**Домашнее задание по эконометрике №3**

**«Выявление драйверов экономического роста российских регионов: исследование чувствительности оценок ключевых переменных модели к методу оценивания»**

**Работу выполнили:**

Громова Дарья

Салимов Райымбек,

Твердов Илья

**Группа**:

МЭЭП212

**Постановка задачи, обзор литературы, теоретическая модель.**

Одним из показателей экономического роста регионов является валовой региональный продукт (ВРП). Это обобщающий показатель экономической деятельности региона, который характеризует процесс производства товаров и услуг для конечного производства.

*Актуальность работы*

Актуальность работы обусловлена возрастающим интересом к показателям регионального развития (прежде всего, к ВРП) со стороны органов государственной власти. ВРП вошел в систему показателей прогнозирования регионального развития, используется Министерством финансов РФ для распределения фондов финансовой поддержки территорий, включен в систему показателей эффективности деятельности органов государственной власти РФ (*Ивченко Ю.С., 2019*).

*Задача исследования:* с помощью эконометрического анализа определить, какое влияние на уровень ВРП оказывают выбранные переменные.

*Обзор литературы*

Существует множество исследований, посвященных анализу влияния различных факторов на уровень ВРП. В работе *Ивченко Ю.С. (2019)* проведен анализ влияния таких факторов, как среднегодовая остаточная стоимость основных фондов; среднегодовые остатки средств на рублевых счетах юридических лиц; доходы населения региона и др. на уровень ВРП субъектов России. Анализ проводился отдельно по нескольким группам регионов, выделенных в зависимости от их уровня ВРП, для получения более однородных статистических совокупностей.

Во многих исследованиях изучается влияние такого фактора, как инвестиции. В работе *Т.П. Скуфьиной, С.В. Баранова, Корчак Е.А. (2018)* исследуется влияние инвестиций на ВРП 8 регионов Севера и Арктики.

Основываясь на изученной литературе, была составлена следующая теоретическая модель:

В качестве объясняемой переменной рассматривается показатель ВРП на душу населения (руб.) – переменная у.

В качестве объясняющих переменных рассматриваются:

1. *Наличие основных фондов на конец года в среднегодовых ценах (млн рублей), нормированный на 1000 жителей – переменная x1.* Увеличение основных фондов может влиять на рост уровня производства. Таким образом, основные фонды влияют на экономический рост, а, следовательно, и на ВРП.
2. *Среднедушевые денежные доходы населения (руб./мес.) – переменная x2.* Среднедушевые денежные доходы населения можно рассматривать как показатель уровня жизни людей. Существует немало исследований, доказывающих взаимосвязь между ростом экономики и ростом уровня жизни людей.
3. *Уровень занятости населения в возрасте 15–72 лет (%) – переменная x3.* Уровень занятости населения является показателем трудовых ресурсов. Увеличение трудовых ресурсов положительно влияет на рост ВРП.
4. *Объем инновационных товаров, работ, услуг (млн. руб.), нормированный на 1000 жителей – переменная x4*. Согласно различным исследованиям, инновационный путь развития признан наиболее эффективным. Таким образом, инновации могут оказывать положительное влияние на экономический рост.
5. *Инвестиции в основной капитал (в млн. руб.), нормированные на 1000 жителей – переменная x5*. Рациональная инвестиционная политика может способствовать поддержанию высокого уровня производства за счет увеличения объемов и обновления основных фондов.

Для проведения исследования были взяты данные из Федеральной службы государственной статистики РФ (ФСГС). Рассматриваемый промежуток времени: 2017*–*2018 года.

Из общей выборки были исключены наблюдения по г. Москва, так как эти данные могут искажать общие результаты, а также регионы, наблюдения по которым содержали пропуски по исследуемым показателям. Общая выборка составила 162 наблюдения.

*Н0: выбранные переменные оказывают значимое влияние на уровень ВРП.*

**Визуальный анализ данных и анализ описательных статистик**

Ниже в таблице 1 представлены описательные статистики данных.

Таблица 1. Описательные статистики

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Obs | Mean | Std. dev. | Min | Max |
| y | 162 | 472890.80 | 353492.50 | 125667.90 | 2517125.00 |
| х1 | 162 | 1960.26 | 1586.15 | 490.99 | 9843.09 |
| x2 | 162 | 28173.09 | 9548.49 | 15011.00 | 78812.00 |
| x3 | 162 | 63.73 | 4.38 | 50.30 | 78.30 |
| x4 | 162 | 368.99 | 349.62 | 8.73 | 2189.57 |
| x5 | 162 | 106.40 | 96.92 | 26.89 | 623.23 |

Таким образом, выборка состоит из 162 наблюдений.

Переменные y, x1, х2, x4 имеют большой средний разброс вокруг средних значений. Переменная х3 имеет маленький средний разброс вокруг среднего значения. По приведенным статистикам сложно оценить поведение рассматриваемых показателей. Поэтому далее рассмотрен графический анализ исследуемых показателей.

Ниже представлены диаграммы рассеяния переменных х1, х2, х3, х4, х5.



Рисунок 1. Диаграмма рассеяния переменной х1



Рисунок 2. Диаграмма рассеяния переменной х2



Рисунок 3. Диаграмма рассеяния переменной х3



Рисунок 4. Диаграмма рассеяния переменной х4



Рисунок 5. Диаграмма рассеяния переменной х5

По представленным диаграммам можно заметить, что между у и переменными х1, х2, х4, х5 наблюдается линейная связь. Однако большая часть наблюдений находятся в одном месте, то есть данные распределены не равномерно. Между у и переменной х3, судя по диаграмме рассеяния, наблюдается нелинейная взаимосвязь.

**Приведите оценки и содержательную интерпретацию модели, признанной наиболее адекватной в ходе выполнения ДЗ1.**

Для определения наилучшей модели были проведены следующие тесты:

**Тест Бокса-Кокса для поиска наилучшей функциональной формы:**

**1.1 Переменные левой и правой частей преобразуются неодинаково**

Результаты теста представлены ниже в таблице 2.

Таблица 2. Результаты теста Бокса-Кокса 1

|  |  |  |  |
| --- | --- | --- | --- |
| Test  H0: | Restricted  log likelihood | chi2 | Prob > chi2 |
| theta=lambda = -1 | -2049.08 | 123.88 | 0.000 |
| theta=lambda = 0 | -2003.14 | 32.00 | 0.000 |
| theta=lambda = 1 | -2049.04 | 123.81 | 0.000 |

Ни одна гипотеза не принимается.

**1.2 Переменные левой и правой частей преобразуются одинаково**

Результаты теста представлены ниже в таблице 3.

Таблица 3. Результаты теста Бокса-Кокса 2

|  |  |  |  |
| --- | --- | --- | --- |
| Test  H0: | Restricted  log likelihood | chi2 | Prob > chi2 |
| lambda = -1 | -2049.08 | 94.70 | 0.000 |
| lambda = 0 | -2003.14 | 2.82 | 0.093 |
| lambda = 1 | -2049.04 | 94.63 | 0.000 |

Гипотеза о равенстве lambda = 0 не отвергается на 5 %-ом уровне значимости. Таким образом можно выбрать линейную в логарифмах модель.

**РЕ-теста МакКиннона, Уайта и Дэвидсона**

На основе данного теста не получилось сделать какие-то выводы о выборе между линейной и линейной в логарифмах модели, так как в результате проведения теста оба коэффициента при дополнительных разностях оказались значимы.

Таким образом, в качестве наилучшей функциональной формы для модели была выбрана линейная в логарифмах модель. В таблице 4 представлены результаты оценивания данной модели.

Таблица 4. Результаты оценивания линейной в логарифмах модели

|  |  |
| --- | --- |
| Переменные | Значения |
| ln\_x1 | 0.295 \*\*\*  (0.036) |
| ln\_x2 | 0.325 \*\*\*  (0.068) |
| ln\_x3 | 0.491 \*\*  (0.237) |
| ln\_x4 | 0.188 \*\*\*  (0.017) |
| ln\_x5 | 0.162 \*\*\*  (0.033) |
| \_cons | 3.623 \*\*\*  (0.881) |
| Количество наблюдений | 162 |
| Prob > x2 | 0.0000 |

\*\*\* p<0,01, \*\* p<0,05, \* p<0,1

В скобках приведены стандартные ошибки.

Можно сделать вывод, что модель в целом значима, имеет высокий R-квадрат (0.93). Все коэффициенты также значимы.

**Тест Чоу** для определения того, существует ли необходимость рассматривать ВРП по отдельным группам.

Предпосылками данного предположения являются исследования других авторов по схожей тематике, а также представленные ниже гистограммы по показателю ВРП рассматриваемых регионов за 2017 и 2018 года.

Рисунок 6. Гистограмма ВРП за 2017 год

Рисунок 7. Гистограмма ВРП за 2018 год

Таким образом, опираясь на рассмотренную выше литературу и на графический анализ данных, можно сформулировать следующую гипотезу о наличии неоднородности в данных: *Данные по ВРП регионов неоднородны*.

Рассмотрим 2 группы регионов:

1) Регионы, ВРП которых <500000 (Присвоим таким регионам дамми-переменную 0)

2) Регионы, ВРП которых >= 500000 (Присвоим таким регионам дамми-переменную 1)

Согласно проведенному тесту Чоу, гипотеза о наличии неоднородности данных подтвердилась, поэтому лучше рассматривать зависимости отдельно по 2 указанным выше категориям.

**Тест Рамсея**

Для понимания того, имеются ли пропущенные переменные в модели или нет, был проведен тест Рамсея. Результаты данного теста: F(3, 153) = 8.50; Prob>F = 0.000.

Нулевая гипотеза отвергается, что говорит о неверной спецификации модели.

Была создана новая переменная х4\_2 = (lnx4)^2 и включена в модель. Результаты повторного теста Рамсея с включенной новой переменной: F(3, 152) = 0.66; Prob>F = 0.581.

Теперь гипотеза о правильной спецификации модели не отвергается.

**Анализ проблемы мультиколлинеарности**

Ниже в таблице 5 представлена корреляционная матрица показателей.

Таблица 5. Корреляционная матрица показателей

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ln\_y | ln\_x1 | ln\_x2 | ln\_x3 | ln\_x4 | ln\_x5 |
| ln\_y | 1.00 |  |  |  |  |  |
| ln\_x1 | 0.87 | 1.00 |  |  |  |  |
| ln\_x2 | 0.81 | 0.72 | 1.00 |  |  |  |
| ln\_x3 | 0.69 | 0.57 | 0.65 | 1.00 |  |  |
| ln\_x4 | 0.84 | 0.65 | 0.59 | 0.64 | 1.00 |  |
| ln\_x5 | 0.84 | 0.79 | 0.74 | 0.55 | 0.63 | 1.00 |

По данным матрицы видно, что между некоторыми регрессорами существует сильная взаимосвязь. Для того, чтобы определить, существует ли в построенной регрессии проблема мультиколлинеарности, были рассчитаны VIF. Результаты представлены ниже в таблице 6.

Таблица 6. VIF показателей

|  |  |  |
| --- | --- | --- |
| Показатель | VIF | 1/VIF |
| ln\_x5 | 3.34 | 0.29 |
| ln\_x1 | 3.24 | 0.31 |
| ln\_x2 | 2.90 | 0.35 |
| ln\_x4 | 2.20 | 0.46 |
| ln\_x3 | 2.10 | 0.48 |
| Mean VIF | 2.76 |  |

Поскольку полученные значения VIF оказались небольшими, можно сделать вывод, что в данной регрессии проблемы мультиколлинеарности нет.

**Тесты Уайта и Бройша Пагана**

Для тестирования гетероскедастичности были проведены тесты Уайта и Бройша Пагана. Оба теста подтвердили наличие гетероскедастичности.

Для решения данной проблемы используем оценки Уайта для дисперсий коэффициентов.

Результаты представлены ниже в таблице 7.

Таблица 7. Оценки Уайта для решения проблемы гетероскедастичности

|  |  |
| --- | --- |
| Переменные | Значения |
| ln\_x1 | 0.271 \*\*\*  (0.038) |
| ln\_x2 | 0.327 \*\*\*  (0.077) |
| ln\_x3 | 0.441 \*  (0.231) |
| ln\_x4 | -0.233\*\*  (0.096) |
| x4\_2 | 0.045 \*\*\*  (0.009) |
| ln\_x5 | 0.097 \*\*\*  (0.036) |
| \_cons | 5.172 \*\*\*  (0.932) |
| Количество наблюдений | 162 |
| Prob > x2 | 0.0000 |

\*\*\* p<0,01, \*\* p<0,05, \* p<0,1

В скобках приведены стандартные ошибки.

Таким образом, с учетом проведенных тестов, данную модель можно считать наиболее адекватной. Можно сделать вывод, что модель в целом значима, имеет высокий R-квадрат (0.94). Все коэффициенты также значимы. Коэффициенты можно проинтерпретировать следующим образом:

* При увеличении показателя наличия основных фондов на 1%, ВРП увеличивается на 0.27%
* При увеличении среднедушевых денежных доходов населения на 1%, ВРП увеличивается на 0.33%
* При увеличении уровня занятости населения на 1%, ВРП увеличивается на 0.44%
* Объем инновационных товаров, работ, услуг имеет нелинейное влияние на ВРП. Так, при увеличении объема инновационных товаров, работ, услуг на 1%, ВРП уменьшается на 0.23%, затем с определенного момента ВРП увеличивается на 0.04%.
* При увеличении инвестиций в основной капитал на 1%, ВРП увеличивается на 0.09%

**Модели с цензурированными и усеченными зависимыми переменными**

Рассмотрим, как оцениваются цензурированные модели и модели с усеченными переменными. Строим регрессию, которая объясняет связь между *Валовым региональным продуктом* *на душу населения* и рядом других показателей: *Наличие основных фондов на конец года в среднегодовых ценах (млн рублей), нормированный на 1000 жителей; Среднедушевые денежные доходы населения (руб./мес.); Уровень занятости населения в возрасте 15*–*72 лет (%); Объем инновационных товаров, работ, услуг (млн. руб.), нормированный на 1000 жителей; Инвестиции в основной капитал (в млн. руб.), нормированные на 1000 жителей*.

Рассмотрим график гистограммы для Валового регионального продукта (рис. 8). Можно заметить, что для большинства наблюдений ВРП на душу населения составляет до 1 000 000 рублей. Видно, что есть регионы с ВРП выше этого значения, и их немного. Характер зависимой переменной для этой группы регионов может отличаться, поэтому стоит рассмотреть модели с их выделением.



Рисунок 8. График гистограммы для переменной Валового регионального продукта (ВРП) на душу населения, руб.

Рассмотрим модели:

* Ordinary Least Squares.
* Усеченную справа регрессию (ВРП на душу населения не менее 1 000 000 рублей).
* Усеченную слева регрессию (ВРП на душу населения более 1 000 000 рублей).
* Цензурированную справа регрессию (в предположении, что доступны наблюдения с ВРП на душу населения менее 1 000 000 рублей).
* Цензурированную слева регрессию (в предположении, что доступны наблюдения с ВРП на душу населения более 1 000 000 рублей).

Сводная таблица (таблица 8) показывает, оценки моделей: Ordinary Least Squares (ols), Усеченная справа регрессия (tr\_ul), Усеченная слева регрессия (tr\_ll), Цензурированная справа регрессия (tob\_ul), Цензурированная слева регрессия (tob\_ll).

Таблица 8. Сводная таблица моделей.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) |
|  | ols | tr\_ul | tr\_ll | tob\_ul | tob\_ll |
| Переменные | ln\_y | ln\_y | ln\_y | ln\_y | ln\_y |
|  |  |  |  |  |  |
| ln\_x1 | 0.295\*\*\* | 0.272\*\*\* | -0.415\*\* | 0.263\*\*\* | -0.408\*\* |
|  | (0.0361) | (0.0358) | (0.185) | (0.0351) | (0.180) |
| ln\_x2 | 0.325\*\*\* | 0.311\*\*\* | 0.451\* | 0.327\*\*\* | 0.407\* |
|  | (0.0680) | (0.0686) | (0.260) | (0.0677) | (0.227) |
| ln\_x3 | 0.491\*\* | 0.630\*\*\* | -1.512 | 0.620\*\*\* | -1.343 |
|  | (0.237) | (0.228) | (0.967) | (0.226) | (0.860) |
| ln\_x4 | 0.188\*\*\* | 0.185\*\*\* | 1.215\*\*\* | 0.185\*\*\* | 1.200\*\*\* |
|  | (0.0166) | (0.0157) | (0.228) | (0.0157) | (0.220) |
| ln\_x5 | 0.162\*\*\* | 0.122\*\*\* | 0.212\* | 0.135\*\*\* | 0.211\* |
|  | (0.0332) | (0.0329) | (0.127) | (0.0322) | (0.119) |
| /sigma |  | 0.135\*\*\* | 0.0508\*\*\* |  |  |
|  |  | (0.00776) | (0.0124) |  |  |
| /var(e.ln\_y) |  |  |  | 0.0182\*\*\* | 0.00247\*\* |
|  |  |  |  | (0.00208) | (0.00116) |
| Constant | 3.623\*\*\* | 3.543\*\*\* | 9.370\*\*\* | 3.422\*\*\* | 9.171\*\*\* |
|  | (0.881) | (0.904) | (2.651) | (0.888) | (2.492) |
|  |  |  |  |  |  |
| Observations | 162 | 153 | 9 | 162 | 162 |
| R-squared | 0.927 |  |  |  |  |

Таблица оценок показывает, что для моделей с усеченными и цензурированными слева зависимыми переменными (tr\_ll и tob\_ll) отличается значимость таких переменных, как Среднедушевые денежные доходы населения (x2), Уровень занятости населения в возрасте 15–72 лет (x3) и Инвестиции в основной капитал (в млн. руб.), нормированные на 1000 жителей (x5). Наличие основных фондов (x1) меняет свой знак для этих регрессий.

Таким образом для регионов с ВРП на душу населения выше 1 000 000 рублей, а это Сахалинская, Тюменская, Магаданская область, Чукотский автономный округ, Республика Саха (Якутия), регрессия различается. Эти регионы отличаются низкой плотностью населения. Для этих регионов факторы значимые для других регионов не подходят, и необходимо строить иные модели, с учетом специфики данных регионов.

**Модели пространственной эконометрики**

В качестве одного из дополнительных методов были выбраны методы пространственной эконометрики. Мы предполагаем, что рассматриваемые переменные коррелируют между соседними регионами, т.е. возможно наличие пространственной зависимости между регионами. Выясним, есть ли пространственная корреляция между рассматриваемыми регионами и если есть, то есть ли различия в пределах изучаемого периода 2017–2018 гг. Проведем также оценки коэффициентов пространственной модели.

Для начала вычислим матрицу пространственных весов для того, чтобы задать направление упорядоченности показателей. В целом существует три группы методов определения пространственного соседства: близость по смежности, по графу и по метрике. В нашем случае был выбран способ поиска соседей по метрике как наиболее простой и надежный способ определения соседства. Поиск соседей по метрике также может осуществляться двумя методами – поиск по количеству и по расстоянию. Нами был выбран наиболее популярный метод – поиск по расстоянию, в первую очередь, из-за наличия таких регионов как Сахалинская и Калининградская области, которые не граничат непосредственно с другими регионами России. На рисунке 9 изображен граф соседства регионов по расстоянию. Москва также была исключена, как и во всей работе.



Рисунок 9. Граф соседства регионов России по расстоянию.

Далее были вычислены веса нормированной матрицы. Весом нормированы на количество регионов-соседей. Чем больше у региона соседей, тем меньшим будет величина веса в соответствующих элементах матрицы. Нормирование по строкам необходимо для соблюдения концепции пространственного «лага». На рисунке 10 представлена визуализация матрицы весов. Полученная матрица весов дает искомую меру потенциальной пространственной связи между всеми парами территориальных единиц. Если сопоставить эту меру со значениями показателей в тех же территориальных единицах, то можно получить оценку пространственной автокорреляции соответствующего показателя.



Рисунок 10. Матрица пространственных весов.

Далее была проведена диагностика пространственной автокорреляции. Для вычисления меры пространственной зависимости используется целый ряд индексов, наиболее используемый из них – индекс Морана.

При проверке пространственной корреляции при расчёте индекса Морана нулевой гипотезой является H0: пространственная корреляция равна нулю, альтернативной же выступает: H1: пространственная корреляция не равна нулю. Рассчитывается Z-статистика, имеющая стандартное нормальное распределение. При отвержении нулевой гипотезы положительное значение Z-статистики говорит о положительной пространственной корреляции – наблюдается группировка схожих (высоких и низких) значений, отрицательное – соответственно, об отрицательной – что означает идеальное перемешивание низких и высоких значений (графически это выглядит как шахматная доска). Значение, равное 0, означает абсолютно случайное распределение, т.е. пространственной зависимости нет. Значение индекса Морана лежит в диапазоне от -1 до +1 для нормально распределенных данных.

Вычислим значение индекса Морана по всем исследуемым переменным за оба года. В таблице 9 указано значение индекса Морана, в скобках – Z-оценка индекса. Все полученные значения индекса являются значимыми на любом разумном уровне значимости.

Таблица 9. Значения индекса Морана для всех переменных

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Переменная | | | | | |
| Год | y | x1 | x2 | x3 | x4 | x5 |
| 2017 | 0.3184  (8.371) | 0.2343  (6.2684) | 0.4691  (11.939) | 0.3586  (8.8535) | 0.1437  (3.8976) | 0.2189  (6.0393) |
| 2018 | 0.2409  (6.5747) | 0.2515  (6.6584) | 0.5386  (13.732) | 0.2990  (7.4001) | 0.1216  (3.3778) | 0.2857  (7.5022) |

Мы видим, что по всем рассматриваемым переменных наблюдается положительная пространственная автокорреляция. В наибольшей степени она сильна по переменной x2 (cреднедушевые денежные доходы населения), что о говорит о существовании выраженных кластеров в пространстве, вероятна сильная кооперация между соседними регионами. Наименьшее значение наблюдается по показателю x4 (объем инновационных товаров, работ, услуг на душу населения) – кластеризация выражена в наименьшей степени, распределение регионов более рассеянное, возможна повышенная конкуренция между ними.

Рассматривая в динамике между годами, можно сказать, что наибольшее относительное увеличение степени кластеризации за год произошло по показателю x5 (инвестиции в основной капитал на душу населения) – на 30%, а наибольшее увеличение рассеяния – по переменной y (ВРП на душу населения), на 24%.

Визуализировать пространственную автокорреляцию возможно с помощью диаграмм рассеяния Морана. По оси X отложено значение показателя в каждом регионе, по оси Y – пространственный лаг региона (средневзвешенное значение по всем его соседям). Наклонная линия – это линейная регрессия, тангенс её наклона равен значению индекса Морана. Для примера изображена диаграмма для зависимой переменной y (рис. 11), а также для независимых переменных с наибольшей положительной автокорреляцией (рис. 12) и с наименьшей (рис. 13).



Рисунок 11. Диаграмма рассеяния Морана по переменной y за 2018 г.

(Индекс Морана = 0.3184)



Рисунок 12. Диаграмма рассеяния Морана по переменной x2 за 2018 г.

(Индекс Морана = 0.5386)



Рисунок 13. Диаграмма рассеяния Морана по переменной x4 за 2018 г.

(Индекс Морана = 0.1216)

Для моделирования пространственной автокорреляции была выбрана модель SAR – модель с пространственным авторегрессионным лагом, поскольку распределение переменной y не является случайным. В нашей работе было использовано следующее уравнение модели SAR:

,

Получим оценки коэффициентов регрессоров () и коэффициент авторегрессии () уравнений за оба года (табл. 10, 11). Оценки коэффициентов регрессоров сравним также с обычными МНК-оценками.

Таблица 10. Оценки параметров SAR-модели за 2017 год

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Оценки  (в скобках указаны стандартные ошибки) | z-value | p-value | LR-статистика о значимости |
| Intercept | -168088.0941  (102945.6487) | -1.6328 | 0.1025143 |  |
| x1 | 43.1649  (8.0408) | 5.3682 | 0.0000\*\*\* |
| x2 | 3.6084  (1.0799) | 3.3415 | 0.0008\*\*\* |
| x3 | 2228.5272  (1804.0080) | 1.2353 | 0.2167114 |
| x4 | 410.2067  (42.2858) | 9.7008 | 0.0000\*\*\* |
| x5 | 788.8089  (131.3299) | 6.0063 | 0.0000\*\*\* |
|  | 0.15932 |  | 0.0001\*\*\* | 16.003 |

\*\*\* значим на любом уровне значимости

Таблица 11. Оценки параметров SAR-модели за 2018 год

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Оценки  (в скобках указаны стандартные ошибки) | z-value | p-value | LR-статистика о значимости |
| Intercept | 57882.8371  (62277.8595) | 0.3567 | 0.7213 |  |
| x1 | 62.3166  (11.4941) | 5.4216 | 0.0000\*\*\* |
| x2 | 4.9534  (1.8887) | 2.6226 | 0.0087\*\*\* |
| x3 | -2179.1726  (2922.5933) | -0.7456 | 0.4559 |
| x4 | 538.0793  (50.7161) | 10.6096 | 0.0000\*\*\* |
| x5 | 604.2404  (190.5153) | 3.1716 | 0.0015\*\*\* |
|  | 0.0629 |  | 0.2155 | 1.5339 |

\*\*\* значим на любом уровне значимости

По таблицам 2 и 3 видно, что на данных оба года являются значимыми оценки коэффициентов переменных x1, x2, x4 и x5. Отличием является то, что на данных за 2018 год коэффициент авторегрессии не значимым, это говорит о том, что модель SAR не является оптимальной, лучше использовать другую модель, например, SEM (модель с пространственным взаимодействием в ошибках). Также из моделей за 2 года лучшей по качеству является за 2017 год (AIC = 2004.3 против 2077.3).

Сравним оценки коэффициентов регрессоров SAR-модели с обычными МНК-оценками (табл. 12 и 13). Во-первых, оценки SAR и МНК не совпадают, что ещё раз подтверждает наличие пространственной автокорреляции, во-вторых, оценки SAR в основном меньше оценок МНК вследствие пространственного лага. Таким образом, из-за неучёта пространственной автокорреляции в линейной модели МНК-оценки являются несостоятельными.

Таблица 12. Оценки параметров линейной модели за 2017 год

|  |  |  |  |
| --- | --- | --- | --- |
|  | Оценки  (в скобках указаны стандартные ошибки) | t-value | p-value |
| Intercept | -227155.000  (117833.321) | -1.928 | 0.0577\* |
| x1 | 43.905  (9.234) | 4.755 | 0.0000\*\*\* |
| x2 | 5.088  (1.174) | 4.334 | 0.0000\*\*\* |
| x3 | 3611.035  (2049.970) | 1.762 | 0.0822\* |
| x4 | 370.536  (47.394) | 7.818 | 0.0000\*\*\* |
| x5 | 903.418  (147.580) | 6.122 | 0.0000\*\*\* |
|  | 0.9636 | | |

\* значим на 10% уровне значимости, \*\*\* значим на любом уровне значимости

Таблица 13. Оценки параметров линейной модели за 2018 год

|  |  |  |  |
| --- | --- | --- | --- |
|  | Оценки  (в скобках указаны стандартные ошибки) | t-value | p-value |
| Intercept | 65693.790  (189083.898) | 0.347 | 0.72924 |
| x1 | 62.050  (12.113) | 5.123 | 0.0000\*\*\* |
| x2 | 5.919  (1.856) | 3.188 | 0.0021\*\* |
| x3 | -2220.616  (3399.985) | -0.653 | 0.51567 |
| x4 | 522.754  (52.175) | 10.019 | 0.0000\*\*\* |
| x5 | 649.678  (196.640) | 3.304 | 0.0015\*\* |
|  | 0.9571 | | |

\* значим на 1% уровне значимости, \*\*\* значим на любом уровне значимости

**Заключение**

Исходя из моделей с усеченными и цензурированными переменными для регионов с ВРП на душу населения выше 1 000 000 рублей, а это Сахалинская, Тюменская, Магаданская область, Чукотский автономный округ, Республика Саха (Якутия), регрессия различается. Эти регионы отличаются низкой плотностью населения. Для этих регионов факторы, значимые для других регионов, не подходят, и необходимо строить иные модели, с учетом специфики данных регионов.

Обычная OLS модель не выделяет выборку с определенными значениями зависимой переменной, а анализ усеченных и цензурированных моделей показал, что такое рассмотрение может быть релевантным, и указывать на иные результаты. С учетом этого рекомендуемая экономическая политика должна различаться для различных кластеров. Для регионов с высоким ВРП на душу на населения не эффектно воздействовать через программы повышения занятости населения, и среднедушевого дохода.

Для моделей пространственной эконометрики мы видим, что по всем рассматриваемым переменных наблюдается положительная пространственная автокорреляция. В наибольшей степени она сильна по переменной x2 (cреднедушевые денежные доходы населения), что о говорит о существовании выраженных кластеров в пространстве, вероятна сильная кооперация между соседними регионами. Наименьшее значение наблюдается по показателю x4 (объем инновационных товаров, работ, услуг на душу населения) – кластеризация выражена в наименьшей степени, распределение регионов более рассеянное, возможна повышенная конкуренция между ними.

Рассматривая в динамике между годами, можно сказать, что наибольшее относительное увеличение степени кластеризации за год произошло по показателю x5 (инвестиции в основной капитал на душу населения) – на 30%, а наибольшее увеличение рассеяния – по переменной y (ВРП на душу населения), на 24%.

При проведении экономической политики также необходимо учитывать взаимодействие и взаимовлияние регионов друг на друга. Особенно полезно учитывать такие межрегиональные эффекты при разработке и проектировании разного рода кластерных структур, таких, например, как индустриальные парки, технопарки, свободные экономические зоны, бизнес-парки и т.п, способствующих появлению выраженной положительной автокорреляции, часто проявляющейся в кооперации регионов между собой и появлению т.н. «полюсов роста». Концентрация экономической активности согласно теориям известных учёных-регионалистов, ведет к диффузии инноваций на периферийные территории – это приводит к конвергенции регионов по социально-экономическим показателям и, в итоге, к уменьшению межрегионального неравенства.

**СПИСОК ЛИТЕРАТУРЫ**

1. Ивченко Ю.С. (2019). Определение основных факторов уровня валового регионального продукта методами эконометрического моделирования по совокупности регионов Российской Федерации. *Статистика и Экономика*. 16(6):4-18.

2. Скуфьина Т.П., Баранов С.В., Корчак Е.А. (2018). Оценка влияния динамики инвестиций на рост валового регионального продукта в регионах Севера и Арктической зоны Российской Федерации. *Вопросы статистики*. 25(6):25-35.