МИНОБРНАУКИ РОССИИ

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

Кафедра дифференциальных уравнений и математической экономики

МНОГОФАКТОРНЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ ЭКОНОМИЧЕСКОГО РОСТА

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 451 группы
направления 38.03.05 — Бизнес-информатика

механико-математического факультета
Хайровой Ренаты Равилевны

Научный руководитель
доцент, к. ф.-м. н., доцент ______________ В. В. Новиков

Заведующий кафедрой
зав.кафедрой, д. ф.-м. н.,
профессор ______________ С. И. Дудов

Саратов 2021

ВВЕДЕНИЕ

В последние годы в РФ по ряду причин наблюдаются невысокие показатели экономического роста, хотя по своему типу ее экономика ближе к экономикам развивающихся стран, для которых характерны высокие темпы экономического развития. В этой связи, наряду с выявлением и анализом экономических, внешнеполитических и социальных факторов, негативно влияющих на экономический рост, приобретает актуальность также и задача анализа зависимостей между темпами роста и различными количественными показателями состояния экономики внутри страны и за ее пределами. Целью такого анализа может быть, в частности, выявление корреляций между темпами роста и теми или иными количественными факторами.

Актуальность данной тематики определяется тем, что на основе результатов корреляционно-регрессионного анализа можно спрогнозировать поведение того или иного социально-экономического процесса, выявить наиболее значимые факторы, влияющие на него.

Целью данной работы является исследование и моделирование некоторых аспектов развития экономики Российской Федерации.

В рамках достижения этой цели решались следующие задачи:

1. На основе учебной и монографической литературы выполнить обзор основных методов регрессионного анализа.

2. Дать краткий обзор существующих подходов к моделированию экономического роста и выполнить эконометрическое моделирование с использованием программы Gretl на основе реальных данных по Саратовской области.

3. Освоить основы программирования на языке R и разработать на нем программу, выполняющую процедуру непараметрической регрессии на основе оценок ортогонального разложения по многочленам Лежандра.

Работа состоит из введения, заключения, четырех разделов и списка источников. В первых двух разделах рассматриваются математические основы регрессионного анализа. Приведены основные факты относительно получения оценок коэффициентов для парной и множественной линейной регрессии по методу наименьших квадратов, статистические свойства этих оценок и критерии качества подгонки регрессионных моделей. Наряду с традиционными параметрическими методами рассмотрены и наиболее распространенные непараметрические подходы. В третьем разделе рассмотрены теоретические аспекты построения моделей экономического роста, а также их практическая реализации средствами статистической программы Gretl для данных по Саратовской области. Четвертый раздел посвящен разработке программы на языке R, выполняющей процедуру непараметрической регрессии на основе оценок ортогонального разложения по многочленам Лежандра.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Многофакторная (многомерная или множественная) регрессия

Данная позволяет построить и проверить связь между одной результирующей переменной и несколькими независимыми факторами, оказывающими на нее влияние.[1]

Многофакторная регрессионная модель (multiple regression model) — это уравнение связи с несколькими независимыми (объясняющими) переменными. В общем виде уравнение связи можно записать следующим образом:

$$y = f(\vec{x}) + \varepsilon,$$

где $\vec{x} = (x_1, x_2, ..., x_n)$ — независимые (объясняющие) переменные, f — неизвестная функция регрессии, подлежащая оценке по эмпирическим данным, ε — ошибка регрессии, случайная величина, дополнительный остаточный член, который отражает влияние случайных факторов. Эти факторы могут быть связаны с особенностями измерений и действий, оказывающих влияние на результирующую переменную, а также с влиянием других объясняющих переменных, которые не были включены в уравнение.

Остаток регрессии — это разность между фактическим и теоретическим значением (оценкой), вычисляется по формуле:

$$e = f(\vec{x}) - \hat{f}(\vec{x}) = y - \hat{y},$$

где $\hat{y} = f\hat{(x)}$ — оценка неизвестной функции регрессии f, полученная тем или иным способом по данным наблюдений. При так называемом параметрическом подходе вид функции $f(\vec{x}, \vec{a})$ считается известным, а оценке подлежат числовые параметры $\vec{a} = (a_1, \ldots, a_k)$, от которых эта функция зависит. Для определения значений $\vec{a}$ используют числовую информацию, которая рассматривается как выборочная. Поэтому рассчитанные на ее основе величины $\hat{\vec{a}} = (\hat{a}_1, \ldots, \hat{a}_k)$, называют оценками параметров, подчеркивая тем самым их возможную неточность из-за неполноты информации.

Оценки параметров могут меняться от выборки к выборке, поэтому они рассматриваются как случайные величины. Так как найденные параметры являются лишь выборочными оценками неизвестных параметров по генеральной совокупности, то возникает вопрос об их качестве. Характеристиками качества полученных оценок параметров регрессии являются их несмещенность, эффективность и состоятельность.[4]

Оценка параметра называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру. Например, математическое ожидание несмещенной оценки $\hat{a}_j$ равно его значению a_j в генеральной совокупности:

$$\mathrm{M}(\hat{a}_j) = a_j.$$

Поскольку оценки являются случайными переменными, их значения лишь по случайному совпадению могут в точности равняться характеристикам генеральной совокупности. Обычно будет присутствовать определенная ошибка, которая может быть большой и малой, положительной и отрицательной в зависимости от чисто случайных составляющих величин x в выборке. Несмещенность означает отсутствие систематической ошибки в большую или меньшую сторону при вычислении оценки. Если условие несмещенности не выполняется, то оценка называется смещенной, а разность $b(\hat{a}_j) = \mathsf{M}\hat{a}_j - a_j$ между ее математическим ожиданием и соответствующей характеристикой генеральной совокупности — *смещением*.

Оценка параметра называется *эффективной*, если она имеет наименьшую дисперсию среди всех возможных несмещенных оценок данного параметра по выборкам одного и того же объема. Желательно получить оценку с максимально возможной вероятностью, близкой к значению теоретической характеристики, что означает получить функцию плотности вероятности, как можно более «сжатую» вокруг истинного значения. Один из способов выразить это требование — сказать, что необходимо получить сколь возможно малую дисперсию.

Оценка параметра называется *состоятельной*, если с увеличением числа наблюдений она стремится (сходится по вероятности) к значению параметра в генеральной совокупности. С содержательной точки зрения состоятельность проявляется в том, что оценка в среднем будет тем точнее, чем большее число наблюдений будет взято для ее вычисления.

Ядерное сглаживание

Идейно простой подход к представлению последовательности весов $\{W_{ni}(x)\}_{x=1}^{n}$ состоит в описании формы весовой функции $W_{ni}(x)$ посредством функции плотности со скалярным параметром, который регулирует размер и форму весов около x. Эту функцию формы принято называть ядром K. Ядро — это непрерывная ограниченная симметричная вещественная функция K с единичным интегралом

$$\int K(u)du = 1.$$

Последовательность весов для ядерных оценок (для одномерного x) определяется как

$$W_{ni}(x) = K_{h_n}(x - X_i)/\widehat{f}_{h_n}(x), \tag{1}$$

где

$$\widehat{f}_{h_n}(x) = n^{-1}\sum_{i=1}^{n} K_{h_n}(x - X_i), \tag{2}$$

а

$$K_{h_n}(u) = h_n^{-1}K(u/h_n)$$

представляет собой ядро с параметром масштаба h_n. Подчеркнув зависимость $h = h_n$ от объема выборки n, условимся сокращенно обозначать последовательность весов (4) через $\{W_{hi}(x)\}_{x=1}^{n}$.

Функция $\widehat{f}_h(\cdot)$ является ядерной оценкой плотности Розенблата — Парзена ([9] и [10]) для (маргинальной) плотности переменной X. Вид (4) ядерных весов $W_{hi}(x)$ был предложен в работах [7] и [8], и, как следследствие,

$$\widehat{m}_h(x) = \frac{n^{-1}\sum\limits_{i=1}^{n} K_h(x - X_i)Y_i}{n^{-1}\sum\limits_{i=1}^{n} K_h(x - X_i)}$$

часто называют оценкой Надарая — Ватсона. форма ядерных весов определяется ядром K, в то время как размер весов параметризируется посредством переменной h, называемой шириной окна.

Построение регрессионных уравнений осуществлялось с помощью эконометрической программы Gretl. Всего построено и проанализировано шесть моделей.

Состав факторных показателей, влияющих на экономический рост

OilPrice - Цена на нефть марки Brent, $/баррель;

Investment in fix assets - Инвестиции в основной капитал, млн рублей;

VRP(SarObl) - Валовой региональный продукт, млн рублей;

L - Численность экономически активного населения;

VRP/L - Производительность общественного труда;

Invest/VRP - Доля инвестиций в ВРП;

Ird - Инвестиции в основной капитал на душу населения, рублей;

K - Стоимость основных фондов, млн. рублей;

K/L - Капиталовооруженность труда;

Inz - Внутренние затраты на научные исследования и разработки, млн рублей;

Inostr Invest - Иностранные инвестиции, тыс. долларов;

HCU - Число студентов высших учебных заведений на 1000 человек.

Многофакторные модели экономического роста

1. Модель зависимости ВРП от цены на нефть и инвестиции в основной капитал:

$$\ln(VRP_t) = 8,8701 - 0,00062 \cdot \ln(Oilprice_t) + 0,9101 \cdot \ln(Investments_t) + \varepsilon_t$$

$$\rightarrow VRP_t = e^{8,8701} \cdot (Oilprice_t)^{-0,00062} \cdot (Investments_t)^{0,9101} \cdot e^{\varepsilon_t}$$

$$\rightarrow VRP_t = 7115,99 \cdot (Oilprice_t)^{-0,00062} \cdot (Investments_t)^{0,9101} \cdot e^{\varepsilon_t}$$

2. Модель зависимости ВРП от цены на нефть:

$$\ln(VRP_t) = 11,852 + 0,258 \cdot \ln(Oilprice_t) + \varepsilon_t \rightarrow VRP_t = e^{11,852} \cdot (Oilprice_t)^{0,258} \cdot e^{\varepsilon_t}$$

$$\rightarrow VRP_t = 140364,79 \cdot (Oilprice_t)^{0,258} \cdot e^{\varepsilon_t}$$

3. Модель взаимосвязи ВРП от инвестиций в основной капитал:

$$\ln(VRP_t) = 8,8678 + 0,91004 \cdot \ln(Investments_t) + \varepsilon_t$$

$$\to VRP_t = e^{8,8678} \cdot (Investments_t)^{0,91004} \cdot e^{\varepsilon_t}$$

$$\to VRP_t = 7099,64 \cdot (Oilprice_t)^{0,91004} \cdot e^{\varepsilon_t}$$

4. Модель зависимости ВРП от производительности труда и инвестиций:

$$\ln(VRP_t) = 13,4641 + 0,9037 \cdot \ln(VRPL_t) - 0,0546 \cdot \ln(InvVRP_t) + \varepsilon_t$$

$$\to VRP_t = e^{13,4641} \cdot (VRPL_t)^{0,9037} \cdot (InvVRP_t)^{-0,0546} \cdot e^{\varepsilon_t}$$

$$\to VRP_t = 703694,78 \cdot (VRPL_t)^{0,9037} \cdot (InvVRP_t)^{-0,0546} \cdot e^{\varepsilon_t}$$

5. Модель взаимосвязи ВРП с объемом иностранных инвестиций и уровнем человеческого капитала:

$$\ln(VRP_t) = 13,0716 - 0,6022 \cdot \ln(HCU_t) + 0,315 \cdot \ln(Inostrinvest_t) + \varepsilon_t$$

$$\to VRP_t = e^{13,0716} \cdot (HCU_t)^{-0,6022} \cdot (Inostrinvest_t)^{0,315} \cdot e^{\varepsilon_t}$$

$$\to VRP_t = 475251,7 \cdot (HCU_t)^{-0,6022} \cdot (Inostrinvest_t)^{0,315} \cdot e^{\varepsilon_t}$$

6. Модель взаимосвязи ВРП с капиталовооруженностью труда, инвестициями на душу населения, внутренними затратами на научные исследования и разработки:

$$\ln(VRP_t) = 4,889 + 0,3577 \cdot \ln(K_t/L_t) + 0,3924 \cdot \ln(Ird_t) + 0,2897 \cdot \ln(Inz_t) + \varepsilon_t$$

$$\to VRP_t = e^{4,889} \cdot (K_t/L_t)^{0,3577} \cdot (Ird_t)^{0,3924} \cdot (Inz_t)^{0,2897} \cdot e^{\varepsilon_t}$$

$$\to VRP_t = 132,82 \cdot (K_t/L_t)^{0,3577} \cdot (Ird_t)^{0,3924} \cdot (Inz_t)^{0,2897} \cdot e^{\varepsilon_t}$$

Статистический анализ полученных уравнений показал, что они значимы: расчетные значения F-критерия больше табличного на 5%-ном уровне значимости. Проверка по t-критерию коэффициентов регрессии показала, что включенные в модель факторы оказывают некоторое влияние на зависимую переменную. Все знаки у коэффициентов регрессии соответствуют экономической сущности влияния аргументов на функцию. Полученные модели характеризуются довольно высокой степенью детерминации, отсутствием гетероскедастичности и автокорреляции в остатках.

Программирование приближающей функции

Последний раздел посвящен реализации программы. Для этого был использован следующий вариант формулы (**??**):

$$\widehat{P}_n(x) = \frac{(2n-1) \cdot x \cdot \sqrt{\frac{2}{2n-1}} \cdot \widehat{P}_{n-1}(x) - (n-1) \cdot \sqrt{\frac{2}{2n-3}} \cdot \widehat{P}_{n-2}(x)}{n \cdot \sqrt{\frac{2}{2n+1}}}, \qquad (3)$$

где $\widehat{P}_0(x) = 1/\sqrt{2}$, $\widehat{P}_1(x) = x/\sqrt{2/3}$.

Программирование формулы для приближающей функции при степени N, где N задается вручную. Сначала необходимо определить функцию K_N, которая задает ядро интегрального оператора частичной суммы Фурье-Лежандра.

```
N<-4
K_N <- function(u,v){
  K <- 0
  for(j in 1:N)
    K <- K + leg(j-1, u) * leg(j-1, v)
  print(K)
}
```

Код построения самой приближающей функции представлен ниже:

```
m1 <- function(u){
  M <- 0
  for(i in 1:n+1)
    M <- M+K_N(u,X[i])
```

```
  print(M)
}
m <- function(w){
  M1 <- 0
  for(i in 1:n+1)
    M1 <- M1+K_N(w,X[i])*Y[i]/m1(w)
  print(M1)
}
```

В результате при $N = 4$ для показателя «Производительность труда» строится следующий график:

Рисунок 1 – Степень приближающей функции = 4

Для сравнения возьмем тот же показатель, но со степенями $N = 5$ и $N = 8$:

Рисунок 2 – Степень приближающей функции = 5

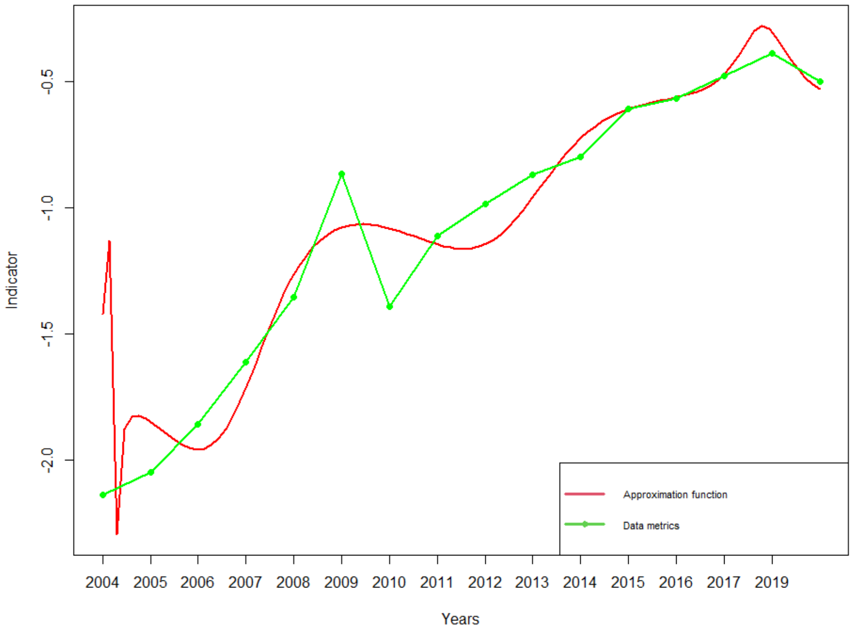

Рисунок 3 – Степень приближающей функции = 8

В рамках работы были построены графики приближающей функции степени $N = 4$ для всех показателей, использованных в моделях взаимосвязи ВРП.

ЗАКЛЮЧЕНИЕ

В ходе выполнения работы был выполнен обзор основных теоретических положений, составляющих математическую основу регрессионного анализа. Был выполнен поиск и проведена подготовка для дальнейшего анализа статистических данных по Саратовской области. На основе этих данных выполнено эконометрическое моделирование зависимостей одного из показателей, характеризующих экономический рост, от различных наборов макроэкономических факторов.

Также были освоены основы программирования на языке R и разработана на нем программа, выполняющую процедуру непараметрической регрессии на основе оценок ортогонального разложения по многочленам Лежандра. В процессе работы также было необходимо реализовать программу, позволяющую вычислять ортонормированный полином Лежандра любой степени.

Анализ построенных регрессионных моделей показал, что при учете совокупного воздействия регрессора на зависимую переменную, наибольшее влияние на экономический рост в Саратовской области за период 2007-2018 гг. оказывли следующие факторы:

— инвестиции в основной капитал;

— инвестиции в основной капитал на душу населения;

— уровень человеческого капитала;

— уровень производительности труда.

Во многих случаях рассмотрение эконометрических моделей экономического роста дает возможность определить, на проведение каких мероприятий следует ориентировать экономическую политику региона, чтобы ускорить его экономическое развитие. В то же время, следует отметить, что глобальные перемены в социальной, экономической и политической ситуации в мире в начале 2021 года позволяют рассматривать полученные в работе зависимости и выводы лишь в ретроспективном плане.

Результаты данной работы имеют практическую ценность в области моделирования показателей экономического роста.