**Министерство науки и высшего образования Российской Федерации**

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ УПРАВЛЕНИЯ»**

Изображение выглядит как текст, гаечный ключ, инструмент

Автоматически созданное описание

|  |  |
| --- | --- |
| Институт | Информационных систем |
| Кафедра | Математических методов в экономике и управлении |

**Семестровая работа**

|  |  |  |  |
| --- | --- | --- | --- |
| Дисциплина | Методы и модели Эконометрики | | |
| Тема работы | Исследование фондоотдачи в отраслях Российской Федерации на основе методов и моделей эконометрики | | |
|  | | | |
| Направление подготовки | 01.03.02 |  | Прикладная математика и информатика | |
| (код) |  | (наименование) | |
| Образовательная программа | Прикладная математика и информатика | | |
| (название образовательной программы) | | |
| Обучающийся | Чернов Антон Георгиевич | | |
|  | (Фамилия, имя, отчество) | | |
| III курс, 2 подгруппа |
| (курс, номер группы) |
| Руководитель работы | Заведующая кафедрой, к.э.н. Писарева О. М. | | |
| (ученая степень, звание, Фамилия и Инициалы) | | |

|  |  |
| --- | --- |
| Оценка: |  |
| Подпись руководителя: |  |

СОДЕРЖАНИЕ

[СОДЕРЖАНИЕ 2](#_Toc162081410)

[ВВЕДЕНИЕ 4](#_Toc162081411)

[1. Использованные данные 5](#_Toc162081412)

[2. Кластеризация 6](#_Toc162081413)

[2.1 Методология кластеризации 6](#_Toc162081414)

[2.2 Интерпретация кластеров 9](#_Toc162081415)

[3. Вывод уравнения регрессии фондоотдачи 10](#_Toc162081416)

[4. Регрессионный анализ 11](#_Toc162081417)

[4.1 Построение значимых моделей на кластерах 11](#_Toc162081418)

[4.2 Анализ свойств моделей 14](#_Toc162081419)

[4.2.1 Проверка на нормальность остатков 14](#_Toc162081420)

[4.2.2 Проверка на мультиколлинеарность 14](#_Toc162081421)

[4.2.3 Проверка на гетероскедастичность 15](#_Toc162081422)

[4.2.4 Устранение гетероскедастичности 16](#_Toc162081423)

[4.2.5 Проверка на автокорреляцию 17](#_Toc162081424)

[4.2.6 Устранение автокорреляции 19](#_Toc162081425)

[4.3 Итоговые модели 21](#_Toc162081426)

[ЗАКЛЮЧЕНИЕ 24](#_Toc162081427)

[СПИСОК ИСТОЧНИКОВ 25](#_Toc162081428)

ВВЕДЕНИЕ

Из экономической теории известно, что производственная функция Кобба-Дугласа и справедлива для предприятий внутри одной отрасли, однако как показал обзор литературы, на межотраслевом уровне их актуальность ещё не была исследована.

Особый интерес представляет оценка эффективности использования основных фондов, в частности, фондоотдачи, которая непосредственно связана с затратами на труд и основные фонды (капитал) через производственную функцию Кобба-Дугласа.

В текущей работе проверяется гипотеза о том, что группировка отраслей по схожести структуры их основных фондов позволит достаточно точно оценивать изменение фондоотдачи этих отраслей в зависимости от изменения затрат на труд и основные фонды. Таким образом, исследование строится в два этапа: сначала производится кластеризация, а затем регрессионный анализ.

1. Использованные данные

В исследовании использовались данные Росстата [1] о 18-ти видах экономической деятельности (далее – ВЭД) согласно общероссийскому классификатору ОКВЭД2 за 6 лет – с 2017 по 2022. Все было отобрано 7 признаков: данные о долях видов ОФ (всего четыре), которые использовалась в кластеризации, и еще три используемые в регрессии, представляющие собой труд капитал и стоимость выпущенной продукции.

Всего различают пять видов ОФ: «Сооружения», «Машины и оборудование», «Жилые здания», «Нежилые здания», однако для увеличения точности кластеризации за счет уменьшения размерности последние два были объединены в единый вид ОФ «Здания».

На основании данных о среднегодовом наличии ОФ для каждого ВЭД сначала рассчитывались значения долей видов ОФ за каждый год с 2017 по 2022. Затем для каждого ВЭД рассчитывались средние доли видов ОФ за 6 лет, которые в итоге использовались как факторы для кластеризации ВЭД.

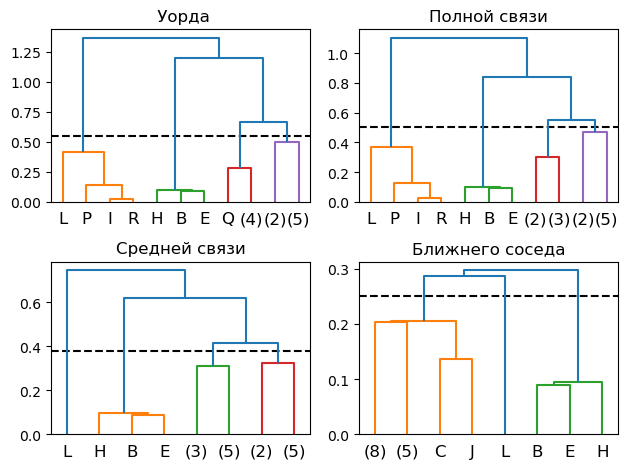
В эконометрической модели (4) использовались годовые показатели ВЭД, описание которых приведено в таблице 1. При этом показатели затраты на рабочую силу и фондоотдача были рассчитаны на основе промежуточных показателей.

Таблица 1 – Переменные (показатели) эконометрической модели

|  |  |  |  |
| --- | --- | --- | --- |
| Тип | Обозначение | Наименование показателя | Формула расчета |
| Промежуточные |  | Среднегодовая численность занятых, чел. | - |
|  | Среднемесячная номинальная начисленная заработная плата, руб./чел. | - |
|  | Валовая добавленная стоимость, млн. руб. | - |
| Основные |  | Среднегодовое наличие основных фондов, млн. руб. | - |
|  | Суммарные годовые затраты на рабочую силу, млн. руб. |  |
|  | Год | - |
|  | Фондоотдача, руб./руб. |  |

1. Кластеризация
   1. Методология кластеризации

Были использованы четыре вида иерархической агломеративной кластеризации, а именно методы Уорда, полной связи, средней связи и ближнего соседа с помощью библиотеки Python «scipy.cluster». В результате работы программы были получены так называемые дендрограммы – диаграммы последовательности объединения кластеров в иерархию (Рис. 1).



Латинскими буквами обозначены отрасли, числами в скобках – количество отраслей, входящих в кластер

Рисунок 1 – Дендрограммы методов кластеризации.

По полученным дендрограммам производился выбор разбиения на кластеры так, чтобы получить как можно более однородные кластеры, и таким образом добиться наилучшего качества регрессии на них. При этом было учтено основное свойство иерархической кластеризации – чем больше кластеров (и меньше элементов в них), тем больше их однородность. Таким образом выбиралось как можно более детализированное разбиение, при этом следя за тем, чтобы в каждом кластере было не слишком маленькое количество элементов, достаточное для статистической значимости соответствующих регрессий.

Минимальное количество элементов в одном кластере для -го признака рассчитывалось по формуле:

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

где – значение критической точки двустороннего распределения Стьюдента с уровнем значимости и степенями свободы ,

– стандартное отклонение -го признака общей выборки,

– количество элементов в общей выборке,

– предельная ошибка (точность), принятая равной .

Можно показать, что если рассчитывать предельную ошибку для каждого признака одинаковым образом, то рассчитанное количество элементов не будет зависеть от конкретных стандартных отклонений различных признаков. Таким образом, минимальное количество элементов в одном кластере будет одинаковым: . В нашем случае при размере исходной выборки минимальное количество наблюдений в одном кластере оказалось равно 16. Учитывая, что по каждой отрасли было собрано по 6 наблюдений (по одному за каждый год с 2017 по 2022), минимальное количество отраслей в одном кластере составило 3.

Итак, для каждого метода кластеризации было выбрано такое разбиение на кластеры, чтобы каждый кластер содержал как можно меньше отраслей, но не менее трех. На дендрограммах (Рис. 1) показаны полученные разбиения: цветом выделены минимальные кластеры, а пунктирной линией – место разреза дерева для получения этих кластеров.

Далее, для каждого метода была посчитана суммарная внутригрупповая дисперсия разбиения в зависимости от количества выбранных кластеров. На соответствующих графиках (Рис. 2) можно увидеть так называемые точки перегиба, соответствующие количеству кластеров, при увеличении дальше которого внутригрупповая средняя уменьшается несущественно. Оказалось, что выбранное нами количество кластеров достаточно близко к оптимальному для всех методов кластеризации, кроме метода ближнего соседа.

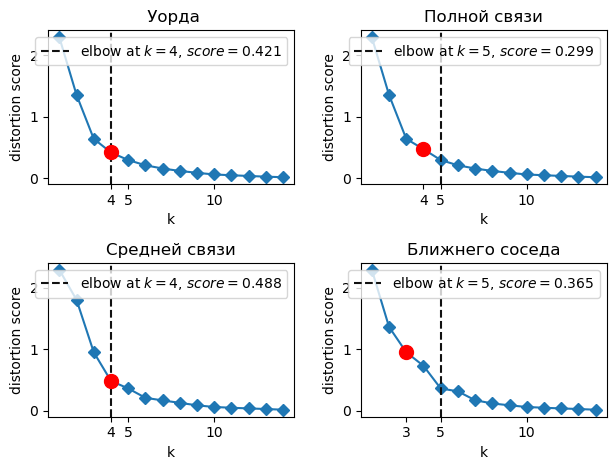


Рисунок 2 – Графики зависимости суммарной внутригрупповой дисперсии от количества кластеров

В итоге лучшим оказалось разбиение на кластеры методом Уорда, так как значение внутригрупповой дисперсии для выбранного количества кластеров было наименьшим (Таблица 2).

Таблица 2 – Зависимость средней внутригрупповой дисперсии от количества кластеров в разбиении для каждого метода

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Количество кластеров | Уорда | Полной связи | Средней связи | Ближнего соседа |
| 1 | 2.29 | 2.29 | 2.29 | 2.29 |
| 2 | 1.36 | 1.36 | 1.81 | 1.37 |
| 3 | 0.64 | 0.64 | 0.95 | 0.95 |
| 4 | **0.42** | 0.48 | 0.49 | 0.73 |

* 1. Интерпретация кластеров

Анализируя центроиды кластеров, полученных методом Уорда, (Таблица 3) были сделаны следующие выводы об их профилях.

Во-первых, выделились два кластера с доминирущим видом основных фондов в структуре основных фондов. Так, в первый кластер попали отрасли, наиболее активно задействующие здания, такие как «Образование» и «Деятельность гостиниц и предприятий общественного питания». Во втором кластере доминирующим видом основных фондов оказались сооружения. В него вошли такие отрасли, как «Добыча полезных ископаемых» и «Транспортировка и хранение».

Остальные два кластера не имеют такой узкой направленности. В третий кластер попали отрасли, в равной степени задействующие здания и оборудование, такие как «Здравоохранение» и «Торговля». В четвертый кластер были объединены отрасли, в равной степени задействующие каждый из видов основных фондов, кроме транспортных средств, например, «Строительство» и «Административная деятельность».

Таблица 3 – Центры полученных кластеров

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Номер кластера | Здания | Машины и оборудование | Сооружения | Транспортные средства |
| 1 | **0.71** | 0.12 | 0.10 | 0.02 |
| 2 | 0.09 | 0.14 | **0.70** | 0.06 |
| 3 | **0.41** | **0.31** | 0.12 | 0.09 |
| 4 | **0.21** | **0.34** | **0.31** | 0.08 |

1. Вывод уравнения регрессии фондоотдачи

Фондоотдача (англ. RoFA, return on fixed assets) по определению есть отношение стоимость произведенной продукции к задействованным основным фондам *К:*

|  |  |  |
| --- | --- | --- |
|  |  | () |

Из экономической теории известно, что стоимость продукции связана с затратами на труд и затратами на основные фонды производственной функцией Кобба-Дугласа. С учетом технического прогресса функция имеет вид:

|  |  |  |
| --- | --- | --- |
|  |  | () |

где – время,

– затраты на трудовые ресурсы,

– темп прироста объема производства благодаря техническому прогрессу.

Разделив обе части уравнения (2) на и учитывая (1), получим производственную функцию для фондоотдачи:

|  |  |  |
| --- | --- | --- |
|  |  | () |

где замена введена для удобства

1. Регрессионный анализ
   1. Построение значимых моделей на кластерах

На каждом из четырех полученных кластеров была оценена линеаризованная функции фондоотдачи:

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Уравнения регрессии по всем кластерам оказались значимыми, причем коэффициенты оказались значимыми во всех кластерах, кроме одного – в модели по кластере с номером 3 коэффициенты (константа) и незначимы.

С учетом того, что кластеризация производилась не по тем же признакам, по которым строилась регрессия, т.е. суммарная внутригрупповая дисперсия минимизировалась по долям видов ОФ, по остальным трем признакам фондоотдаче, труду и капиталу всё ещё возможно улучшить качество разбиения за счет перенесения наиболее аномальных значений в ближайшие к ним кластеры. Вместе с увеличением однородности кластеров возможно улучшение показателей качества значимых уравнений регрессии, в том числе и значимости коэффициентов.

Качество данных в кластере 3 оценивалось по графикам разброса (рисунок 3) и было сделано предположение об аномальности наблюдений отрасли G. Соответствующая гипотеза была проверена с помощью расстояния Махаланобиса, которое отражает расстояние от объекта до центра многомерного нормального распределения в стандартных отклонениях и вычисляется по формуле:

где – наблюдение, проверяемое на аномальность,

– матожидание выборочного распределения,

– симметричная положительно определенная матрица ковариации выборки.

Квадрат расстояния Махаланобиса – случайная величина, распределенная по закону с количеством степеней свободы, равным количеству признаков в выборке (в нашем случае ). Гипотеза об аномальности одного наблюдения формулируется следующим образом:

На уровне значимости была проверена гипотеза для всех наблюдений кластера 3, сравнивая соответствующие рассчитанные расстояния с критическим значением . Оказалось, что только два наблюдения отрасли G являются аномальными по расстоянию Махаланобиса (на рисунке 3 отмечены крестиками).

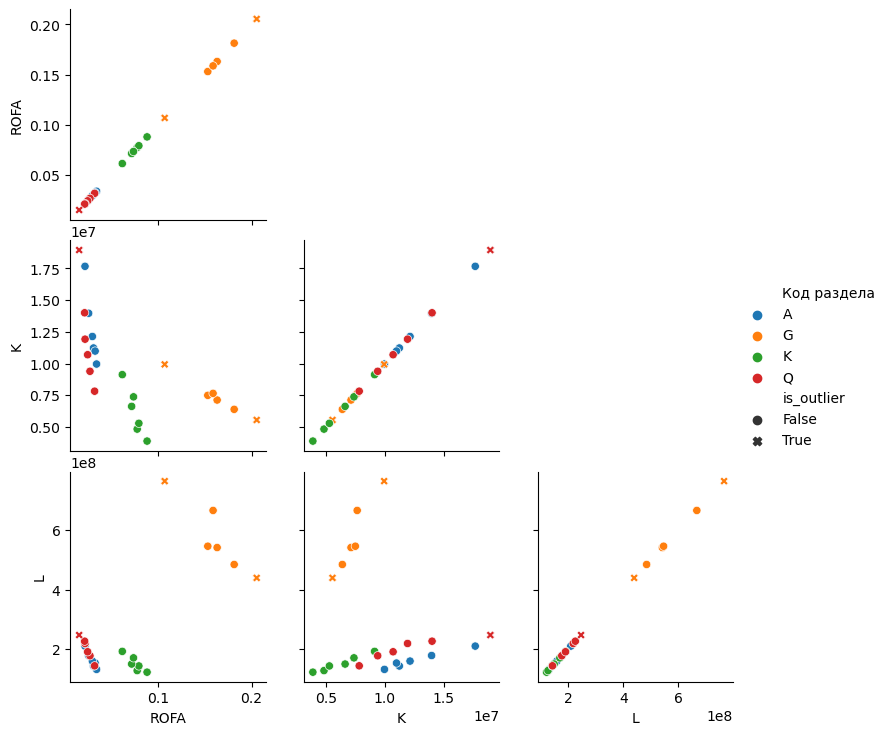


Рисунок 3 – В кластере 3 все наблюдения отрасли G находятся далеко от остальных по основным переменным уравнения регрессии

Несмотря на то, что гипотеза об аномальности отклоняется для большинства наблюдений отрасли G, наблюдения этой отрасли в целом всё ещё находятся существенно вдалеке от остальных, и было принято решение перенести эту отрасль в другой ближайший кластер. Сравнив рассчитанные евклидовы расстояния от центра этой отрасли до центров всех кластеров, оказалось, что отрасль G ближе всего находится к кластеру 4, в связи с чем была перенесена в него.

Была измерена суммарная внутригрупповая дисперсия кластера 3 до и после перенесения и оказалось, что после перенесения она уменьшилась. Также были пересчитаны регрессии по кластерам 3 и 4 и оказалось, что оба уравнения значимы и оба имеют значимые коэффициенты.

* 1. Анализ свойств моделей

После оценки коэффициентов моделей на кластерах они были проверены нормальность остатков, а также на наличие мультиколлинеарности, гетероскедастичности и автокорреляции.

Проверка на нормальность остатков производилась с помощью теста Шапиро-Уилка, суть которого заключается в сравнении упорядоченных и стандартизованных квантилей выборочного распределения с квантилями нормального распределения.

Нулевая гипотеза о нормальности выборочных значений остатков не была отвергнута только для модели по первому кластеру. Таким образом, для моделей по кластерам 2, 3, 4 нельзя судить о значимости коэффициентов, а также нельзя оценить доверительные интервалы коэффициентов.

Таблица 4 – Результаты теста Шапиро-Уилка

|  |  |
| --- | --- |
| cluster\_id | p-значение |
| 1 | **0.71** |
| 2 | 0.02 |
| 3 | 0.03 |
| 4 | 0.06 |

Проверка на мультиколлинеарность производилась с помощью коэффициента инфляции дисперсии :

где – коэффициент детерминации уравнения регрессии на все остальные экзогенные переменные модели.

Значения -коэффициента лежат в интервале от 1 до бесконечности, причем значения, лежащие близко к 1, свидетельствуют об отсутствии мультиколлинеарности, а значения, большие 3 – о её наличии.

Для каждой переменной оцениваемой модели были рассчитаны -коэффициенты на данных каждого кластера (таблица 3). Только в регрессии, построенной по второму кластеру, была обнаружена мультиколлинеарность между переменными и .

Таблица 3 – -коэффициенты

|  |  |  |  |
| --- | --- | --- | --- |
| Номер  кластера | для | для | для |
| 3 | 1.6002 | 1.1659 | 1.7519 |
| 2 | **28.8948** | **28.8565** | 1.0253 |
| 4 | 1.1772 | 1.1957 | 1.0996 |
| 1 | 1.1123 | 1.1415 | 1.0875 |

Вследствие мультиколлинеарности в модели по второму кластеру оцененные коэффициенты имеют сравнительно большие стандартные ошибки (таблица 4).

Также из-за мультиколлинеарности сильно коррелированные переменные имеют неправильные с точки зрения экономической интерпретации знаки. Фондоотдача, по определению (1) обратно зависит от капитала и, согласно выведенной формуле (3), зависит прямым образом от труда, однако во второй модели коэффициент при логарифме капитала имеет положительный знак, а при логарифме труда – отрицательный (таблица 4).

Таблица 4 – Оцененные параметры моделей и их ошибки

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Коэффициенты | | | | Стандартная ошибка | | | |
| Номер кластера | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| const | 251,07 | 230,94 | -92,02 | 166,27 | 56,00 | 83,38 | 72,64 | 35,93 |
| ln\_K | -0,36 | **1,80** | -1,67 | -0,81 | 0,03 | **0,32** | 0,15 | 0,04 |
| ln\_L | 0,26 | **-1,99** | 0,78 | 0,79 | 0,08 | **0,37** | 0,09 | 0,05 |
| t | -0,13 | -0,11 | 0,05 | -0,08 | 0,03 | 0,04 | 0,04 | 0,02 |

Проверка на гетероскедастичность остатков построенных моделей производилась с помощью теста Уайта, в котором предполагается, что дисперсии ошибок регрессии представляют собой одну и ту же функцию от наблюдаемых значений регрессоров:

Если такая функциональная зависимость существует, то существует статистическая зависимость между ошибками регрессии и значениями регрессоров:

В тесте Уайта включает все регрессоры, их квадраты и все их попарные произведения.

Проверяемая гипотеза об отсутствии гетероскедастичности есть то же самое что и гипотеза о значимости этого уравнения регрессии и проверяется по статистике множителей Лагранжа и имеет распределение с степенями свободы.

По полученным -значениям -статистики теста Уайта был сделан вывод о наличии гетероскедастичности в кластерах с номерами 1 и 4 на уровне значимости 5%.

Таблица 4 – Результаты теста Уайта

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Номер кластера | 1 | 2 | 3 | 4 |
| LM-статистика | **0.0328** | 0.1053 | 0.0664 | **0.0009** |

Устранение гетероскедастичности моделей, построенных на кластерах, производилось с помощью взвешенного метода наименьших квадратов, в рамках которого оценивается регрессия с переменными, нормированными на свои стандартные отклонения:

Так как неизвестны, они заменяются состоятельными оценками согласно предположениям теста Уайта, для чего оценивается регрессия вида:

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

где – ошибки исходной гетероскедастичной модели.

В отличии от оригинального теста Уайта, в регрессии (5) в качестве зависимой переменной использовались не квадраты ошибок, а их логарифмы, чтобы прогнозные значения не принимали отрицательные значения и возможно было выполнить обратное преобразование .

Модели для кластеров 1 и 4, оцененные взвешенным МНК, оказались обладающими гетероскедастичностью по тесту Уайта, так как соответствующие -значения и оказались меньше и в дальнейшем исследовании не участвовали.

Проверка на автокорреляцию предварительно производилась по коррелограммам с максимальным лагом на уровне (Рис. 4). Из них можно предположить наличие автокорреляции второго порядка со значением примерно в моделях по кластерам 1 и 3, а также первого порядка в модели по кластеру 2 и третьего – по кластеру 4.

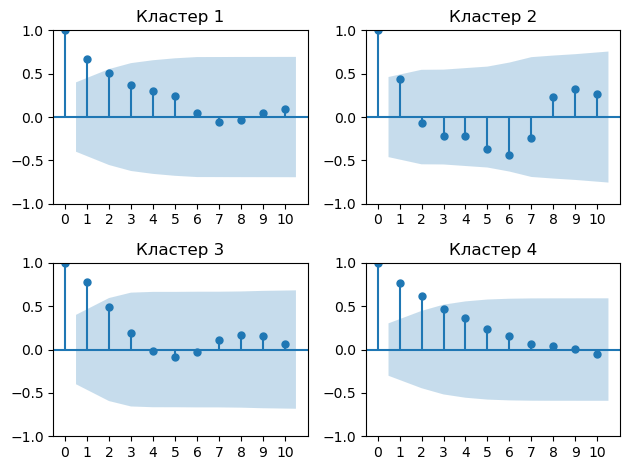


Рисунок 4 – Коррелограммы моделей, построенных по кластерам

Формальная проверка на автокорреляцию остатков построенных моделей производилась с помощью статистики Дарбина-Уотсона:

где и – ошибки соседних по времени наблюдений.

Несмотря на то, что тест Дарбина-Уотсона не является статистическим критерием, так как распределение его статистики зависит не только от числа наблюдений, но и от количества регрессоров, для него существуют пороговые значения и , по которым можно судить о наличии автокорреляции и о направлении её связи. Соответствующие интервалы здесь и далее условимся обозначать цветами, как показано на рисунке 5.

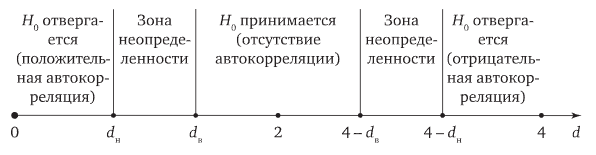


Рисунок 5 – Интерпретация статистики Дарбина-Уотсона

На основании применения теста Дарбина-Уотсона были сделаны следующие выводы. В построенных на кластерах 1, 3, и 4 моделях присутствует авторегрессия, а о наличии или отсутствии автокорреляции в кластере 2 судить невозможно.

Таблица 4 – Рассчитанные и табличные значения -статистики

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| cluster\_id | n | DW |  |  |  |  |
| 1 | 24 | 0.59 | 0.88 | 1.41 | 2.59 | 3.12 |
| 2 | 18 | 1.01 | 0.71 | 1.42 | 2.58 | 3.29 |
| 3 | 24 | 0.33 | 0.88 | 1.41 | 2.59 | 3.12 |
| 4 | 42 | 0.38 | 1.15 | 1.46 | 2.54 | 2.85 |

Чтобы уточнить наличие автокорреляции в регрессии по второму кластеру был проведен тест Бреуша-Годфри, который основан на оценке вспомогательной регрессии остатков от своих лаговых значений порядков :

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

Статистика теста представляет собой -статистику этой регрессии – если она значима, то имеет место автокорреляция. При этом наличие лагов определенного порядка будет определяться значимостью соответствующих коэффициентов.

В результате проведенного теста оказалось, что на уровне значимости 5% гипотеза о наличии автокорреляции в регрессии по второму кластеру впервые отклоняется при порядке автокорреляции , а значит максимальный размер лага автокорреляции равен 1. При этом для моделей по остальным кластерам тест показал наличие автокорреляции не меньше 6 порядка, что согласуется с тестом Дарбина-Уотсона.

Таблица 5 – Рассчитанные -значения -статистики теста Бреуша-Годфри для различных максимальных порядков автокорреляции

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер кластера | Порядок лага | | | | | Показатели |
| 1 | 2 | 3 | 4 | 5 |
| 1 | **0,00** | 0,26 | 0,35 | 0,12 | 0,15 | BG t-stat p-value |
| **0,00** | **0,00** | **0,01** | **0,01** | **0,01** | BG F-stat p-value |
| 2 | **0,04** | 0,89 | 0,44 | 0,92 | 0,93 | BG t-stat p-value |
| 0,08 | 0,10 | 0,22 | 0,17 | 0,13 | BG F-stat p-value |
| 3 | **0,00** | 0,80 | 0,71 | 0,49 | 0,17 | BG t-stat p-value |
| **0,00** | **0,00** | **0,00** | **0,00** | **0,00** | BG F-stat p-value |
| 4 | **0,00** | 0,25 | 0,54 | 0,43 | 0,67 | BG t-stat p-value |
| **0,00** | **0,00** | **0,00** | **0,00** | **0,00** | BG F-stat p-value |

Устранение автокорреляции производилось с помощью итеративной процедуры Кохрейна-Оркатта, в рамках которого предполагается наличие авторегрессии первого порядка:

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

В рамках алгоритма используется преобразование исходных переменных:

,

где оценка коэффициента автокорреляции , получена по уравнению (7) для остатков исходной модели.

Для первого наблюдения вводится так называемая поправка Прайса-Уинстона:

Новая модель по преобразованным переменным проверяется на наличие автокорреляции. Если гипотеза об отсутствии автокорреляции все еще отвергается, то процедура применяется к преобразованным переменным, при этом устраняя автокорреляцию порядка на один выше. Таким образом, процедура Кохрейна-Оркатта рекурсивно понижает порядок автокорреляции.

В процессе применения итеративной процедуры Кохрейна-Оркатта к моделям, построенным на кластерах на каждой итерации, наличие автокорреляции проверялось сразу по двум тестам – Дарбина-Уотсона и Бреуша-Годфри. Всего было проведено 5 итераций.

Результаты тестов приведены в таблице 6. Для теста Бреуша-Годфри для каждого порядка лага указывается -значения для -статистики уравнения (6) с максимальным порядком лага равным , а также для -статистики коэффициента при лаге этого уравнения. Для теста Дарбина-Уотсона приводятся значения статистики теста, обозначенные цветом в соответствии с принятым выше обозначением (см. рисунок 5).

Таблица 6 – Результаты тестов Дарбина-Уотсона (DW) и Броуша-Годфри (BG) для различных порядков лага

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Номер кластера | Статистики тестов | Порядок лага | | | | | |
| 0 | 1 | 2 | 3 | 4 | 5 |
| 1 | DW-stat | 0,59 | 1,87 | 2,77 | 3,06 | 3,24 | 3,37 |
| BG t-stat p-value |  | **0,00** | 0,26 | 0,35 | 0,12 | 0,15 |
| BG F-stat p-value |  | **0,00** | **0,00** | **0,01** | **0,01** | **0,01** |
| 2 | DW-stat | 1,01 | 1,35 | 1,54 | 1,57 | 1,15 | 1,15 |
| BG t-stat p-value |  | **0,04** | 0,89 | 0,44 | 0,92 | 0,93 |
| BG F-stat p-value |  | 0,08 | 0,10 | 0,22 | 0,17 | 0,13 |
| 3 | DW-stat | 0,33 | 1,24 | 2,70 | 3,22 | 3,18 | 3,12 |
| BG t-stat p-value |  | **0,00** | 0,80 | 0,71 | 0,49 | 0,17 |
| BG F-stat p-value |  | **0,00** | **0,00** | **0,00** | **0,00** | **0,00** |
| 4 | DW-stat | 0,38 | 1,69 | 2,94 | 3,36 | 3,53 | 3,63 |
| BG t-stat p-value |  | **0,00** | 0,25 | 0,54 | 0,43 | 0,67 |
| BG F-stat p-value |  | **0,00** | **0,00** | **0,00** | **0,00** | **0,00** |

По полученным результатам тестов были определены порядки автокорреляции, устраняемой процедурой Кохрейна-Оркатта. Для моделей по кластерам 1 и 4 выводы обоих тестов согласуются – устраняемая автокорреляция имеет первый порядок.

Для модели по кластеру 3 также возможно устранить автокорреляцию первого порядка. Несмотря на то, что для первого лага -статистика находится в промежутке неопределенности, тест Броуша-Годфри обнаруживает автокорреляцию.

В модели по второму кластеру подтверждается сделанный ранее вывод об отсутствии автокорреляции, несмотря на то что для лагов порядков 2 и 3 тест Дарбина-Уотсона показывает наличие автокорреляции. Значения -статистики для этих порядков несущественно преодолевают зону неопределенности, в то время как по тесту Броуша-Годфри ни для одного порядка лага нельзя отвергнуть гипотезу об отсутствии автокорреляции.

* 1. Итоговые модели

Итак, после перенесения наблюдений одной отрасли в другой кластер были получены значимые модели со значимыми коэффициентами.

Затем была произведена попытка устранить гетероскедастичность в кластерах 1 и 4 с помощью ВМНК, которая не дала результатов, поэтому полученные с помощью него модели в дальнейшем не рассматривались.

Далее была успешно устранена автокорреляция в моделях по кластерам 1, 3, 4, которые, помимо сохранения значимости в целом и значимости коэффициентов по отдельности, как оказалось имеют больший , ввиду чего были выбраны в качестве окончательных для этих кластеров. Для кластера 2 автокорреляция отсутствовала изначально, поэтому финальной моделью для него стала самая первая модель.

Показатели качества и свойств указанных моделей обобщены в таблице 7, где код модели означает этап, на котором модель была получена: 0 для моделей после скорректированной кластеризации, и 1 – для моделей, полученных после устранения автокорреляции. Жирным выделены финальные модели.

Таблица 7 – Показатели качества и свойств построенных моделей

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| cluster\_id | model\_code | R\_sq | R\_sq\_adj | Shapiro-Wilk | D-W | White\_pval |
| 1 | 0 | 0.89 | 0.87 | 0.71 | 0.59 | 0.03 |
| **1** | **0.77** | **0.73** | **0.01** | **1.80** | **0.01** |
| 2 | **0** | **0.73** | **0.67** | **0.02** | **1.01** | **0.11** |
| 1 | 0.44 | 0.32 | 0.99 | 1.31 | 0.25 |
| 3 | 0 | 0.93 | 0.92 | 0.03 | 0.33 | 0.07 |
| **1** | **0.94** | **0.93** | **0.00** | **1.65** | **0.01** |
| 4 | 0 | 0.95 | 0.94 | 0.06 | 0.38 | 0.00 |
| **1** | **0.91** | **0.90** | **0.00** | **2.17** | **0.00** |

Рассчитанные параметры моделей приведены в таблице 8. Так как нормальность остатков нарушена во всех финальных моделях, нельзя учитывать значимость коэффициентов и доверять построенным по ним интервальным оценкам.

Таблица 8 – Рассчитанные коэффициенты моделей и их стандартные ошибки

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| cluster\_id | model\_code | indicator | const | ln\_K | ln\_L | t |
| 1 | 0 | coef | 251.07 | -0.36 | 0.26 | -0.13 |
| se | 56.00 | 0.03 | 0.08 | 0.03 |
| **1** | **coef** | **-0.22** | **-0.39** | **0.20** | **-0.00048** |
| **se** | **0.22** | **0.06** | **0.14** | **0.00** |
| 2 | **0** | **coef** | **230.94** | **1.80** | **-1.99** | **-0.00036** |
| **se** | **83.38** | **0.32** | **0.37** | **0.04** |
| 1 | coef | -1.11 | 0.97 | -1.01 | -0.00 |
| se | 0.51 | 0.50 | 0.57 | 0.00 |
| 3 | 0 | coef | -92.02 | -1.67 | 0.78 | 0.05 |
| se | 72.64 | 0.15 | 0.09 | 0.04 |
| **1** | **coef** | **-0.07** | **-1.42** | **0.86** | **0.00174** |
| **se** | **0.08** | **0.11** | **0.09** | **0.00** |
| 4 | 0 | coef | 166.27 | -0.81 | 0.79 | -0.08 |
| se | 35.93 | 0.04 | 0.05 | 0.02 |
| **1** | **coef** | **-0.15** | **-0.99** | **0.67** | **0.00031** |
| **se** | **0.11** | **0.06** | **0.06** | **0.00** |

Таким образом, с учетом обратного преобразования логарифмированных моделей, уравнение фондоотдачи для группы отраслей, наиболее активно задействующей здания, имеет вид:

|  |  |  |
| --- | --- | --- |
|  |  |  |

Для группы отраслей, наиболее активно задействующей сооружения:

|  |  |  |
| --- | --- | --- |
|  |  |  |

Для группы отраслей, наиболее активно задействующей в равной степени здания и оборудования:

|  |  |  |
| --- | --- | --- |
|  |  |  |

Для группы отраслей, наиболее активно задействующей в равной степени здания и оборудования:

|  |  |  |
| --- | --- | --- |
|  |  |  |

ЗАКЛЮЧЕНИЕ

В текущей работе отрасли экономики РФ были сгруппированы по схожести структуры их основных фондов на четыре группы. Это позволило достаточно точно оценивать изменение фондоотдачи в зависимости от изменения затрат на труд и основные фонды для отраслей, принадлежащих различным группам.

Так, в первый кластер были отнесены отрасли, наиболее активно задействующие здания, во втором кластере доминирующим видом основных фондов оказались сооружения.

Остальные два кластера не имеют такой узкой направленности. В третий кластер попали отрасли, в равной степени задействующие здания и оборудование. В четвертый кластер были объединены отрасли, в равной степени задействующие каждый из видов основных фондов, кроме транспортных средств.

СПИСОК ИСТОЧНИКОВ

1. https://rosstat.gov.ru/statistics/accounts
2. ЭКОНОМЕТРИКА 4-е изд., испр. и доп. Учебник и практикум для вузов (Н. Ш. Кремер, Б. А. Путко ; под редакцией Н. Ш. Кремера.)
3. ЭКОНОМЕТРИКА. Учебник для вузов (Под ред. Елисеевой И.И.)