Федеральное государственное образовательное бюджетное учреждение высшего образования

«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ РФ»

Департамент анализа данных и машинного обучения

Отчет по практике №5

по дисциплине «эконометрика»

Студента группы ПМ23-1

Факультета информационных технологий и анализа больших данных

Тищенко И.С.

Преподаватель

Кудрявцев К.Н.

Москва 2025

1. Введение

В рамках работы исследовались временные ряды, характеризующие две переменные: численность кроликов (R) и численность лис (F). Данные загружались из файла Excel (fox_rabbit.xls) и содержали 1000 наблюдений для каждой группы. Целью исследования являлось построение векторной авторегрессии (VAR) с оптимальным порядком лага, оценка параметров модели, проверка значимости оцененных коэффициентов, получение доверительных интервалов, прогнозирование будущих значений и анализ динамики системы посредством импульсного отклика.

2. Подготовка данных и построение VAR-модели

2.1. Импорт библиотек и загрузка данных

Вначале были импортированы необходимые библиотеки:
- pandas и numpy – для обработки данных,
- scipy.stats – для вычисления квантилей нормального распределения при построении доверительных интервалов,
- matplotlib – для визуализации результатов,
- IPython.display – для форматированного вывода уравнений в LaTeX.

Далее данные были загружены из Excel-файла:

Python:
```
data = pd.read_excel(os.path.join(data_dir, 'fox_rabbit.xls'), header=0)
```

После загрузки набор данных имел размерность (1000 x 2) с переменными fox и rabbit. Значения временных рядов были разделены:

Python:
```
R_data = data.rabbit.values
F_data = data.fox.values
```

2.2. Формирование матриц для VAR-модели

Использовалась функция prepare_var_data(R, F, p), которая на основании заданного порядка лага p формирует:
- X – матрицу объясняющих переменных, состоящую из лаговых значений временных рядов (первые p наблюдений для R и F) и константы,
- Y – матрицу зависимых переменных для текущих значений R и F.

Это позволяет сформулировать систему уравнений вида:
$Y = XA,$
где A – матрица коэффициентов размером $(2p + 1) \times 2$.

2.3. Оценка модели методом наименьших квадратов (OLS)

Функция fit_var_ols(X, Y) осуществляет оценку параметров модели. Расчет производился по формуле:

$$\hat{A} = \left(X^\top X\right)^{-1} X^\top Y,$$

после чего были вычислены остатки E и оценена ковариационная матрица ошибок:

$$\hat{\Sigma} = \frac{E^\top E}{T - p}.$$

Данные результаты являются основой для дальнейшего анализа значимости коэффициентов и формировании доверительных интервалов.

2.4. Выбор оптимального порядка лага

Для того чтобы определить лучший порядок лага, был использован критерий Байеса (BIC). Функция compute_bic вычисляет BIC по формуле:

$$BIC = (T - p) \cdot \ln\left(\det(\hat{\Sigma})\right) + k \cdot \ln(T - p),$$

где k – число оцениваемых параметров (в данном случае $2(2p+1)$). Функция var_select_order(R, F, p_{\text{max}}) перебирает лаги от 1 до заданного максимума и определяет оптимальный порядок p_{opt}, при котором значение BIC минимально.

В отчете выводятся значения BIC для различных лагов и оптимальный порядок, что позволяет обеспечить сбалансированность между качеством подгонки модели и количеством параметров.

3. Оценка параметров модели и интерпретация результатов

3.1. Итоговая оценка модели

После выбора оптимального лага модель переоценивается для получения итоговой матрицы коэффициентов $\hat{A}$. Из матрицы коэффициентов были выделены два уравнения:

- Для уравнения, описывающего динамику кроликов R_t:

$$R_t = \underbrace{(0.4938)\,R_{t-1}}_{\text{lag R}} + \underbrace{(-0.5127)\,F_{t-1}}_{\text{lag F}} + \underbrace{(10.2148)}_{\text{constant}},$$

- Для уравнения, описывающего динамику лис F_t:

$$F_t = \underbrace{(0.4729)\,R_{t-1}}_{\text{lag R}} + \underbrace{(0.5192)\,F_{t-1}}_{\text{lag F}} + \underbrace{(3.8944)}_{\text{constant}}.$$

Эти результаты были визуализированы с использованием функции display(Math(...)), что позволило отобразить уравнения в удобном LaTeX-формате.

3.2. Проверка значимости коэффициентов

Для оценки статистической значимости коэффициентов были вычислены стандартные ошибки и соответствующие t-статистики для каждого оцененного параметра. Стандартные ошибки определялись как корень из диагональных элементов матрицы $\mathrm{Var}(\hat{A}) = \sigma^2 \cdot (X^\top X)^{-1}$ для каждого уравнения (с использованием соответствующих элементов диагонали оцененной ковариационной матрицы ошибок).

Оцененные t-значения сравнивались с критическим значением (примерно 2.58 при уровне значимости 1% для большого числа степеней свободы), что позволяло определить, какие коэффициенты статистически значимы.

3.3. Построение доверительных интервалов

На основе стандартных ошибок и квантиля нормального распределения ($z_{0.995} \approx 2.575829$) были рассчитаны 99%-ные доверительные интервалы для каждого коэффициента обоих уравнений. Это дало возможность оценить диапазон значений, в котором с высокой вероятностью находится истинное значение параметра.

4. Прогнозирование и анализ динамики системы

4.1. Прогнозирование будущих значений

Функция var_forecast использовалась для итеративного построения прогноза на 1, 2 и 3 шага вперёд. На каждом шаге прогноз строился с использованием уже известных (или ранее спрогнозированных) значений и итоговой матрицы коэффициентов $\hat{A}$. Итоговые прогнозные значения для R и F были выведены и представлены в отчете.

4.2. Определение равновесного состояния (стационарного уровня)

Для поиска стационарного равновесного состояния системы были агрегированы коэффициенты из оцененной матрицы:
- A_{11} – суммарный коэффициент при лаговых значениях R для уравнения R_t,
- A_{12} – суммарный коэффициент при лаговых значениях F для уравнения R_t,
- A_{21} и A_{22} – аналогичные коэффициенты для уравнения F_t,
- c_R и c_F – константные члены для соответствующих уравнений.

Система уравнений для стационарного равновесия задавалась как:

$$\begin{cases} (1 - A_{11})R - A_{12}F = c_R \\ -A_{21}R + (1 - A_{22})F = c_F \end{cases}$$

Решение данной системы позволило найти равновесные уровни R и F для кроликов и лис соответственно.

4.3. Анализ импульсного отклика (IRF)

Импульсный отклик позволяет проанализировать, как система реагирует на внешнее возмущение в одном из уравнений. В работе был построен IRF для VAR(1)-модели:

- На начальном этапе к равновесному состоянию был добавлен единичный шок к R_t.
- Итеративно рассчитывалось дальнейшее поведение системы при помощи матрицы агрегированных коэффициентов A_{var} и постоянного вектора C.
- Полученные отклонения от равновесия для обеих переменных на горизонте прогноза (H = 10 шагов) были визуализированы в виде графика, позволяющего отследить динамику реакции системы.

5. Заключение

Выполненная работа включает комплексный анализ временных рядов двух взаимосвязанных переменных (численности кроликов и лис) с использованием VAR-модели. Ключевые этапы исследования можно суммировать следующим образом:

- Подготовка данных: Загрузка и предварительная обработка выборки.
- Оценка VAR-модели: Формирование регрессионных матриц, выбор оптимального порядка лага по критерию BIC и оценка матрицы коэффициентов с помощью метода ОЛС.
- Анализ результатов: Интерпретация оцененных коэффициентов, проведение проверки значимости, вычисление доверительных интервалов.
- Прогнозирование и динамика: Построение краткосрочных прогнозов, определение равновесного состояния системы и анализ импульсных откликов.