

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ
Н.Г.ЧЕРНЫШЕВСКОГО»**

Кафедра теории функций и стохастического анализа

Регрессионные модели динамики факторов экономического роста

АВТОРЕФЕРАТ

студентки _____ 2 _____ курса _____ 248 _____ группы

направления _____ 09.04.03 «Прикладная информатика» _____

_____ механико-математического факультета _____

_____ Лю Светланы Вадимовны _____

Научный руководитель
доцент, к.ф.-м. наук _____

_____ Новиков В.В. _____

Зав. кафедрой
доцент, д.ф.-м. наук _____

_____ Сидоров С.П. _____

Саратов 2019

ВВЕДЕНИЕ

Актуальность работы объясняется тем, что на основе результатов регрессионного анализа можно спрогнозировать «поведение» того или иного социально-экономического процесса, выявить наиболее значимые факторы, влияющие на него и принять продуманное, эффективное решение.

Работа посвящена построению регрессионных моделей конкретных экономических зависимостей средствами языка R и интерпретации полученных результатов. Рассмотрены основные конструкции языка R, используемые в регрессионном анализе, построены регрессионные модели зависимости валового регионального продукта от таких показателей, как численность экономически активного населения, инвестиции в основной капитал, производительность труда и др. Оценена статистическая значимость полученных моделей.

Целью работы является создание программы для анализа данных, сочетающую в себе необходимые для анализа функции и параметры.

Для создания данной программы перед нами стоят следующие задачи:

- поиск пакета для создания интерфейса;
- разработка приложения для анализа данных;
- поиск и формирование таблицы данных для анализа;
- анализ полученных данных и вывод.

При работе над магистерской работой использовался статистический язык программирования R и пакет Shiny.

Данная работа состоит из двух разделов:

- Математические аспекты эконометрического моделирования
- Моделирование экономического роста в среде программирования R

Работа прошла апробацию на различных конференциях, в частности,

- в XIX Международной Саратовской зимней школе «Современные проблемы теории функций и их приложения», посвященной 90-летию со дня рождения академика П. Л. Ульянова, январь 2018 года
- на ежегодной студенческой конференции "Актуальные проблемы математики и механики которую проводил механико-математический факультет СГУ в апреле 2019 года, в секции «Анализ данных»,
- в VII Международной молодежной научно-практической конференции

«Математическое и компьютерное моделирование в экономике, страховании и управлении рисками», ноябрь 2018 года.

Основное содержание работы

Первый раздел включает в себя математические аспекты эконометрического моделирования, а именно: этапы эконометрического моделирования, типы моделей, типы данных. Так же описание регрессионного анализа: модель парной и множественной регрессии.

Эконометрика — наука, изучающая количественные и качественные экономические взаимосвязи с помощью математических и статистических методов и моделей.

Объект изучения эконометрики — экономико-математические модели, которые строятся с учетом случайных факторов. Такие модели называются эконометрическими моделями.

Основным предметом исследования эконометрики являются массовые экономические явления и процессы [1].

Можно выделить несколько этапов эконометрического моделирования [2].

1. Постановочный. Можно выделить следующие цели эконометрического исследования:
 - анализ изучаемого экономического процесса (явления, объекта);
 - моделирование поведения процесса при различных значениях независимых (факторных) переменных;
 - прогноз экономических показателей, характеризующих изучаемый процесс;
 - выработка управленческих решений. Факторные переменные не должны быть связаны функциональной или тесной корреляционной связью, присутствие в модели условия мультиколлинеарности может привести к негативным последствиям всего процесса моделирования.
2. Априорный.
3. Параметризация. Помимо этого, на этапе моделирования решается задача спецификации модели путем:

- аппроксимации математической формы выявленных связей и соотношений между переменными;
- определения зависимых и независимых переменных;
- формулировки исходных предпосылок и ограничений модели.

4. Информационный.

5. Идентификация модели.

6. Оценка качества модели.

7. Интерпретация результатов моделирования. Среди наиболее известных эконометрических моделей можно выделить:

- модели потребительского и сберегательного потребления;
- модели взаимосвязи риска и доходности ценных бумаг;
- модели предложения труда;
- макроэкономические модели (модель роста);
- модели инвестиций;
- маркетинговые модели;
- модели валютных курсов и валютных кризисов и др.

Основными задачами эконометрики являются:

- выявление связей между количественными характеристиками экономических объектов в целях построения математических правил прогноза
- определить значения всех числовых параметров, входящих в модель и обеспечить соответствие ее реальному поведению объекта
- получение наилучших оценок параметров экономико-математических моделей, конструируемых в прикладных целях;
- проверка теоретико-экономических положений и выводов на фактическом (эмпирическом) материале;
- создание универсальных и специальных методов для обнаружения статистических закономерностей в экономике.

Можно выделить три основных класса моделей, которые применяются в эконометрике для анализа и/или прогноза:

I. Модели временных рядов:

1. модели тренда: $y(t) = T(t) + \epsilon_i$, где $T(t)$ - временной тренд (например, линейный $T(t) = a + bt$), ϵ_i - случайная компонента;
2. модели сезонности: $y(t) = S(t) + \epsilon_i$; где $S(t)$ - периодическая сезонная

компонента, ϵ_i , – случайная компонента

3. тренда и сезонности: $y(t) = T(t) + S(t) + \epsilon_i$ (аддитивная модель);

При моделировании экономических процессов встречаются два типа исходных данных:

- данные, характеризующие совокупность различных объектов в определенный момент (период) времени;
- данные, характеризующие один объект за ряд последовательных моментов (периодов) времени.

Модель парной регрессии:

Предположим есть набор значений двух переменных $X_t, Y_t, t = 1, \dots, n$; можно отобразить пары (X_t, Y_t) точками на плоскости $X - Y$ в соответствии с рисунком (1)

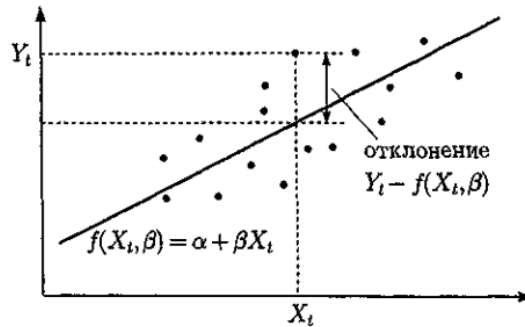


Рисунок 1 – График функции двух переменных $X_t, Y_t, t = 1, \dots, n$;

Надо подобрать функцию $Y = f(X)$ из параметрического семейства функций $f(X, \beta)$, «наилучшим» способом описывающую зависимость Y от X . Подобрать функцию в данном случае означает выбрать «наилучшее» значение параметра β [?]. (Примером параметрического семейства может служить семейство линейных функций $f(X, \beta) = \alpha + \beta X$.)

В качестве меры отклонения функции $f(X, \beta)$ от набора наблюдений можно взять:

1. сумму квадратов отклонений $F = \sum_{t=1}^n (Y_t - f(X_t, \beta))^2$,
2. сумму модулей отклонений $F = \sum_{t=1}^n |Y_t - f(X_t, \beta)|^2$, или, в общем случае
3. $F = \sum_{t=1}^n g(Y_t - f(X_t, \beta))$, где g – «мера», с которой отклонение $Y_t - f(X_t, \beta)$ входит в функционал F .

Модель множественной регрессии:

Естественным обобщением линейной регрессионной модели с двумя переменными является многомерная регрессионная модель (multiple regression model), ИЛИ модель множественной регрессии:

$$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \epsilon_t, \quad t = 1, \dots, n,$$

или

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \epsilon_t, \quad t = 1, \dots, n, \quad (1)$$

где x_{tp} – значения регрессора x_p в наблюдении t , а $x_{t1} = 1$, $t = 1, \dots, n$. С учетом этого замечания не будем далее различать модели вида (1) со свободным членом или без свободного члена.

Основные гипотезы, лежащие в основе модели множественной регрессии, являются естественным обобщением модели парной регрессии

1. $y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \epsilon_t$, $t = 1, \dots, n$ – спецификация модели.
2. x_{t1}, \dots, x_{tk} – детерминированные величины. Векторы $x_s = (x_{1s}, \dots, x_{ns})'$, $s = 1, \dots, k$ линейно независимы в R^n .
- 3а. $E\epsilon_t = 0$, $E(\epsilon_t^2) = V(\epsilon_t) = \sigma^2$ – не зависит от t
- 3б. $E(\epsilon_t \epsilon_s) = 0$ при $t \neq s$, статистическая независимость (некоррелированность) ошибок для разных наблюдений.

Часто добавляется условие:

- 3с. Ошибки ϵ_t , $t = 1, \dots, n$, имеют совместное нормальное распределение: $\epsilon_t \sim N(0, \sigma^2)$.

Так же рассмотрен класс непараметрических оценок функции регрессии, использующих интерполяционные коэффициенты Фурье-Лагранжа.

Дана регрессионная модель

$$Y_i = m(X_i) + \epsilon_i, \quad i = 1, \dots, n,$$

где $m(x) = E(Y|X=x)$ – неизвестная функция регрессии, подлежащая оцениванию на основе эмпирических данных $\{(X_i, Y_i)\}_{i=1}^n$, а $\{\epsilon_i\}_{i=1}^n$ – случай-

ные ошибки. Один из непараметрических методов построения оценки $\hat{m}(x)$ основан на разложении функции $m(x)$ в ряд Фурье

$$m(x) = \sum_{j=0}^{\infty} \beta_j \varphi_j(x) \quad (2)$$

по некоторой ортонормированной системе $\{\varphi_j(x)\}_{j=0}^{\infty}$ с последующей заменой бесконечного ряда (2) его частичной суммой подходящего порядка $N(n)$, в которой коэффициенты Фурье β_j заменены их оценками $\hat{\beta}_j$. Например, в [1] рассмотрена следующая конструкция. Предположим, что система $\{\varphi_j(x)\}_{j=0}^{\infty}$ ортонормирована на отрезке $[-1; 1]$ относительно скалярного произведения $(f, g) = \int_{-1}^1 f(x)g(x)dx$, и что переменная X принимает равноотстоящие значения из $[-1; 1]$. Рассмотрим разбиение этого отрезка на непересекающиеся интервалы $\{A_i\}_{i=1}^n$ такие, что $X_i \in A_i, i = 1, \dots, n$. Тогда в качестве оценки функции регрессии берется выражение

$$\hat{m}_N(x) = \sum_{j=0}^{N(n)} \hat{\beta}_j \varphi_j(x),$$

где

$$\hat{\beta}_j = \sum_{i=1}^n Y_i \int_{A_i} \varphi_j(x) dx. \quad (3)$$

Там же приведены условия на $\{\varphi_j\}$, $N(n)$ и $\{\epsilon_i\}_{i=1}^n$ при которых эта оценка будет состоятельной.

Второй раздел включает в себя моделирование экономического роста в среде программирования R.

R — это язык программирования и среда для статистических вычислений и графического анализа, первоначально разработанным в Белл Лабораториз (Bell Labs). Это программа для анализа данных с открытым кодом, которая поддерживается большим и активным исследовательским сообществом по всему миру. Однако существует много распространенных программ для статистической и графической обработки данных (таких как Microsoft Excel, SAS, IBM SPSS, Stata и Minitab).

R возник как свободный аналог среды S-PLUS, которая, в свою очередь, является коммерческой реализацией языка расчётов S. Язык S – довольно старая разработка (почти как TeX). Он возник в 1976 году в компании Bell Labs, и был назван, естественно, «по мотивам» языка C. Начиная с третьей версии (1988 г.), коммерческая реализация S называется S-PLUS.

В августе 1993 г. двое молодых новозеландских ученых из университета Окленда анонсировали свою новую разработку, которую они назвали R. По замыслу создателей, Роберта Джентльмена (Robert Gentleman) и Росса Ихака (Ross Ihaka), она должна была стать новой реализацией языка S, отличающейся от S-PLUS некоторыми деталями, например, обращением с глобальными и локальными переменными, а также работой с памятью. Фактически, они создали не полный аналог S-PLUS, а новую «ветку» на «дереве S». Многие вещи, которые отличают R от S-PLUS, связаны с влиянием языка Scheme (функциональный язык программирования, один из наиболее популярных диалектов языка Lisp).

На середину 2016 года R догнал SAS и SPSS (которые являются платными) и вошел в тройку самых распространенных систем для обработки статистической информации. Также следует отметить, что R входит в 10 языков программирования общего назначения.

В R используется интерфейс командной строки, хотя доступны и несколько графических интерфейсов пользователя, например пакет R Commander, RKWard, RStudio, Weka, Rapid Miner, KNIME, а также средства интеграции в офисные пакеты.

Shiny — R пакет, с помощью которого можно построить любое интерактивное приложение, используя язык R.

В простейшем виде приложение, созданное при помощи Shiny, состоит из двух файлов, которые хранятся в одной папке. Один из этих обязательных файлов (должен иметь имя ui.R) содержит определения элементов пользовательского интерфейса; второй обязательный файл - server.R - содержит скрипт, определяющий работу сервера.

Особенности пакета:

- Создание удобного веб-приложения всего несколькими строками кода - JavaScript не требуется;

- Приложения, созданные с использованием Shiny автоматически “живые” также как и исходные таблицы. Выход меняется интерактивно по мере изменения входных данных не требуя перезагрузки браузера;
- Пользовательский интерфейс Shiny может быть создан полностью в R или написан прямо в HTML, CSS или JavaScript для большей гибкости;
 - *Работает в любой среде R (консольная версия R, Rgui для Windows или Mac, ESS, StatET, RStudio и другие);
- Привлекательный UI по умолчанию на основании темы Twitter Bootstrap; Настраиваемый виджет слайдера с встроенной поддержкой для анимации;
- Предварительно установленные выходные виджеты для отображения графиков и таблиц, а также вывода на печать объектов R;
- Быстрая двухсторонняя коммуникация между веб браузером и R с использованием пакета websockets;
- Использует реактивную модель программирования которая исключает беспорядочные события при работе с кодом, так что вы можете сосредоточиться на том, что действительно важно;
- Разрабатывайте и размещайте свои собственные Shiny виджеты чтобы другие разработчики могли легко их включить в свои приложения (ожидается).

Для разработки приложения использовались шаблоны, а так же было пройдено обучение для самостоятельного построения приложений. Шаблон для создания загрузки был предназначен для файлов формата .txt, однако для анализа используются чаще всего данные в формате .csv, поэтому был изменен код шаблона для корректной работы приложения. Представленный ниже виджет в соответствии с рисунком ?? показывает, что файлы в формате .csv считываются, а значит можно сделать анализ.

В соответствии с рисунком 2 представленным ниже можно увидеть окно приложения, где у пользователя есть возможность загрузить файл со своими данными не более 5 Мб. Особенность этого приложения состоит в том, что пользователь может изменить названия переменных в зависимости от названия столбцов в загруженных данных.

Также в соответствии с рисунком 3 и рисунком 4 отображены дополни-

App: 1 - Shay
 http://127.0.0.1:3434 Open in Browser Publish

Multiple Linear Regression

File must have a header with variables' names.

Choose CSV File


Browse... 123.csv
 Upload complete

Id column

Loaded variables:
 x1, x2, x3, x4, y

Regression formula

y ~ x1+x2+x3



Document format

PDF HTML Word

Download

Summary & ANOVA Residual Plots Multicollinearity

Call:
`lm(formula = as.formula(input$formula), data = dataLoaded())`

Residuals:

Min	1Q	Median	3Q	Max
-64.558	-32.031	6.262	27.049	85.526

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-193.07	35.78	-5.396	2.03e-05 ***
x1	107.80	22.92	4.702	0.000109 ***
x2	304.24	165.17	1.842	0.078998 .
x3	143.13	44.13	3.243	0.003731 **

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 38.05 on 22 degrees of freedom
 Multiple R-squared: 0.8602, Adjusted R-squared: 0.8412
 F-statistic: 45.14 on 3 and 22 DF, p-value: 1.435e-09

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
x1	1	152591	152591	105.375	7.481e-10 ***
x2	1	28274	28274	19.525	0.0002169 ***
x3	1	15232	15232	10.519	0.0037307 **
Residuals	22	31858	1448		

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Рисунок 2 – Summary & ANOVA

тельные возможности приложения.

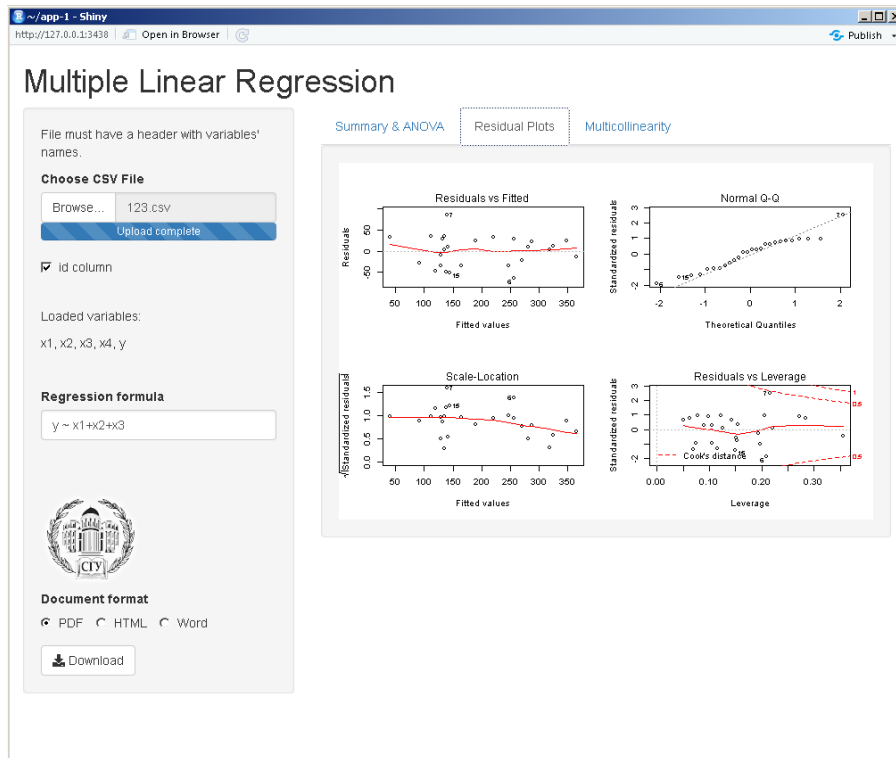


Рисунок 3 – Residual Plots

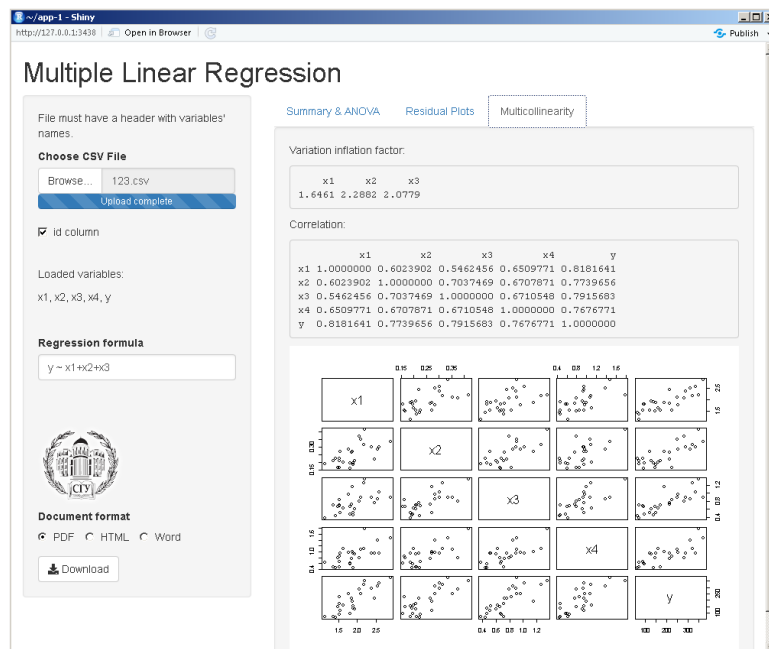


Рисунок 4 – Multicollinearity

Multiple Linear Regression

File must have a header with variables' names.

Choose CSV File


Browse... larch.csv
Upload complete

id column

Loaded variables:
x1, x2, x3, x4, y

Regression formula
y ~ x1+x2+x3

n
1 5 10


Document format
 PDF HTML Word
Download

Summary & ANOVA Residual Plots Multicollinearity

Nonparametric Regression

a:

```
[1] 13.0507166 -4.1105534 -1.1162377 -0.5712399 -0.3943146 -0.3330126
```

b:

```
[1] 0.000000e+00 4.996004e-16 2.220446e-15 1.776357e-15 2.109424e-15  
[6] -6.106227e-16
```

Coefficients:

```
[1] 26.1014331 -8.2211068 -2.2324755 -1.1424798 -0.7886293 -0.6660252  
[1] 0.000000e+00 1.387779e-15 4.440892e-15 3.552714e-15 4.107825e-15  
[6] -1.276756e-15  
[1] 10.99547
```

Plot

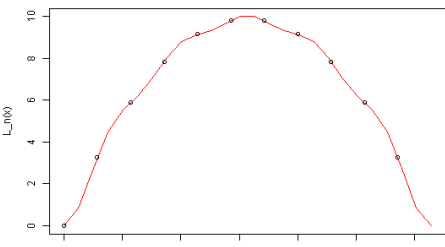


Рисунок 5 – Nonparametric Regression

ЗАКЛЮЧЕНИЕ

Стратегическая цель экономической политики России состоит в обеспечении устойчивого экономического роста, важнейшим фактором которого являются инвестиции в основной капитал. Существенное влияние на экономический рост в России оказывают также внешние факторы, важнейшими из которых являются мировые цены на углеводородное сырье и политически мотивированная санкционная политика отдельных стран. В этой связи значительную роль, наряду с другими, приобретают исследования на основе эконометрического моделирования количественных и качественных закономерностей влияния основных макроэкономических показателей на экономический рост. В рамках данного направления исследований и была выполнена практическая работа. Построенные регрессионные зависимости могут быть использованы при анализе влияния некоторых макроэкономических показателей на динамику валового регионального продукта.