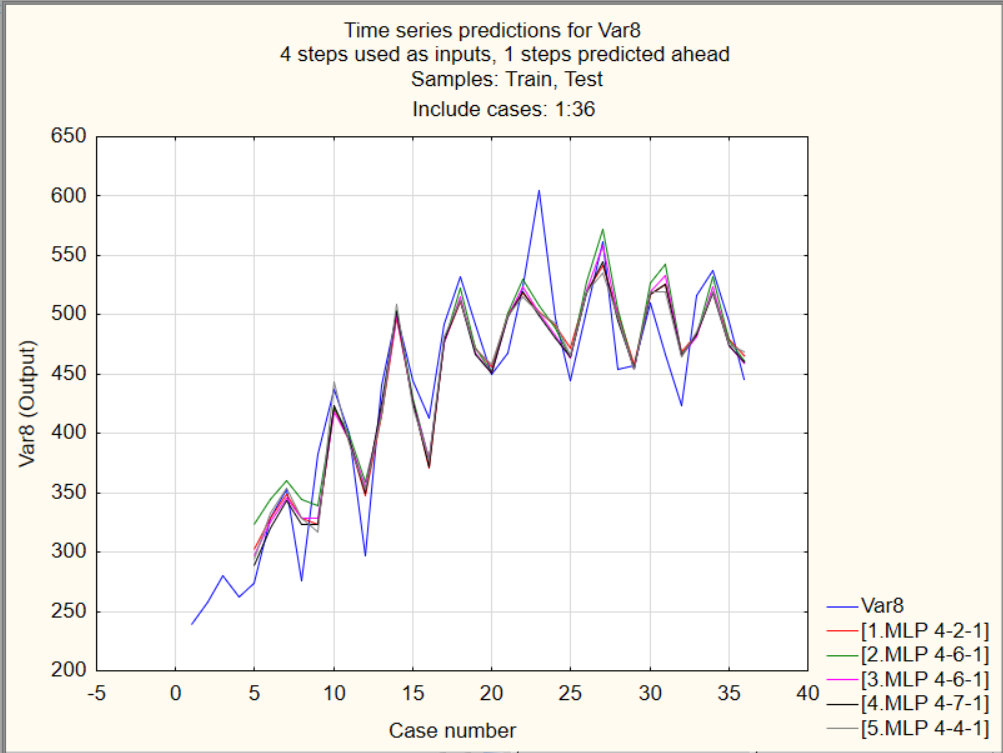


Рисунок 9. График временного ряда

На графике видим синим цветом исходный график и предсказанные ряды для каждой модели. В этом случае мы не можем посмотреть, как будет прогнозироваться этот ряд дальше. Построим теперь модель включающую прогнозы

Получаем следующий график показывающий прогнозы

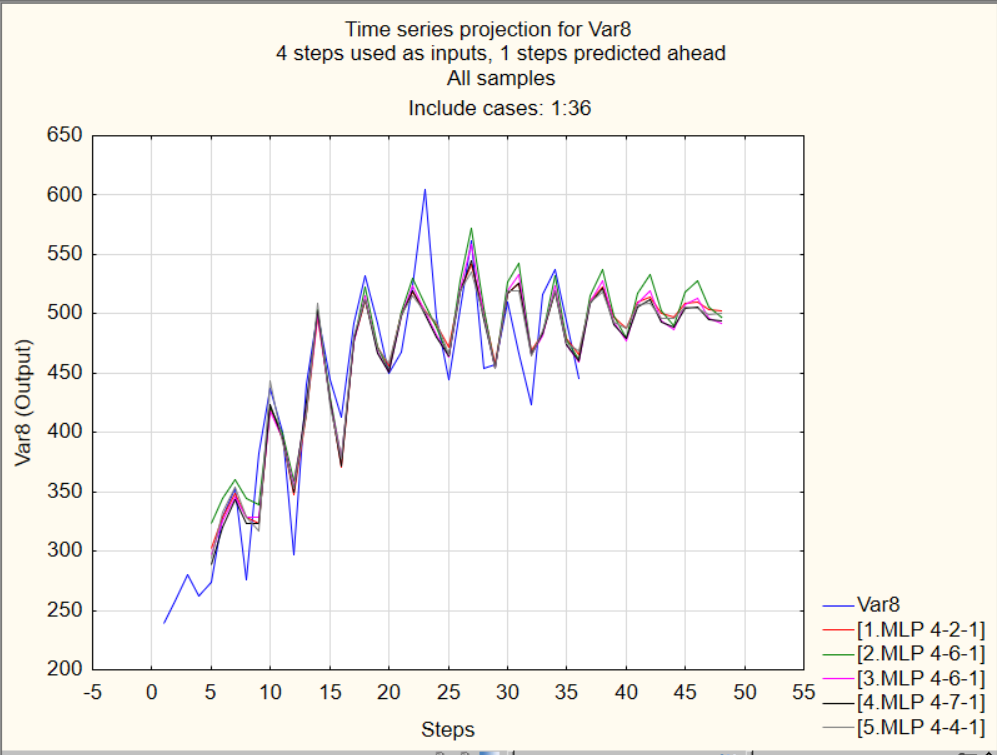
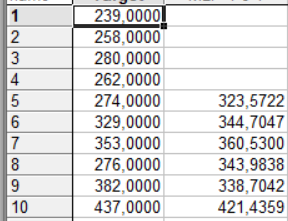


Рисунок 26. График проекции

Все модели улавливают зависимость

В данном случае видим, что наиболее оптимальной является 4-ая модель.

Далее можно смотреть и анализировать качество сети. Таким образом, теперь мы можем сравнить прогноз нашей построенной сети с тестовой выборкой, которую мы оставили в самом начале.



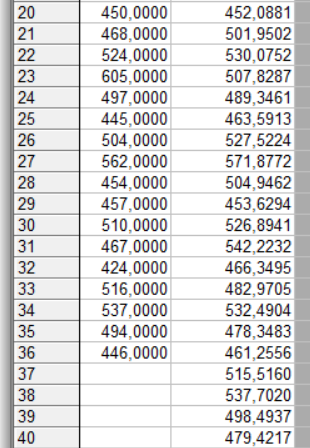


Рисунок 30. Прогнозируемые значения

Необходимо сравнить полученные значения с исходными

Получаем следующий результат:

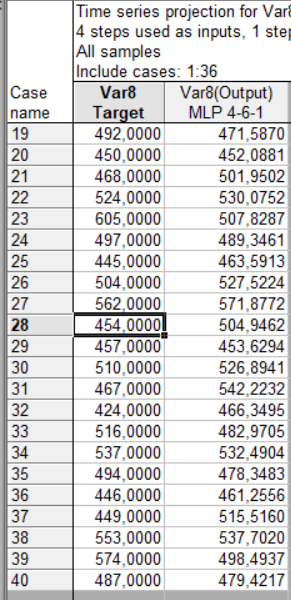


Рисунок 33. Вставка исходных данных

Теперь можно построить график, Построим линейный график

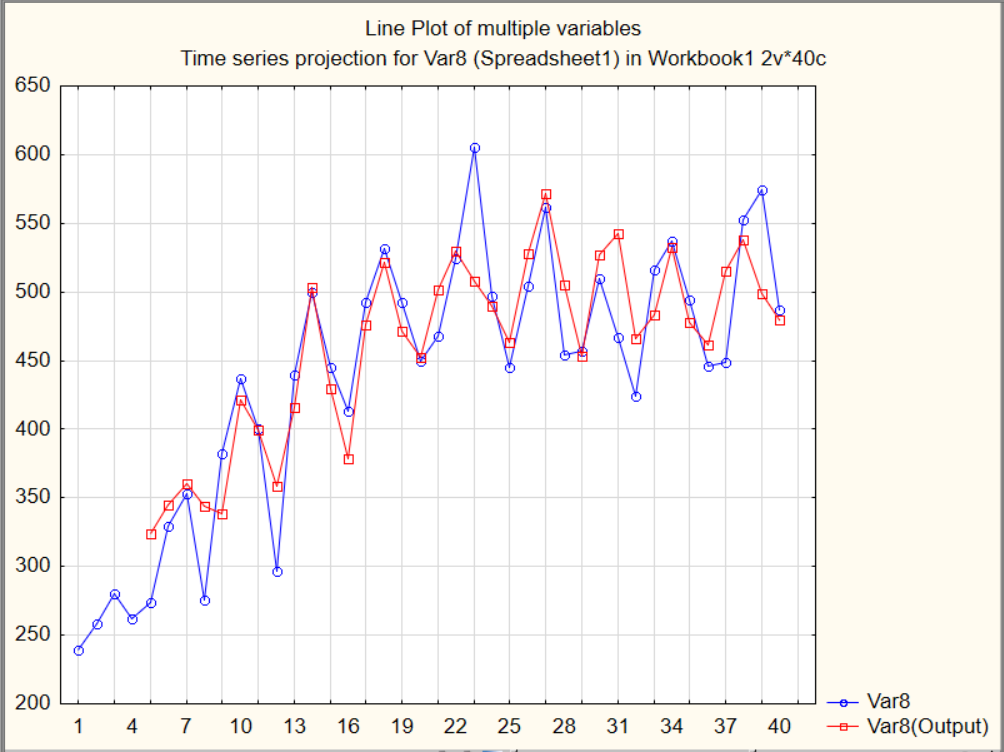


Рисунок 35. Линейные графики

Здесь тот же график, что и на рисунке 2, только ещё с учетом кросс-проверочной выборки. Мы видим, что качество прогноза оставляет желать лучшего, но связано это с тем, что мы провели недостаточное количество испытаний.

Попытаемся улучшить прогнозы путем построения более простых моделей. Построим дополнительные сети.

Возьмем количество скрытых нейронов равное 1 и в качестве функции активации оставим только **Тождественную**

Рисунок 37. Выбор параметров для линейной модели

Получаем ещё 5 моделей. Выбираем из них ту, у которой произвольность контрольной выборки (Test Pref.) наибольшая. А также оставляем предыдущую 1-ую модель.

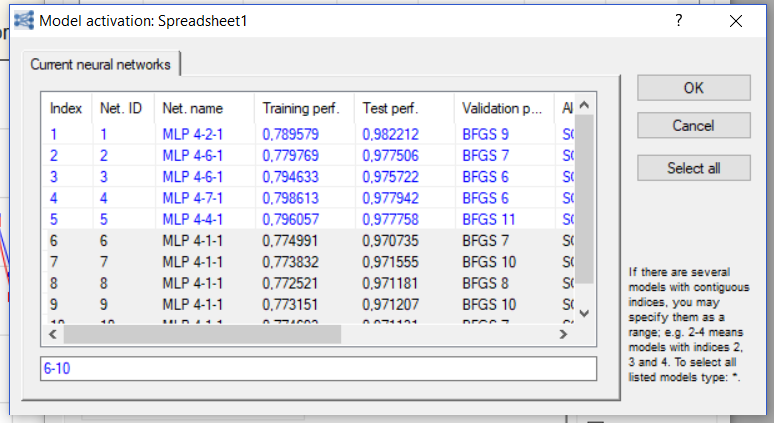


Рисунок 38. Выбор наилучших моделей

Теперь построим график проекции:

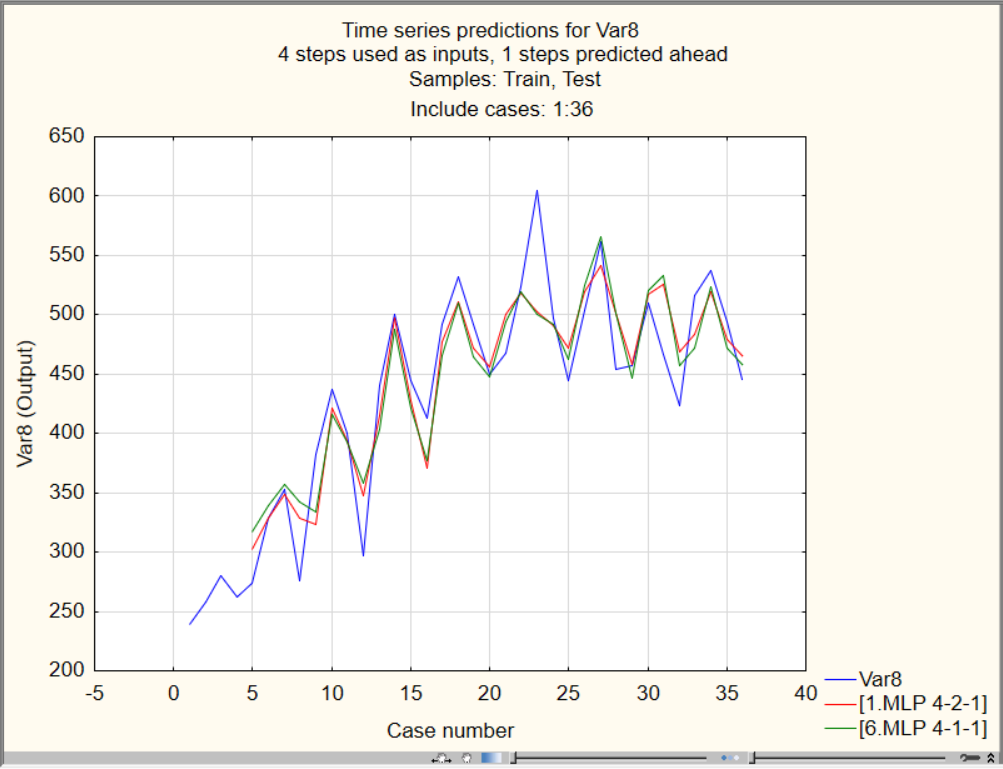


Рисунок 39. Графики проекций для старой и новой модели

. Построим следующую таблицу которая будет включать результаты по полученным двум моделям, а также рассчитаем их ошибки:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 449 | 511,12931 | 508,284601 | 0,138372628 | 0,132036973 |
| 553 | 522,22386 | 531,082991 | 0,0556530561 | 0,0396329277 |
| 574 | 496,768356 | 494,048617 | 0,134549902 | 0,139288124 |
| 487 | 487,323262 | 473,390832 | 0,000663782341 | 0,0279449035 |

Рисунок 45. Ошибки для моделей

Последним этапом остается найти среднее значение ошибок

Получаем таблицу результатов:

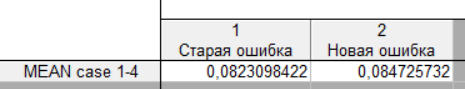


Рисунок 47. Средние относительные ошибки для старой и новой модели

Таким образом, видим, что вторая модель немного хуже первой, так как её ошибка составляет 0,084 (8,4%) в то время как для первой модели ошибка составляет 8,2%.

Следовательно, используем первую модель, чтобы сделать прогноз на два года вперед (8 кварталов).

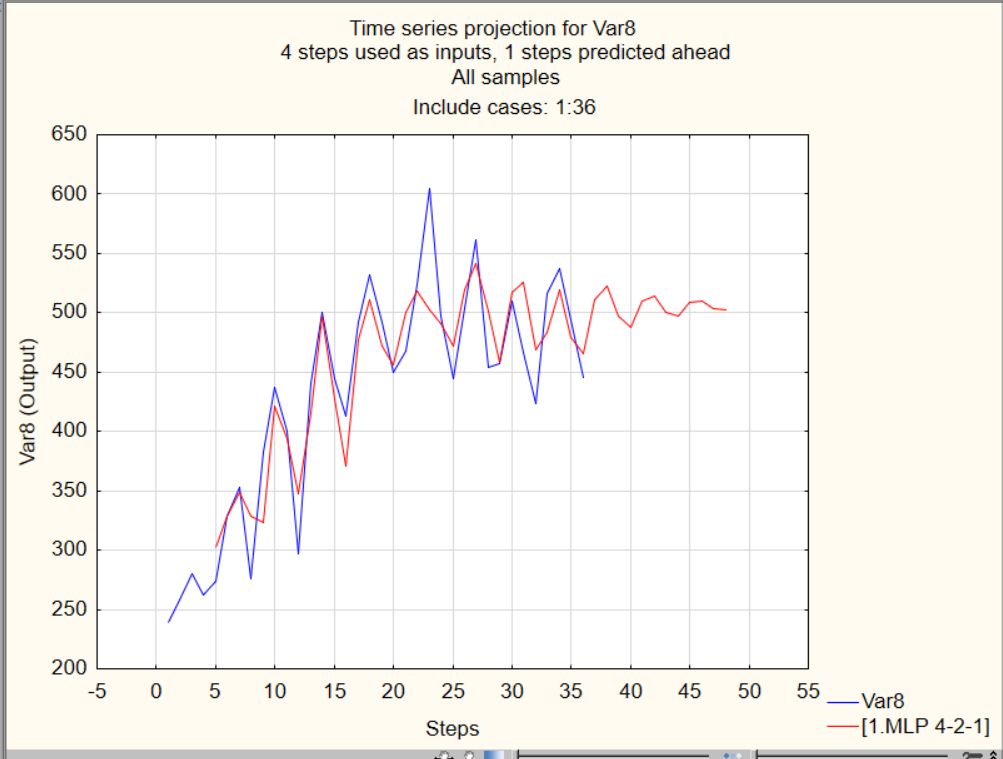


Рисунок 49. График проекции

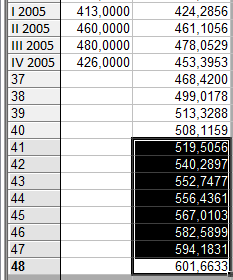


Рисунок 50. Полученные прогнозируемые значения

Теперь сравним данные значения с теми, которые были получены в Лабораторной работе №1 на основе метода экспоненциального сглаживания.

|  |  |  |
| --- | --- | --- |
| Период | Экспоненциальное сглаживание | Нейронные сети |
| I 2007 | 522,4472 | 509,6629 |
| II 2007 | 559,1767 | 514,1797 |
| III 2007 | 586,7978 | 500,4743 |
| IV 2007 | 548,0917 | 497,7515 |
| I 2008 | 573,6160 | 508,6560 |
| II 2008 | 610,3455 | 510,0674 |
| III 2008 | 637,9666 | 503,0847 |
| IV 2008 | 599,2605 | 502,5841 |

Видим что значения разнятся не существенно. Разница заметна лишь в 3-ем квартале в каждом году.

**Приложение А**

**(Обязательное)**

**Исходные данные для анализа**

Таблица А.1 – Выборочные данные по курсам ценных бумаг

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Кварталы | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 | x9 | x10 |
| I 1997 | 37,26 | 17,92 | 48,73 | 17,69 | 5 | 73,20 | 73,20 | 239 | 239,11 | 22,06 |
| II 1997 | 73,11 | 32,58 | 84,55 | 30,95 | 24 | 93,77 | 60,00 | 258 | 232,83 | 34,16 |
| III 1997 | 38,93 | 51,36 | 45,76 | 33,72 | 60 | 99,17 | 46,50 | 280 | 210,83 | 32,29 |
| IV 1997 | 70,39 | 113,09 | 60,38 | 32,16 | 32 | 64,40 | 70,20 | 262 | 262,04 | -29,99 |
| I 1998 | 10,99 | 137,24 | 35,70 | 93,02 | 53 | 123,83 | 54,40 | 274 | 227,14 | 2,23 |
| II 1998 | 66,99 | 151,01 | 87,01 | 86,57 | 97 | 64,98 | 52,00 | 329 | 245,74 | -0,14 |
| III 1998 | 64,81 | 157,07 | 85,94 | 101,94 | 137 | 150,11 | 49,10 | 353 | 233,99 | -1,09 |
| IV 1998 | 107,93 | 236,49 | 143,51 | 118,75 | 118 | 110,68 | 76,00 | 276 | 237,94 | 20,97 |
| I 1999 | 45,85 | 223,48 | 33,28 | 106,07 | 113 | 113,11 | 63,20 | 382 | 233,33 | 6,43 |
| II 1999 | 99,16 | 282,30 | 133,64 | 124,24 | 137 | 110,92 | 57,90 | 437 | 234,13 | 13,29 |
| III 1999 | 87,87 | 284,93 | 112,82 | 125,72 | 158 | 72,66 | 72,66 | 400 | 315,21 | 10,61 |
| IV 1999 | 88,45 | 328,65 | 72,74 | 126,31 | 126 | 95,85 | 91,20 | 297 | 212,45 | 1,20 |
| I 2000 | 63,85 | 406,55 | 103,65 | 134,17 | 134 | 119,01 | 78,90 | 440 | 211,86 | 14,68 |
| II 2000 | 105,66 | 367,87 | 124,23 | 157,67 | 157 | 114,24 | 72,80 | 500 | 266,59 | 4,43 |
| III 2000 | 105,62 | 370,08 | 124,49 | 173,40 | 189 | 120,64 | 66,70 | 445 | 249,84 | -4,29 |
| IV 2000 | 129,29 | 430,08 | 170,27 | 158,34 | 158 | 86,98 | 100,00 | 413 | 295,14 | -28,64 |
| I 2001 | 132,70 | 395,57 | 161,13 | 230,67 | 211 | 135,56 | 89,50 | 492 | 249,39 | 15,39 |
| II 2001 | 137,93 | 428,27 | 181,21 | 235,01 | 253 | 132,91 | 72,80 | 532 | 292,30 | -23,44 |
| III 2001 | 137,84 | 490,12 | 150,95 | 213,47 | 263 | 121,16 | 93,90 | 492 | 262,41 | 5,84 |
| IV 2001 | 158,59 | 502,39 | 197,75 | 285,03 | 240 | 112,34 | 112,34 | 450 | 311,71 | 30,60 |
| I 2002 | 154,43 | 528,37 | 176,88 | 199,19 | 284 | 119,13 | 102,60 | 468 | 272,09 | 46,37 |
| II 2002 | 174,30 | 592,22 | 199,82 | 268,18 | 316 | 126,83 | 86,80 | 524 | 256,52 | 24,95 |
| III 2002 | 183,70 | 594,53 | 244,20 | 287,37 | 316 | 102,94 | 102,94 | 605 | 286,45 | -0,86 |
| IV 2002 | 162,16 | 599,75 | 189,02 | 292,67 | 292 | 112,14 | 128,90 | 497 | 272,77 | 28,89 |
| I 2003 | 191,31 | 625,82 | 208,40 | 307,94 | 307 | 138,96 | 111,40 | 445 | 260,80 | 28,40 |
| II 2003 | 226,16 | 681,13 | 211,19 | 310,86 | 353 | 103,24 | 93,00 | 504 | 316,45 | 31,72 |
| III 2003 | 262,08 | 731,02 | 258,91 | 347,52 | 397 | 160,03 | 116,70 | 562 | 303,12 | 31,21 |
| IV 2003 | 291,44 | 694,18 | 321,57 | 355,10 | 355 | 138,28 | 138,28 | 454 | 355,63 | 23,02 |
| I 2004 | 272,92 | 745,36 | 293,26 | 357,05 | 380 | 98,78 | 116,00 | 457 | 354,30 | 12,64 |
| II 2004 | 302,23 | 790,22 | 333,60 | 348,87 | 426 | 149,00 | 104,00 | 510 | 380,25 | 13,69 |
| III 2004 | 323,66 | 772,85 | 283,81 | 389,70 | 460 | 124,27 | 124,27 | 467 | 362,29 | -4,33 |
| IV 2004 | 382,30 | 869,02 | 375,38 | 383,26 | 383 | 148,74 | 148,74 | 424 | 402,19 | 13,62 |
| I 2005 | 377,94 | 871,44 | 363,82 | 402,61 | 413 | 142,25 | 93,00 | 516 | 368,14 | 46,50 |
| II 2005 | 409,12 | 878,85 | 404,27 | 432,87 | 460 | 170,11 | 86,80 | 537 | 375,11 | 10,23 |
| III 2005 | 418,23 | 898,02 | 434,47 | 432,19 | 480 | 157,76 | 104,00 | 494 | 403,62 | -5,08 |
| IV 2005 | 475,85 | 932,06 | 482,71 | 426,01 | 426 | 133,88 | 153,50 | 446 | 446,38 | -4,35 |
| I 2006 | 452,06 | 938,44 | 479,89 | 468,04 | 468 | 154,38 | 124,60 | 494 | 407,30 | 13,92 |
| II 2006 | 528,85 | 944,70 | 542,70 | 470,03 | 505 | 141,17 | 100,00 | 553 | 431,81 | 13,03 |
| III 2006 | 561,82 | 1013,30 | 589,11 | 483,11 | 555 | 158,00 | 136,80 | 574 | 502,04 | 78,10 |
| IV 2006 | 582,46 | 1052,46 | 607,87 | 495,30 | 495 | 178,50 | 185,10 | 487 | 487,48 | 11,39 |

х1 - численность занятого в экономике населения, тыс.человек;

х2 - общая численность безработных (на конец периода), тыс.человек;

х3 - численность официально зарегистрированных в службе занятости безработных (на конец периода), тыс.человек;

х4 - из них получают пособие по безработице;

х5 - численность граждан, незанятых трудовой деятельностью, состоящих в службе занятости, человек;

х6 - потребность предприятия в работниках, заявленная в службу занятости, человек;

х7 - нагрузка незанятого населения на одну заявленную вакансию;

х8 - номинальная среднемесячная зарплата на 1 работника, руб.;

х9 - среднемесячные денежные доходы в расчете на душу населения, тыс.рублей;

х10 - число выбывших, человек.