

Использование больших данных при анализе производительности труда

О. И. Дранко

*Институт проблем управления им. В. А. Трапезникова
Российской академии наук*

olegdranko@gmail.com

К. Ю. Трушин

*Институт проблем управления им. В. А. Трапезникова
Российской академии наук*

trushin@ipu.ru

Аннотация: В работе представлена модель роста ВВП, учитывающая ресурсные ограничения по инвестициям и трудовым ресурсам. В рамках модели задача распределения мероприятий по развитию сводится к задаче о дробном рюкзаке, оптимальное решение которой достигается ранжированием мероприятий по эффективности использования дефицитного ресурса. При ограничениях на трудовые ресурсы критерием эффективности выступает производительность труда. Проведены расчёты производительности труда для 993 тыс. организаций всех отраслей. Для обработки исходных данных использованы методы анализа больших массивов на основе открытой финансовой отчётности и сведений о численности персонала Федеральной налоговой службы России. Установлена статистически значимая степенная зависимость производительности труда от размера организации. Предложенный подход позволяет обосновать приоритеты развития отраслей и регионов на основе объективных данных о производительности труда, а также может быть использован для корректировки программ развития с учётом ограничений по кадровым ресурсам.

Ключевые слова: моделирование, индикативное планирование, управление эффективностью, показатели эффективности, прогнозирование, факторы роста, добавленная стоимость, производительность труда, анализ больших данных, мягкие вычисления, нечеткие множества

I. ВВЕДЕНИЕ

Производительность труда является одним из основных индикаторов экономического развития, определяющим динамику валового внутреннего продукта (ВВП) через добавленную стоимость. В условиях реализации национальных целей развития Российской Федерации ключевым ориентиром выступает устойчивое увеличение экономики, что непосредственно интерпретируется как рост ВВП. Достижение этого показателя в значительной степени обеспечивается повышением производительности труда, которое позволяет наращивать выпуск без пропорционального увеличения затрат ресурсов. В условиях ограниченности ресурсов и необходимости устойчивого экономического развития понимание факторов влияющих на производительность, а также возможность ее прогнозирования и управления приобретают стратегическое значение.

В последние годы для решения подобных задач все чаще используются методы мягких вычислений: совокупность вычислительных подходов, ориентированных на обработку неточной, неопределенной и частично формализованной информации. К ним относят различные аппроксиматоры,

нечетко-логические системы, алгоритмы оптимизации и их комбинации. В отличие от жестко детерминированных моделей, мягкие вычисления позволяют выявить сложные зависимости между совокупностью факторов и результативными показателями, адаптироваться к изменению структуры данных и учитывать экспертные знания в виде лингвистических правил.

Особый интерес представляет моделирование производительности труда в ее взаимосвязи с выручкой предприятия. Выручка выступает интегральным индикатором, аккумулирующим как объемные, так и ценовые факторы, при этом ее связь с производительностью не является прямолинейной: рост производительности может опережать или отставать от динамики выручки в зависимости от рыночных условий.

Цель работы – разработка методологии моделирования производительности труда на основе мягких вычислений, ориентированных на выявление количественных зависимостей между производственными факторами.

II. ОБЗОР

В работе Дранко О.И., Резчикова А.Ф., Степановской И.А., Богомолова А.С., Кушникова В.А. [1] исследуется многоуровневая модель индикативного планирования, увязывающая ВВП, ВДС и производительность труда с ресурсными ограничениями. Подход использован для обоснования модели роста ВВП и ранжирования мероприятий.

В исследовании Цвиркуна А.Д., Дранко О.И., Резчикова А.Ф., Богомолова А.С., Кушникова В.А., Степановской И.А. [2] описывается формализация задачи повышения производительности труда с введением коэффициента эффективности. Результаты применены для ранжирования мероприятий по эффективности использования кадровых ресурсов.

В статье Авдеевой З.К. и Ковриги С.В. [3] рассказывается о когнитивных картах как инструменте стратегического планирования. Аппарат использован для формализации качественных факторов и учёта неопределённости в рамках мягких вычислений.

В работе Авдеевой З.К. и Ковриги С.В. [4] рассматриваются методы интеграционного прогнозирования нестационарных временных рядов. Обосновано применение нейросетевых и гибридных моделей для обработки больших данных.

В исследовании Ковриги С.В. и Авдеевой З.К. [5] представлена модель убеждений и аттитудов субъекта

на основе когнитивных карт. Подход позволяет формализовать экспертные знания для нечёткого моделирования.

В материалах Дранко О.И., Резчикова А.Ф., Степановской И.А. [6] описывается подход к распределению мероприятий по эффективности использования ресурсов. Схема корреспондирует с задачей о дробном рюкзаке и ранжированием по производительности труда.

В работе Чернова И.В. [7] исследуется методология сценарного планирования. Разработанная процедура применена для построения прогнозных сценариев развития отраслей с учётом ограничений по производительности труда.

Классическая работа Заде Л.А. [8] закладывает теоретическую основу для нечёткого моделирования качественных факторов через введение понятия лингвистической переменной.

В издании Борисова В.В., Круглова В.В., Федулова А.С. [9] излагаются методы нечёткого логического вывода и нейро-нечётких сетей, использованные для построения гибридной модели производительности труда.

III. ИСХОДНЫЕ ДАННЫЕ

В качестве исходных данных использовались данные Росстата [10] и открытые данные ФНС России [11]

- по финансовой отчетности организаций (2,2 млн. организаций в 2024 г.);
- по численности организаций (993 тысячи организаций).

На основе открытых данных (далее выборка) по численности и выручке организаций была рассчитана производительности труда как отношение выручки к среднесписочной численности работников. Итоговый массив после объединения источников составил порядка 993 тысяч организаций, представляющих себе все секторы экономики. Учитывая масштаб и разнородность данных, особое внимание было уделено процедуре очистки данных. Для исключения аномальных значений, обусловленных в основном ошибками в отчетности, применялся отраслевой пороговый метод. Порог для выбросов использован $K = 10$ к среднеотраслевому значению. Это сохранило репрезентативность выборки.

Обработка таких больших объемов данных потребовала применения технологий распределенных вычислений и оптимизированных алгоритмов, что соответствует концепции работы с большими данными. После очистки данные были стратифицированы по отраслевой принадлежности по кодам ОКВЭД-2, и по категориям организации (по размеру средней выручки), что позволило строить модели с учетом структурной неоднородности и повысить точность прогнозирования.

IV. МОДЕЛЬ

Формализованная постановка задачи – это увеличение ВВП

$$GDP \rightarrow \max, \quad (1)$$

где GDP – валовый внутренний продукт.

ВВП состоит из суммы валовой добавленной стоимости (ВДС) по видам экономической деятельности

$$GDP = \sum_{i=1}^n GVA_i + \varepsilon, \quad (1)$$

где GVA_i – валовая добавочная стоимость отрасли i , а ε – прочие компоненты (чистые налоги, амортизация и фонд оплаты труда), которыми в данном исследовании не исследуются.

Максимизация ВВП сводится к увеличению ВДС в отраслях.

Валовая добавленная стоимость в отраслях принимается как некая функция от двух основных факторов производства: трудовые ресурсы и инвестиции. В общем виде можно представить как:

$$GVA_i = F_i(L_i, I_i, \theta_i), \quad (2)$$

где L_i – трудовые ресурсы (численность занятых, качество рабочей силы), I_i – инвестиции (объем основных фондов, капиталовложения), θ_i – совокупность структурных, технологических и институциональных факторов, отражающих специфику отрасли.

Производительность труда, определяемая как отношение добавленной стоимости к численности занятости – это ключевое звено, которое связывает ресурсы и ВДС. Однако данные о добавленной стоимости нам не доступны из отчетности компаний. Традиционно используется показатель отношения выручки к численности. В дальнейшем показатели можно дезагрегировать до предприятий по формуле (3)

$$GVA_i = \sum_{j=1}^n VA_{ij}, \quad (3)$$

где VA – добавочная стоимость j предприятия i -й отрасли.

Производительность труда будет рассчитываться по формуле

$$P_i = R_i/L_i, \quad (4)$$

где P – производительность труда, R – выручка отрасли, L – количество людей, занятых в отрасли, i – отрасль.

При ограниченных трудовых ресурсах L

$$\sum_i L_i = L, \quad (5)$$

наилучший суммарный результат достигается приоритизацией (ранжированием) отраслей по убыванию (задача о дробном рюкзаке)

$$P_i = R_i/L_i \rightarrow \max. \quad (6)$$

V. РЕЗУЛЬТАТЫ РАСЧЕТОВ

В результате вычислений была получена табл. I, которая представляет собой агрегированные показатели по видам экономической деятельности по классификации ОКВЭД-2. Для каждого сектора указаны: количество организаций, суммарная выручка отрасли (в тысячах рублей), общая численность работников (человек) и рассчитанная производительность труда – отношение выручки к численности (тыс.руб./чел.).

Масштаб выборки – более 990 тыс. организаций, что обеспечивает высокую репрезентативность для анализа производительности труда на уровне всей экономики. Распределение организаций по отраслям существенно различается: наибольшее количество сосредоточено в оптовой и розничной торговле (раздел G) – 300,8 тыс. (30,3% от общего числа), строительстве (раздел F) – 121,0 тыс., обрабатывающих производствах (раздел C).

Средневзвешенная производительность труда по выборке составила 12,4 млн. руб./чел. Значительно выделяется финансовая деятельность (Раздел K), что не удивительно для специфического вида деятельности – в 15 раз выше среднего. Торговля (оптовая и розничная) значительно выше среднего уровня. Отметим, что к торговле относятся крупнейшие организации по торговле газом (ПАО «Газпром»).

Значительный разброс показателей указывает на неоднородность отраслевой структуры и необходимость учёта отраслевой специфики при формировании программ повышения производительности труда.

Производительность труда значительно различается по видам деятельности (рис. 1, табл. I). Широкий разброс значений наглядно демонстрирует существенную неоднородность отраслевой структуры экономики. Для обеспечения наглядности ось абсцисс (значения выручки) построена в логарифмическом масштабе, что позволяет корректно отобразить многопорядковый диапазон показателей.

Полученные оценки согласуются с агрегированными данными Росстата по валовой добавленной стоимости и численности занятых, а их отклонения находятся в пределах погрешности, связанной с использованием выручки в качестве основного результативного показателя. Представленная дифференциация

подтверждает необходимость отраслевой стратификации при построении моделей производительности, поскольку единые для всей экономики зависимости не способны уловить специфику каждого сектора.

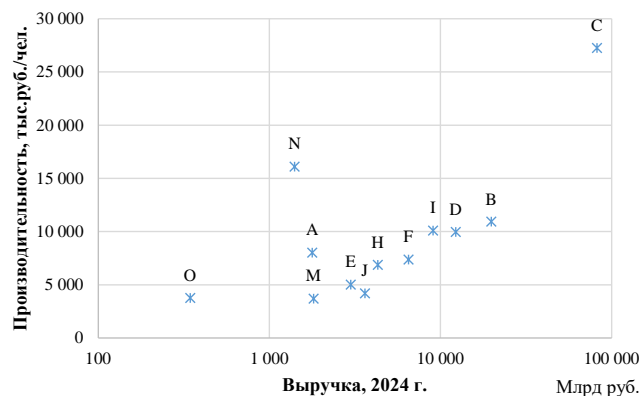


Рис. 1. Производительность труда по видам деятельности.

Зависимость производительности труда от размера организации по полной выборке показана на Рис. 2. Второй график иллюстрирует связь между размером организации (в качестве прокси-переменной использована выручка) и производительностью труда.

На всём массиве данных выявлена устойчивая закономерность, которая с высокой точностью аппроксимируется степенной функцией (коэффициент детерминации $R^2 = 0,9226$):

$$P = A R^{0,5039} + \varepsilon, \quad (7)$$

где P – средневзвешенная производительность труда, R – выручка организации, A – нормировочный коэффициент, ε – погрешность.

ТАБЛИЦА I. Производительность труда по отраслям, по выборке организаций, 2024 г.

Вид деятельности	Кол-во организаций	Выручка, млн. руб.	Численность, чел.	Производительность
А. Сельское, лесное хозяйство, охота, рыболовство и рыбоводство	21 328	5 920 694	910 248	6 504
В. Добыча полезных ископаемых	4 779	3 837 224	347 394	11 046
С. Обрабатывающие производства	112 738	32 686 086	3 480 895	9 390
Д. Обеспечение электрической энергией, ...	6 153	3 109 031	446 737	6 959
Е. Водоснабжение; ...	7 074	1 780 322	221 962	8 021
Ф. Строительство	121 014	19 803 838	1 812 545	10 926
Г. Торговля оптовая и розничная; ...	300 846	82 081 574	3 011 829	27 253
Н. Транспортировка и хранение	60 898	12 264 365	1 231 711	9 957
И. Деятельность гостиниц и предприятий общественного питания	30 657	2 986 460	594 476	5 024
Ж. Деятельность в области информации и связи	42 767	6 496 520	882 226	7 364
К. Деятельность финансовая и страховая	5 026	8 405 287	44 986	186 842
Л. Деятельность по операциям с недвижимым имуществом	83 837	4 296 971	625 837	6 866
М. Деятельность профессиональная, научная и техническая	90 188	9 040 013	895 261	10 098
Н. Деятельность административная ...	46 340	3 624 462	862 135	4 204
О. Государственное управление ...	826	48 217	19 660	2 453
Р. Образование	3 880	83 649	20 858	4 010
Q. Деятельность в области здравоохранения и социальных услуг	32 014	1 814 442	490 262	3 701
Р. Деятельность в области культуры, спорта, организации досуга и развлечений	8 507	1 403 235	87 118	16 107
С. Предоставление прочих видов услуг	14 392	344 597	91 521	3 765
Т. Деятельность домашних хозяйств ...	38	557	286	1 947
У. Деятельность экстерриториальных организаций и органов	2	101	8	12 565
Общий итог	993 304	200 027 643	16 077 955	12 441

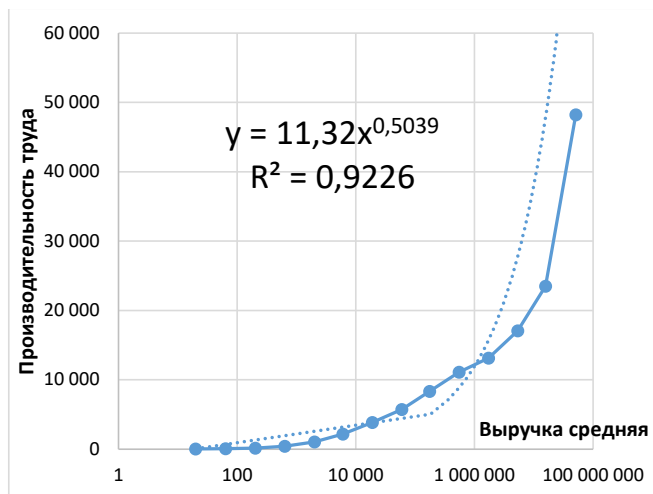


Рис. 2. Зависимость производительности труда от выручки

VI. ЗАКЛЮЧЕНИЕ

В работе предложена модель роста ВВП, учитывающая ресурсные ограничения по инвестициям и трудовым ресурсам, которая формализована как задача о дробном рюкзаке. Оптимальное распределение мероприятий достигается ранжированием по эффективности использования дефицитного ресурса, где при лимитировании трудовых ресурсов ключевым критерием выступает производительность труда. Для эмпирической верификации модели обработан массив данных, включающий 993 тыс. организаций всех отраслей, сформированный на основе открытой финансовой отчетности ФНС России и сведений о численности персонала. Использование методов обработки больших данных (big data) позволило провести агрегацию, фильтрацию выбросов и стратификацию по видам деятельности и масштабу предприятий.

В ходе анализа выявлена статистически значимая степенная зависимость производительности труда от размера организации, что подтверждает наличие эффекта масштаба и служит основой для количественной оценки эффективности использования трудовых ресурсов.

Применение методов мягких вычислений (soft computing), в частности нечётких множеств (fuzzy sets), позволило учесть неопределённость исходных данных, характерную для больших массивов разнородной информации. Гибридные алгоритмы, сочетающие

эволюционную оптимизацию и нечёткую логику, были использованы для идентификации параметров степенной модели.

Предложенный подход, базирующийся на интеграции больших данных и мягких вычислений, позволяет обосновать приоритеты развития отраслей и регионов на основе объективных показателей производительности труда, а также корректировать программы развития с учётом ограничений по кадровым ресурсам.

СПИСОК ЛИТЕРАТУРЫ

- [1] Дранко О.И., Резчиков А.Ф., Степановская И.А., Богомолов А.С., Кушников В.А. Сценарное моделирование развития страны на основе индикативного планирования // Проблемы управления. 2024. № 5. С. 25–41.
- [2] Цвиркун А.Д., Дранко О.И., Резчиков А.Ф., Богомолов А.С., Кушников В.А., Степановская И.А. Задача повышения производительности труда на производстве // Управление большими системами: сборник трудов. 2025. Вып. 116. С. 6–30.
- [3] Авдеева З.К., Коврига С.В. Когнитивные карты для анализа и моделирования в системе стратегического планирования государства // Управление большими системами: сборник трудов. 2024. Вып. 111. С. 147–178.
- [4] Авдеева З.К., Коврига С.В. Интеграционное прогнозирование нестационарных процессов, представленных временными рядами. Обзор // Управление большими системами: сборник трудов. 2024. Вып. 112. С. 129–167.
- [5] Коврига С.В., Авдеева З.К. Модель представления убеждений и аттитудов субъекта с использованием когнитивных карт // Управление большими системами: сборник трудов. 2025. Вып. 117. С. 19–51.
- [6] Дранко О.И., Резчиков А.Ф., Степановская И.А. Подход к распределению мероприятий по развитию на основе эффективности использования ресурсов // Труды 17-й Международной конференции «Управление развитием крупномасштабных систем» (MLSD*2024). М.: ИПУ РАН, 2024. (в печати).
- [7] Чернов И.В. Методология сценарного планирования и управления в системе обеспечения национальной безопасности // Управление большими системами: сборник трудов. 2024. Вып. 112. С. 187–232.
- [8] Заде Л.А. Понятие лингвистической переменной и его применение к принятию приближенных решений. М.: Мир, 1976. 165 с.
- [9] Борисов В.В., Круглов В.В., Федулов А.С. Нечёткие модели и сети. 2-е изд. М.: Горячая линия – Телеком, 2012. 284 с.
- [10] Федеральная служба государственной статистики (Росстат). Российский статистический ежегодник. 2024: Стат. сб. М.: Росстат, 2024. URL: <https://rosstat.gov.ru/folder/210/document/12994> (дата обращения: 30.03.2026).
- [11] Единая межведомственная информационно-статистическая система (ЕМИСС). Показатели производительности труда. URL: <https://www.fedstat.ru> (дата обращения: 10.03.2026).