В условиях современной экономики как бы ни была организована государственная система пенсионного обеспечения, ее основными задачами выступают предотвращение бедности среди пенсионеров и компенсация заработка, утраченного в связи с наступлением событий, перечисленных в законодательстве, - достижением определенного возраста, наступлением инвалидности, потерей кормильца и пр. Для того чтобы успешно решать эти задачи в пенсионной системе должны быть сбалансированы поступления и обязательства. Поэтому важнейшими целями любой пенсионной реформы будут достижение долгосрочной финансовой сбалансированности и обеспечение социально приемлемого уровня пенсионного обеспечения.

Бюджет Пенсионного фонда РФ формируется за счет страховых взносов работающего населения. В связи с их недостаточностью он восполняется средствами государственного бюджета, в виде дотаций, займов из федеральной казны. Данная тенденция является негативной, поскольку сокращение доли собственных поступлений негативно сказывается не только на бюджете фонда, но и государственных финансов в целом.

В связи с увеличивающимся дефицитом бюджета Пенсионного фонда РФ проблема реформирования пенсионной системы приобрела в последнее время еще большую актуальность.

В рамках данной работы были определены те социально-экономические факторы, которые наиболее влияют на уровень страховых взносов, построены модели зависимости уровня страховых взносов от данных факторов, что в свою очередь позволит повысить собираемость взносов в регионах.

Для этого средствами R-studio были построены несколько эконометрических моделей, проведено сравнение данных моделей и выявление из них наилучшей. По результатам моделирования также были сделаны выводы по оптимизации деятельности Пенсионного фонда.

Для выявления и формализации взаимосвязей между взносами в Пенсионный Фонд и социально-экономическими показателями развития региона в данном исследовании используются многомерные статистические методы: кластерный анализ и модели панельных данных.

Перед построением эконометрических моделей зависимости уровня пенсионных взносов в ПФР от социально-экономических факторов, в первую очередь была сформирована база статистических данных, в которую вошли следующие факторы:

- внутренний региональный продукт;

- среднемесячная заработная плата;

- средняя численность занятых в текущем году, человек;

- количество безработных;

- средняя численность экономически активного населения;

- миграционный прирост населения;

- численность пенсионеров;

- количество предприятий на конец года.

Данные были взяты из общедоступного ресурса сети интернет «Федеральной Службы Государственной Статистики» за 10 лет с 2008 по 2017гг.

После создания базы данных был произведен кластерный анализ. Для этого данные были расчитаны на доли, а в програмном средстве R-studio были построены соответствующие каждому из годов дендрограммы.

Дендрограммы были построены для каждого года. Анализ каждой из них показал, что для каждого года регионы находятся примерно в одних и тех же кластерах.

Далее методом k-средних осуществлялось разбиение выборки на два, три и четыре кластера соответственно. Обоснованность кластеризации была проверена с помощью индекса Ранда, который оказался самым высоким для данных 2017 года, поделенных на три кластера и равен 0.85.

В результате проведенной кластеризации оптимальным оказалось разбиение данных на три кластера.

На втором этапе работы были построены эконометрические модели панельных данных по полной выборке и с учетом кластеризации.

Для построения моделей по полной выборке была создана матрица с данными, описанными в начале статьи. В связи с тем, что численность взносов напрямую зависит от численности населения в регионе, чтобы избежать некорректного расчета и уйти от этой зависимости был произведен перерасчет исходных данных на душу населения, также набор данных был нормирован двумя способами:

1. Линейная нормировка по "минимаксу":



где Xij – порядковый номер

Xmin j – минимальное значение в столбце данных

Xmax j - максимальное значение в столбце данных.

1. Нормировка на “единичную дисперсию”:



где Xi – порядковый номер

X – среднее значение по столбцу данных

σ - стандартное отклонение для столбца данных.

В результате было получено две расчетные матрицы, по каждой из которых было построено три вида моделей панельных данных:

- Модель сквозной регрессии (с объединенными эффектами)

- Модель панельных данных со случайными эффектами. (модель со стохастическим индивидуальным эффектом),

- Модель панельных данных с фиксированными эффектами (модель с детерминированным индивидуальным эффектом)

Проведя сравнительный анализ, получили результат, представленный в таблицах:

Для первой матрицы:

Таблица 1 – Результаты моделирования для первой матрицы

|  |  |  |  |
| --- | --- | --- | --- |
| **Вид модели** | **R2** | **p-value** | **Критерий Фишера** |
| Модель сквозной регрессии | 0,56 | 2,2 × $10^{-16}$, менее | 93,7 |
| Модель со случайными эффектами | 0,53 | 2,2 × $10^{-16}$, менее | 459,0 |
| Модель с фиксированными эффектами | 0,57 | 2,2 × $10^{-16}$, менее | 81,5 |

Для второй матрицы:

Таблица 2 – Результаты моделирования для второй матрицы

|  |  |  |  |
| --- | --- | --- | --- |
| **Вид модели** | **R2** | **p-value** | **Критерий Фишера** |
| Модель сквозной регрессии | 0,56 | 2,2 × $10^{-16}$, менее | 144,1 |
| Модель со случайными эффектами | 0,51 | 2,2 × $10^{-16}$, менее | 458,3 |
| Модель с фиксированными эффектами | 0,57 | 2,2 × $10^{-16}$, менее | 152,7 |

Как видно из вышеприведенных таблиц, характеристики моделей построенных по разным нормировкам не сильно отличаются. Также можно увидеть, что наилучшей будет являться модель с наибольшим коэффициентом детерминации (R2) - это третья модель с фиксированными эффектами.

Для каждой модели и каждой нормировки были проведены три теста:F-test, Тест Хаусмана, Тест Бройша-Пагана

По итогам проведенных тестов, было сделано заключение, что именно модель с фиксированными эффектами при учете структуры панельных данных и с учетом долей позволяет получить значимый и обоснованный вариант моделирования, который можно использовать для оценки основного индикатора - уровня страховых взносов в ПФР.

Далее были построены модели с учетом кластерного анализа. В связи с тем, что модели, построенные по двум способам нормирования, не сильно отличаются друг от друга, для построения моделей с учетом кластерного анализа решено было взять только данные расчетной матрицы для первого способа нормирования.

Характеристики лучших моделей для каждого кластера приведены в таблице 3:

Таблица 3 - Характеристики лучших моделей для каждого кластера

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Номер кластера** | **Вид модели** | **R2** | **p-value** | **Критерий Фишера** |
| Первый кластер | Модель сквозной регрессии | 0,73 | 2,2 × $10^{-16}$, менее | 154,9 |
| Второй кластер | Модель с фиксированными эффектами | 0,63 | 2,2 × $10^{-16}$, менее | 146,8 |
| Третий кластер | Модель со случайными эффектами | 0,68 | 2,2 × $10^{-16}$, менее | 573,259 |

Как видно из таблицы, для каждого кластера лучшими являются разные модели, с различным набором значимых факторов, что говорит о том, что у каждого кластера имеются свои особенности, которые и были отражены в результатах моделирования.

Однако часть значимых факторов повторяется в каждой из построенных моделей. Повторяющимися являются следующие факторы:

* Внутренний региональный продукт (валовая добавленная стоимость в основных ценах), в рублях
* Среднемесячная заработная плата, в рублях
* Средняя численность занятых в текущем году, в долях на душу населения
* Численность пенсионеров, в долях на душу населения

Это позволяет предположить, что именно данные факторы наиболее влияют на размер взносов в каждом кластере.

По результатам исследований данной работы был сделан следующий вывод: гипотеза о том, что кластеризация позволяет повысить качество моделей, а также о том, что в каждом кластере различное влияние имеют определенные социально-экономические факторы, подтвердилась. Также результаты проведенных исследований позволяют сделать вывод о возможности применения моделей панельных данных для построения зависимости уровня страховых взносов в ПФР от социально-экономических факторов развития региона. Кластеризация регионов значительно улучшила качество построенных моделей.

Стоит отметить тот момент, что заранее предположить какая из моделей окажется наиболее подходящей для конкретного кластера и выборки без кластеризации, затруднительно. Поэтому, необходимо построение всех моделей и после этого выбор лучшей.

Также некоторые из построенных моделей имеют относительно низкий коэффициент детерминации. Таким образом, возможно, имеет смысл исключить из списка социально-экономических факторов некоторые из них или же, наоборот, дополнить данный список факторами, не рассматриваемыми в данной работе.