

---

---

# ТЕОРИЯ И МЕТОДЫ РЕГИОНАЛЬНЫХ ИССЛЕДОВАНИЙ

---

---

УДК 332.1

Земцов С.П. (Москва)

## ОБЗОР СТАТИСТИЧЕСКИХ МЕТОДОВ РЕГИОНАЛЬНОГО АНАЛИЗА ИННОВАЦИОННОЙ ДЕЯТЕЛЬНОСТИ

Zemtsov S.P.

### OVERVIEW OF STATISTICAL METHODS FOR REGIONAL ANALYSIS OF INNOVATIVE ACTIVITIES

**Аннотация.** В статье дан анализ основных статистических методов региональных исследований инновационной деятельности. Особое внимание уделяется модели производственной функции знаний, являющейся разновидностью метода множественной регрессии, служащей для моделирования производства инновационной продукции, в первую очередь, на региональном уровне. Приведены обзоры наиболее значимых эмпирических зарубежных и российских публикаций с описанием методик и процедур исследований. Кроме того, рассматриваются новые методы регионального анализа: бинарные регрессии, многоуровневые модели, анализ стохастической границы эффективности и оболочечный анализ данных.

**Abstract.** The article analyzes the main statistical methods for regional research of innovation. Particular attention is given to the model of the production function of knowledge, which is a variant of the method of multiple regressions, serving to simulate the production of innovative products, especially at the regional level. There is an overview of the most important empirical Russian and foreign publications with description of its methodology. It also covers new techniques for regional studies: binary regressions, multi-level analysis, stochastic frontier analysis and data envelopment analysis.

**Ключевые слова:** статистические методы, региональный анализ, регионы России, производственная функция знаний, анализ стохастической границы, оболочечный анализ данных.

**Keywords:** statistical methods, regional analysis, regions of Russia, the knowledge production function,

---

**Введение.** Анализ инновационных процессов как предмет региональных исследований за рубежом ассоциируется с географией инноваций – крупным и плодотворным направлением регионалистики, заложенным в начале 90-х гг. XX в. Мэрианн Фельдман [21, 35, 37]. За последние десятилетия вышли сотни работ, посвященные данной тематике [37], а на крупнейших форумах регионалистов<sup>1</sup> обязательно присутствует соответствующая тематическая секция [21]. В отечественной региональной науке основные работы опубликованы В.Л. Бабуриным [3, 4, 5, 12], А.Н. Пилясовым [21] и представителями новосибирской школы [13, 20, 24].

Изначально сама необходимость регионального анализа инновационных процессов ставилась под сомнение, и до сих пор ведется дискуссия, какой территориальный уровень исследований предпочтителен (см., например, [21, 28]). Необходимость применения методов регионального анализа обосновывается в рамках концепций перетока и неявных знаний. Особенность знаний в их неделимости, возможности использовать неограниченное число раз и ограниченной возможности исключить других агентов от пользования ими, поэтому инновационная деятельность одного агента порождает положительные внешние эффекты для других,

---

<sup>1</sup> Конференции Международной ассоциации региональных исследований (<http://www.regionalstudies.org/conferences/conference/regional-studies-association-annual-conference-2015-piacenza-italy>) и Конгрессы Международной и Европейской ассоциации региональной науки (<http://www.ersalisbon2015.org/#!about3/cdv2>).

так называемые знаниевые экстерналии, или перетоки знания (от англ. knowledge spillover [23])<sup>2</sup>. Часть знаний, неявные знания, не могут быть полностью формализованы, а передаются только «от учителя к ученику». В обоих случаях локализация знаний и их генерация происходят на локальном и региональном уровнях.

Основные проблемы географии инноваций, решение которых служит выработке рекомендаций по совершенствованию региональной инновационной политики:

- определение условий и факторов, влияющих на инновационную деятельность в регионах, для их последующего стимулирования;
- типология регионов по их инновационным характеристикам, в частности выявление регионов-лидеров, для определения территориальных приоритетов государственной поддержки или размещения научно-исследовательских центров крупных компаний;
- изучение эффективности региональных инновационных систем (далее – РИС) для определения территориальных приоритетов и выявления эффективных инструментов региональной политики.

Изучение инновационных процессов сопряжено с рядом трудностей, так как объект исследования характеризуется сложной и многофакторной структурой с множеством нечетких и скрытых связей. Статистические методы<sup>3</sup> регионального анализа играют решающую роль, так как позволяют решать многофакторную задачу определения переменных, влияющих на инновационную активность. Но при эмпирической проверке гипотез не все факторы могут быть корректно интерпретированы с помощью существующих показателей [7, 41]. Ученые часто сталкиваются с отсутствием или ограниченным доступом к качественной и верифицируемой информации, характеризующей инновационные процессы, из-за сложности сбора данных, их высокой степени секрет-

ности, высокой скорости изменчивости и других причин. Зачастую эти ограничения вынуждают исследователей использовать косвенные индикаторы, экспертные оценки, интегральные индексы и т.д. Область знаний об инновационных процессах в рамках региональных исследований активно развивается, заимствуя методы и подходы из смежных дисциплин. В частности активно используются картографические методы (картограммы, картодиаграммы, карты-анаморфозы, линии движения и т.д.) для описания структуры инновационного пространства регионов, сетей взаимодействия между ними, процессов диффузии инноваций и т.д. (подробнее см., например, [4, 12]).

Цель данной статьи – формирование универсального свода методов, применимых для регионального анализа инновационной деятельности<sup>4</sup>. Структура работы выстроена в соответствии с основными проблемами географии инноваций, описанными выше.

Исследование инновационных процессов значимо для современной социально-экономической географии, так как позволяет выявить новые факторы территориальной организации общества и хозяйства. В современной отечественной экономико-географической литературе данному научному направлению уделяется недостаточное внимание.

**Методы анализа факторов инновационной деятельности в регионах.** При выявлении зависимостей между инновационной активностью и факторами ее определяющими наиболее применяемым является *метод множественной регрессии*<sup>5</sup> [19, 22, 33], позволяющий определить влияние отдельных независимых переменных–регрессоров на исследуемую зависимую величину. Условием применения данного метода является однородность выборки, что нарушается при изучении инновационной деятельности в России, которая высоко сконцентрирована в столичном регионе [4, 12]. Решение об изменении выборки (исключении выбо-

<sup>2</sup> По данным [44] объем перетоков знаний в США, индикатором которых выступают патентные цитаты, существенно уменьшается при достижении расстояния в 100 миль (около 160 км).

<sup>3</sup> Статистические методы широко распространены в социальных науках, довольно часто используются в экономической географии [2, 10, 23], очень широкое распространение получили в региональных исследованиях как за рубежом, так и в России [16, 20, 24, 38, 44]. Их применение существенно упрощено использованием специальных программных сред (MatLab, R+ и др.) и многочисленных статистических пакетов (SPSS, Statistica, EViews, GRETL, STADIA, Stata и др.) [6, 18, 22].

<sup>4</sup> Для наглядности изложения материала, где это было возможным, использованы примеры исследований российских регионов с подробным описанием процедур.

<sup>5</sup> Для получения наилучших результатов требуется выполнение четырех условий Гаусса-Маркова [11].

сов) принимается исходя из целей работы, но предпочтительным является нормальное распределение данных<sup>6</sup>. Рекомендуется также, чтобы число независимых переменных было в шесть-семь раз меньше, чем число исследуемых объектов, то есть для региональных исследований в России возможно использование в регрессии не более 12-13 независимых переменных [11]. Дискуссионным является вопрос о достаточности данных по 85 субъектам федерации для выявления зависимостей методом, ориентированным на сотни и тысячи наблюдений, поэтому во многих исследованиях предпочтение отдается панельным данным, которые увеличивают число наблюдений, но могут исказить результат иным образом [19].

Наиболее применяемыми в региональном анализе инновационной деятельности являются модели *производственной функции знаний* (ПФЗ) [41, 42], в которой зависимой переменной выступают результаты инновационной деятельности, а независимыми – затраты ресурсов. Модель ПФЗ была предложена Цви Грилихесом в 80-е гг. XX в. [41] и изначально служила для изучения влияния затрат на научные исследования и опытно-конструкторские разработки (RnD\_exp – НИОКР) на инновационную (на практике – патентную) активность (Innov)<sup>7</sup>:

$$\ln Innov_i = \beta_0 + \beta_1 \times \ln RnD\_exp_i + \varepsilon \quad (1),$$

где  $i$  – регион,  $\beta_j$  – эмпирические коэффициенты;  $\varepsilon$  – остаток.

Важным моментом при построении подобного рода моделей является выбор зависимой переменной, отражающей объем инновационного «выпуска». В наибольшем числе исследований используются патенты и иные объекты интеллектуальной собственности как прокси для видимой и оформленной части изобретений [3, 25, 41, 42, 48, 51]. Но не все патенты могут быть коммерциализованы и воплощены в инновационной продукции [41, 42, 44, 47].

Информация о патентах, как и в целом статистика по инновационной деятельности в России, не всегда может считаться надежной [5, 7, 12]. Для российских регионов в среднем характерна низкая доля коммерциализированных патентов, в 2000-х годах не превосходившая 7%<sup>8</sup>. Невысокий уровень коммерциализуемости российских патентов может говорить о невысоком качестве разрабатываемых технологий и о низком качестве проверки патентов. Дополнительным свидетельством их низкого качества может служить тот факт, что для отдельных регионов характерна высокая волатильность числа патентных заявок во времени и завышенное число патентов на человека. Примером является Ивановская область, где число поданных патентных заявок возросло на порядок в течение двух лет. В ряде субъектов, наоборот, патентная активность крайне мала и, возможно, носит случайный характер, например в Ненецком автономном округе.

В монографии [13] зависимой переменной выступает число инновационно-активных предприятий ( $N$ ) и применяется ПФЗ, предложенная П. Ромером [50]:

$$\dot{A}_i = \delta H_i^\mu A_i^s \quad (2),$$

где  $\dot{A}$  – прирост новых знаний в регионе  $i$ ,  $H$  – человеческий капитал<sup>9</sup>,  $A$  – накопленный уровень знаний,  $\delta$ ,  $\mu$  и  $s$  – параметры модели.

Оценка величины накопленных знаний была исключена в виду отсутствия достоверных данных по регионам, а также предположения о том, что запас накопленных за советский период знаний исчерпан. Величина человеческого капитала сведена к численности занятых в сфере НИОКР (HR). Уравнение представлено в логлинейной форме

$$\ln(\Delta N)_{i,t} = a_0 + a_1 \ln(H_R)_{i,t} + \eta_i + \mu_i + \xi_{i,t} \quad (3),$$

где  $a_i$  – эмпирические коэффициенты,  $\mu_i$ ,  $\xi_{i,t}$ ,  $\eta_t$  – идеи, распространяющиеся из других

<sup>6</sup> Наличие выбросов отражает естественную структуру данных, обусловленную процессами экономико-географической концентрации. Впрочем, критерий однородности выборки подразумевает единую природу и единый масштаб исследуемых объектов, что очевидно не выполняется при сравнении среднестатистических регионов России с крупнейшими агломерациями страны и со слабозаселенными территориями.

<sup>7</sup> В работах М. Фельдман [21, 35] показано, что основные предпосылки ПФЗ выполняются только на региональном и локальном уровнях, а на уровне фирм и стран зависимости между инновационной деятельностью и затратами на НИОКР могут не выявляться.

<sup>8</sup> Патентная активность России и США: аналитическое исследование из цикла «Индикаторы инновационного развития российской экономики», 2013 г. Режим доступа: [http://www.nbkg.ru/researches/patent\\_activity\\_russia\\_vs\\_usa.pdf](http://www.nbkg.ru/researches/patent_activity_russia_vs_usa.pdf)

<sup>9</sup> Чаще всего репрезентативными индикаторами человеческого капитала служат доля занятых с высшим образованием, среднее число лет обучения или занятые в НИОКР.

регионов, состоящие из фиксированного во времени индивидуального эффекта, фиксированного для всех регионов временного эффекта и стохастической компоненты.

Авторы высказали несколько гипотез. Первая гипотеза о существовании положительных внешних эффектов от межрегионального взаимодействия не подтвердилась; их отсутствие свидетельствует о низкой степени горизонтального взаимодействия между регионами. Вторая гипотеза о положительном влиянии человеческого капитала на создание инновационных идей подтвердилась, но только для кандидатов наук и аспирантов; влияние докторов наук и технического персонала на прирост числа инновационных компаний оказалось отрицательным.

К недостаткам исследования следует отнести использование независимой переменной, которая является статистически недостоверной [7]<sup>10</sup>, а также использование численности занятых в НИОКР как переменной, оценивающей человеческий капитал, хотя многие исследователи не занимаются созданием инноваций, да и сами инновации могут создаваться за пределами сектора НИОКР.

В работе [24] в качестве зависимой переменной выступает число поданных заявок на изобретения и полезные модели. Выявление факторов инновационной активности осуществляется с помощью панельной регрессии – модели с фиксированными ошибками. Использована выборка из 76 субъектов РФ в период с 1998 по 2003 г. (349 наблюдений), исключен ряд регионов в связи с отсутствием отдельных показателей. Было выявлено, что затраты на НИОКР в регионах России оказывают положительное влияние на инновационную активность. Значимым оказалось качество человеческого капитала, выраженное через число занятых в НИОКР. Показано, что импорт технологий преобладает над экспортом, но сделан неправомерный вывод о том, что экспорт товаров может являться негативным фактором создания патентов. В реальности региональные инновационные системы (РИС) слабо развиты в северных регионах, где ведется добыча полезных ископаемых на экспорт. Выводы исследова-

ния спорны, в том числе из-за использования показателя «число заявок на патенты» без дополнительной верификации. Возможна эндогенность ряда переменных, в частности числа исследователей и числа занятых в НИОКР, поэтому использование их в одной модели не оправдано. Критика подхода также может быть связана с недоучетом временных лагов, очевидно возникающих между этапом финансирования НИОКР и регистрацией патентов.

В статье [17] в качестве зависимой переменной используется объем выпуска инновационной продукции; исследуется влияние эффектов диффузии инноваций и качества человеческого капитала. Для построения регрессии применен обобщенный метод моментов. В разных спецификациях положительно значимы оказались: выпуск инновационной продукции в предшествующий период, что свидетельствует о дивергенции регионов; число патентов, что может показывать влияние сектора НИОКР на создание инновационной продукции; объем прямых иностранных инвестиций, показывающий влияние глобальных потоков знаний (от англ. *global knowledge pipeline* [21; 37]); инвестиции в основной капитал, что доказывает необходимость обновления оборудования. Сомнение вызывает определение зависимой переменной [7].

В работе [5] авторами использована следующая спецификация ПФЗ:

$$\begin{aligned} \ln(Pat\_act_i) = & \beta_0 + \beta_1 \times \ln(Human\_Cap_i) + \\ & + \beta_2 \times \ln(Pat\_stock_i) + \beta_3 \times \ln(RnD\_exp_i) + \\ & + \beta_4 \times \ln(RnD\_empl_i) + \beta_5 \times \ln(Pat\_potential) \\ & + \beta_6 \times \ln(Urban) + \beta_7 \times \ln(Unr\_variety) + \beta_8 \times \\ & \times (Ind\_conc) + \beta_9 \times \ln(Pop\_dens) + \varepsilon \end{aligned}$$

где  $i$  – регион;  $Pat\_act$  – число заявок на патенты на 10 млн чел.;  $Human\_cap$  – человеческий капитал;  $Pat\_Stock$  – накопленные знания: число использованных патентов кумулятивно с 1994 г.;  $RnD\_exp$  – затраты на НИОКР;  $Pat\_potential$  – оценка перетока знаний (формула 5);  $Urban$  – доля городских жителей,  $Unr\_Variety$  – агломерацион-

<sup>10</sup> Компании не заинтересованы в заполнении сложной и трудоемкой статистической формы №4 Инновация, значительно проще либо саботировать ее заполнение, либо проставить нулевые значения во всех графах. Крупным компаниям выгоднее не заполнять форму в виду низких штрафных санкций, в частности форма не заполняется большинством иностранных компаний в России. В этой связи как объем инновационной продукции, так инновационная активность компаний в регионах сильно занижаются. При этом если для патентов можно определить регионы, в которых статистическая информация наиболее недостоверна, а соответственно исключить их из рассмотрения, то в случае с инновационной продукцией степень искажения в различных регионах неизвестна.

ный эффект разнообразия занятости (индекс энтропии Шеннона по отраслевой структуре занятости); *Ind\_conc* – локализационные эффекты концентрации промышленности (индекс Херфиндала-Хиршмана по отраслевой структуре промышленного выпуска).

Трудность оценки ПФЗ заключается в том, что регионы являются сложными открытыми системами, а на исследовательскую деятельность в отдельном регионе влияние могут оказывать научные связи с другими регионами. В работах [4, 5] в качестве индикатора возможных перетоков знания выступает *патентный потенциал*<sup>11</sup>, рассчитанный с помощью гравитационного уравнения, предложенного У. Айзардом [43]:

$$V_j = P_j + \sum P_i / D_{ji} \quad (5),$$

где  $V_j$  – потенциал в столице региона  $j$ ;  $P_j$  и  $P_i$  – число выданных патентов на изобретения на 100 тыс. городских жителей регионов  $j$  и  $i$ ;  $D_{ji}$  – расстояние от центра  $j$  до центра  $i$ , км.

На рис. 1 показана региональная структура патентного потенциала: возможные перетоки знаний преимущественно концентрируются в Европейской части России<sup>12</sup> в регионах, соседствующих с крупнейшими агломерациями и научными центрами.

В работе [5] выявлено, что на число российских патентных заявок на изобретения, скорректированного с помощью статистических методов<sup>13</sup>, наибольшее влияние оказывают человеческий капитал (доля занятых с высшим образованием), накопленное число использованных патентов, общие затраты на НИОКР и плотность населения, а на международную патентную активность (*Pat\_PCT*)<sup>14</sup> (табл. 1) – среднее число лет обучения занятых, затраты на прикладные НИОКР и потенциальный переток знаний (рис. 1).

Для обеих зависимых переменных показатель человеческого капитала оказался более значим, чем затраты на НИОКР, то есть наличие «умных» людей – более важный

фактор патентной активности в регионах России, чем наличие финансовых ресурсов. Результаты также могут свидетельствовать о неэффективности затрат на НИОКР.

Основной проблемой применения метода множественной регрессии является отсутствие качественных, доступных и верифицируемых данных о зависимой переменной. Для решения данной проблемы часто используются различные методы «уточнения» переменной, например, можно предложить показатель числа потенциально коммерциализируемых патентов, рассчитываемый по формуле

$$Innov_i = 0,05 \times Pat\_rus_i + 0,5 \times Pat\_PCT_i \quad (6),$$

где *Innov* – число коммерциализируемых патентов региона  $i$ , *Pat\_rus* – число российских патентных заявок, *Pat\_PCT* – число поданных РСТ-заявок. Коэффициенты в данном случае отражают уровень коммерциализируемости: для российских заявок в среднем он не превосходит 5%, для международных – около 50%. Но недостатком такого подхода является недоучет региональной дифференциации уровня коммерциализации патентов.

Когда неизвестен реальный объем инновационного выпуска или он вызывает сомнения, зависимая переменная может иметь или принимать бинарную структуру: 1 – инновация создана и 0 – отсутствие инновации. В этом случае применяются специализированные методы: логит-, пробит- или тобит-регрессии [6, 22, 33].

Большинство подобных моделей используется на уровне фирм с учетом региональных факторов [52, 53]. Например, в работе [53] на примере 1800 компаний Европейского Союза показано, что характеристики региональных инновационных систем (РИС) оказывают сильное воздействие на инновационное поведение (где 1 – внедрение новых технологий; 0 – их отсутствие) малых фирм,

<sup>11</sup> Для расчета потенциальных перетоков знаний также используются методы на основе подсчета суммы и среднего арифметического значения инновационной активности соседних регионов.

<sup>12</sup> Картограммы и картодиаграммы активно применяются для понимания факторов территориальной дифференциации отдельных переменных инновационной деятельности, так как объяснение взаимовлияния переменных зачастую следует искать не в их реальном взаимодействии, а в их взаимном расположении.

<sup>13</sup> Для получения правомерных результатов из выборки исключены регионы, в которых коэффициент вариации патентных заявок по годам был выше 0,3.

<sup>14</sup> Использование данных о РСТ-заявках (англ. Patent Cooperation Treaty – Договор о Патентной Кооперации) может дать более достоверную информацию об уровне и качестве изобретательской активности, но в этом случае результаты искажены низкой патентной активностью в большинстве регионов. Получить РСТ-патент сложнее, так как процесс проверки патента требует нескольких лет, а взносы составляют в совокупности около \$3000. РСТ-патенты обладают значительно большей степенью коммерциализируемости.





Рис. 1. Патентный потенциал регионов России

Таблица 1

Результаты расчета регрессии с зависимой переменной *Pat\_PCT*

Панельная регрессия с фиксированными эффектами. 984 наблюдения.				
Переменные	Коэффициенты (стандартная ошибка)			
Константа	-27,42*** (3,53)	-12,53** (5,81)	-12,94** (5,84)	-13,06** (5,78)
Среднее число лет обучения занятых	11,8*** (1,37)	5,26** (2,42)	5,57** (2,44)	5,37** (2,44)
Число использованных патентов кумулятивно с 1994 г.		0,27*** (0,07)	0,17* (0,1)	0,07 (0,1)
Затраты на прикладные НИОКР			0,08* (0,04)	0,08* (0,04)
Патентный потенциал, или переток знаний				0,55** (0,23)
$R^2$	0,69	0,7	0,7	0,71
Скорректированное на число переменных значение $R^2$	0,66	0,67	0,67	0,67

Значимость (*p-value*) на уровне: \*\*\* - 0,01; \*\* - 0,05; \* - 0,1

но для крупных фирм более значимым оказываются внутрифирменные факторы.

Логит-регрессия применяется также для предсказания вероятности возникновения некоторого события в зависимости от фактора  $x$ . Регрессионное уравнение имеет вид:

$$Pr\_Innov_i = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_i x_i)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_i x_i)} \quad (7),$$

где  $Pr\_Innov_i$  – вероятность создания инновации в регионе  $i$  ( $Pr\_Innov_i \in [0;1]$ ),  $b_0$  – свободный член,  $b_i$  – коэффициенты уравнения регрессии,  $\varepsilon$  – случайная величина.

В статье [12] приведён пример расчета зависимости между инновационным потенциалом регионов России, рассчитанным с помощью интегрального индекса  $[0; 1]$ , и вероятностью создания инновационной технологии в регионе, определяемой через бинарную переменную числа заявок на РСТ-патенты. Небинарные по своей природе данные о заявках в связи с нулевыми и малыми значениями в большинстве регионов России преобразованы в бинарный вид, где значение 1 придавалось регионам со средним числом заявок более 3-х в 2009–2011 гг.,

а 0 – в иных случаях<sup>15</sup> (рис. 2). Полученная функция служит для расчета вероятности возникновения новой технологии в регионах России в зависимости от их инновационного потенциала. Инновационный потенциал Москвы равен 1, поэтому и вероятность возникновения инновации близка к единице, а в Чукотском автономном округе с минимальным потенциалом вероятность близка к нулю.

В ряде случаев для проведения взвешенной инновационной политики требуется понять, на какой пространственный уровень необходимо воздействовать. Нужно ли поддерживать конкретные фирмы, кластеры или регионы? Для оценки влияния пространственного уровня используются многоуровневые модели (от англ. multilevel modeling) [40, 52, 53]. Предположим, модель имеет двухуровневую структуру с фирмами на первом уровне, расположенными в регионах на втором уровне. Тогда для первого уровня уравнение регрессии выглядит следующим образом:

$$I_{ij} = \beta_{0j} + \beta_{1j} \times x_j + e_j \quad (8)$$

А второй уровень моделируется следующими соотношениями:

$$\beta_{0j} = \alpha_0 + \alpha_0 \times \omega_j + u_{0j} \quad (9),$$

$$\beta_{1j} = \alpha_1 + \alpha_1 \times \omega_j + u_{1j} \quad (10),$$

где  $I_{ij}$  является зависимой переменной (в нашем случае инновационная активность),  $x_j$  – объясняющая переменная (предиктор) на первом уровне;  $\omega_j$  – предикторы второго уровня;  $e_j$  и  $u_{ij}$  – случайные эффекты;  $i$  – фирма;  $j$  – регион. В статье [52] показано, что факторы развития региональной инновационной системы заметно превышают внутрифирменные факторы, что является важнейшим выводом для обоснования изучения и поддержки инновационной деятельности на региональном уровне.

**Методы редукции данных, типологии и классификации регионов.** Для выявления приоритетов и проведения вариативной

инновационной политики необходима типология регионов и выделение лидеров и аутсайдеров. А при определении конкретных целей инновационной политики требуется сокращение числа рассматриваемых индикаторов инновационной деятельности. Для последнего применяются методы редукции данных: факторный анализ<sup>16</sup>, компонентный и кластерный, которые в региональных исследованиях одновременно используются и для типологии регионов.

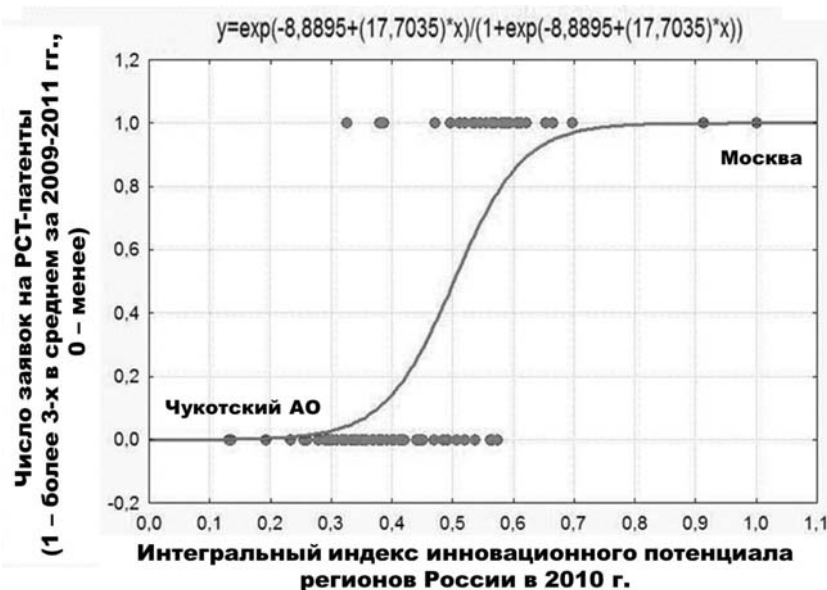
Цель компонентного анализа заключается в преобразовании набора независимых переменных  $\{x_1, x_2 \dots x_n\}$  в новый набор главных компонент  $\{f_1, f_2 \dots f_m\}$ , чтобы набор каждой следующей компоненты соответствовал все более точному объяснению суммарной дисперсии [11]. Каждая исходная переменная определенным образом коррелирует с новой компонентой, что определяется через показатель нагрузки. Первая компонента соответствует самой большой доле дисперсии. Преимуществом применения анализа на региональном уровне является возможность выявления связанных друг с другом показателей и их сочетаний в одних и тех же регионах. Таким образом, регионы-лидеры будут обладать максимально благоприятным сочетанием факторов.

В докладе [32] применялся метод главных компонент для выявления факторов инновационного потенциала на примере 115 стран по 25 показателям в период с 1992 по 2004 гг. В результате выявлено, что первая (главная) компонента инновационного потенциала коррелирует с индикаторами патентования, научными публикациями, инфраструктурой информационно-коммуникационных технологий, центрами сертификации ISO и доступом к финансированию. Можно утверждать, что указанные показатели необходимо использовать в качестве целевых в инновационной политике.

В упомянутой статье [12] изложена методика типологии регионов и верификации патентной активности с помощью метода главных компонент. На основе анализа более 30-ти показателей по регионам России в 2011 г. была определена первая компо-

<sup>15</sup> Расчет бинарной переменной основан на допущении, что если в регионе ежегодно создается несколько международных патентов, то хотя бы один будет коммерциализирован и станет инновационной продукцией, а значит вероятность создания инновации в таком регионе близка к единице (значение 1 на оси ординат).

<sup>16</sup> Для применения метода требуется однородность выборки и соответствие распределения данных нормальному закону, в противном случае необходимо преобразование данных или поиск иных методов анализа [1]. Предварительно, на основе анализа матриц парных корреляций один из двух показателей с корреляцией более 0,9 следует исключить из дальнейшего исследования.



Примечание: на графике пунсонами обозначены конкретные значения регионов, а линией – регрессионная функция, уравнение которой приведено в верхней части рисунка

Рис. 2. График зависимости вероятности возникновения новой технологии в регионах России от их инновационного потенциала

нента, объяснявшая более 50% дисперсии и включавшая патентную активность. При этом число выданных патентов имело максимальную нагрузку, то есть максимально коррелировало с компонентой один, поэтому можно говорить, что остальные показатели «уточняли», или верифицировали патентную активность. Индикаторы с нагрузкой выше 50% в первой компоненте использованы авторами для типологии регионов России путем построения интегрального индекса инновационного потенциала<sup>17</sup>. Впоследствии на основе индекса с помощью метода изолиний были выявлены ареалы концентрации инновационного потенциала (Окско-Волжское междуречье, Балтийский, Волжский и др.), где поддержка инновационной деятельности и создание инновационных центров наиболее эффективны.

Среди методов классификации регионов выделяются: кластерный, дискриминантный и анализ таблиц сопряженности. Каждый регион обладает набором признаков, выраженных в  $n$  переменных, то есть регионы

можно представить в виде точек в  $n$ -мерном пространстве и измерить расстояние между ними. Близко расположенные точки включаются в одну группу (кластер) [1, 15]. Анализ, проводимый с помощью статистических пакетов<sup>18</sup>, имеет серьезный недостаток – невозможность точной интерпретации выявленных групп регионов. Кроме того, сложно формализовать и обосновать выбор метода кластеризации и способа оценки расстояния.

В работе [8] критерием выбора метода кластеризации является предположение, что чем более равномерно регионы расположены между кластерами, тем лучше. Для определения равномерности используется информационный критерий энтропии Шеннона. Оптимальны для регионов России признан метод Варда с использованием квадрата евклидова расстояния<sup>19</sup>. Но в случае исследования инновационных процессов критике можно подвергнуть саму предпосылку о необходимости равномерного распределения регионов между кластерами, так как очевидно, что Москва и Санкт-Петербург резко

<sup>17</sup> Авторы также использовали для типологии регионов конкретные значения первой и второй компоненты, рассчитанные по результатам компонентного анализа на основе нагрузок всех использованных индикаторов.

<sup>18</sup> SPSS, STATISTICA, Stata.

<sup>19</sup> Исследования на примере европейских регионов показали, что в целом для типологии регионов этот метод и является оптимальным [54].



выделяются из всех регионов и не могут попасть в один кластер с другими.

В публикациях [3, 4] используется особый вид кластерного анализа с заранее выявленными пороговыми значениями (основаны на средних значениях) для определения границы между регионами-инноваторами и регионами-акцепторами новых технологий. Показано, что число креативных регионов в России в сравнении с 1988 г. сокращается, а растет число регионов, заимствующих зарубежные технологии, а также относящихся к инновационной периферии, где инновации не создаются и не заимствуются<sup>20</sup>.

**Методы анализа эффективности региональных инновационных систем.** Для оценки эффективности инновационной политики необходим особый комплекс методов.

Первый из используемых подходов основан на построении индекса путем соотнесения результата и используемых ресурсов: чем выше соотношение инновационной продукции к расходам на инновационную деятельность, например, патентов к затратам на НИОКР, тем более эффективна региональная инновационная система (РИС) [5].

*Оболочечный анализ*, или анализ среды функционирования (англ. Data Envelopment Analysis, DEA) – метод сравнительного анализа эффективности сложных систем [29], соотносящий результаты к задействованным ресурсам. В работе [34] предложен непараметрический подход к измерению эффективности в случае одного выпуска методом дробно-линейного программирования при постоянной отдаче от масштаба<sup>21, 22</sup>. Регион считается эффективнее другого, если он достигает по крайней мере не меньшего выпуска, чем другой,

при использовании не большего количества ресурсов.

В пространстве «затраты – выпуск» ( $y$  и  $x$ ) можно построить график границы производственных возможностей (CRS; максимальный выпуск при заданных ресурсах) [29]. На рис. 3 показано соотношение двух факторов затрат и одного индикатора выпуска. Регионы, находящиеся на границе CRS, наиболее эффективны (2 и 5), оценка их эффективности ( $\theta$ ) = 1. Для других регионов  $\theta \in (0; 1)$ . Оценить эффективность регионов 3 и 4 можно, проведя линию к началу координат, тогда соотношение отрезков  $O3^*$  к  $O3$  и  $O4^*$  к  $O4$  и будет индексом эффективности.

Методом преимущественно оценивается эффективность национальных и региональных инновационных систем [30, 46] по созданию новых технологий<sup>23</sup>.

Третий метод оценки технической эффективности регионов и их инновационных систем основан на *анализе стохастической границы* производственных возможностей (SFA – Stochastic Frontier Analysis) [26, 45]. Регион технически эффективен, если невозможно производить больше инноваций при заданном количестве ресурсов. Для определения функции границы в нашем случае может использоваться ПФЗ:

$$y_i = f(X_i\beta) + V_i - U_i \quad (13),$$

где  $y$  – объем выпуска (например, число патентов) в регионе  $i$ ,  $X$  – факторы производственной функции,  $\beta$  – коэффициенты,  $V$  – нормально распределенная случайная ошибка,  $U$  – неэффективность, имеющая одностороннее нормальное распределение.

В логарифмированном виде [16] ПФЗ примет вид:

$$\ln P_i = \beta_0 + \beta_1 \times \ln L_i + \beta_2 \times \ln K_i + \varepsilon \quad (14),$$

<sup>20</sup> Описанные выше статистические закономерности в пространстве анализируются также методами пространственной эконометрики, находящимися на стыке между статистическими и экономико-географическими методами. [36]. Они применяются в анализе расположения многочисленных дискретных объектов, например, университетов, инновационных компаний и т.д. К сожалению, в большинстве случаев модели на уровне регионов не обладают более высокой объясняющей способностью по сравнению со стандартными регрессиями.

<sup>21</sup> В статьях [27, 29] были предложены способы с учетом возможности нескольких выпусков и метод оценки относительной эффективности, не находящегося на границе производственных возможностей. В [27, 31] эти идеи были обобщены на случай переменной отдачи от масштаба.

<sup>22</sup> Среди программных продуктов DEA-анализа выделим MaxDEA (<http://www.maxdea.cn/>), имеющее бесплатную версию, и OpenSourceDEA ([http://www.opensourcidea.org/index.php?title=Open\\_Source\\_DEA](http://www.opensourcidea.org/index.php?title=Open_Source_DEA)) со сходными характеристиками, но также позволяющая построить границу эффективности.

<sup>23</sup> Метод разработан для оценки эффективности «единиц принятия решений» (от англ. decision-making units) [29], коими в нашем случае выступают регионы, точнее региональные инновационные системы (РИС). Но даже если говорить об оценке эффективности РИС по созданию инноваций, важно понимать, что в России многие университеты и научные институты находятся в ведении федеральных ведомств, а потому рассматривается полностью как управляемые компоненты РИС не могут.

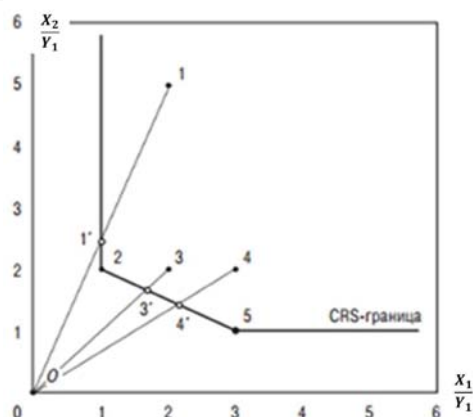


Рис. 3. Граница производственной эффективности [46]

где  $P_i$  – число патентов,  $L_i$  – число исследователей,  $K_i$  – объем затрат на НИОКР,  $\varepsilon$  – ошибка. Согласно формуле (13) получаем зависимость вида

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j \times x_i^{(j)} + \varepsilon \quad (15)$$

Тогда преобразованная модель (14) в линейном виде

$$P_i = \exp\left\{\beta_0 + \sum_{j=1}^p \beta_j \times x_i^{(j)} + V_i - U_i\right\} \quad (16)$$

описывает случайную величину, характеризующую фактический инновационный выпуск  $i$ -го региона. Если исключить из инновационного процесса все факторы неэффективности, то данный объем производства повысится до потенциального уровня

$$P_i^{pot} = \exp\left\{\beta_0 + \sum_{j=1}^p \beta_j \times x_i^{(j)} + V_i\right\} \quad (17)$$

Описанную таким образом зависимость результата инновационной деятельности от значений основных факторов при исключенном воздействии факторов неэффективности принято называть *граничным производственным потенциалом*. Техническая эффективность региона или региональной инновационной системы

$$E_i = \frac{P_i}{P_i^{pot}} = \exp\{-U_i\} \quad (18)$$

Заметим, что ТЕ является случайной величиной, принимающей значения из интервала

(0; 1]. Чем выше ТЕ, тем выше эффективность РИС [38, 39]. Но численно оценить  $U$  невозможно, поэтому используются оценки на основе подбора характеристик распределения случайной величины  $U^{24}$ . Преимуществом данного метода является возможность выявления факторов, влияющих на максимальный потенциальный выпуск.

Недостатком всех методов анализа эффективности является то, что чем больше факторов рассматривается, тем меньше различия в эффективности между регионами.

**Заключение.** Статистические методы региональных исследований инновационных процессов активно применяются как в России, так и за рубежом. Основные проблемы, которые призваны решить данные группы методов: выявление факторов инновационной активности, выявление различных типов регионов и оценка эффективности их инновационных систем. Для каждой из проблем были проанализированы основные методы анализа с описанием их достоинств и недостатков.

Для анализа и выявления основных условий и факторов инновационной деятельности в регионах России следует использовать метод множественной регрессии, основанный на применении модели производственной функции знаний (ПФЗ). Основной проблемой применения ПФЗ является отсутствие качественных, доступных и верифицируемых данных о зависимой переменной, что ведет к различного рода смещениям последующих оценок. Патентная актив-

<sup>24</sup> Реализация метода стохастической границы осуществлена в статистическом пакете Stata.

ность, чаще всего применяемая в качестве объясняемой переменной, не полностью соответствует инновационной активности, так как не все патенты могут быть воплощены в новой продукции. Для верификации авторы применяют разные индикаторы инновационной деятельности: российские и зарубежные патенты, инновационный выпуск, доля инновационных компаний и т.д., разные способы уточнения спорного индикатора: оценка вариации индикатора, методы факторного анализа, построение интегральных индексов, бинарные переменные и т.д., а также разные математические подходы: бинарные модели, осреднение данных по годам, метод аппроксимации на основе максимального правдоподобия, обобщенный метод моментов и т.д.

Для определения влияния факторов различных пространственных уровней для лучшего понимания механизмов стимулирования инновационной деятельности применяется метод многоуровневых регрессий. Последние работы показывают существенную значимость региональных факторов инновационной деятельности в сравнении с внутрифирменными, а соответственно необходимость поддержки региональных инновационных систем.

Проведение региональной инновационной политики требует выявления регионов-лидеров, в которых концентрируется инновационный потенциал, и поддержка которых стала бы наиболее эффективной. Для выявления разных типов регионов используются сочетания методов компонентного и кластерного анализов, когда на первом этапе выявляются наиболее значимые компоненты инновационного потенциала, а на втором проводится типология регионов исходя из набора выявленных компонент. Подобного

рода методика позволяет выявлять регионы-лидеры с оптимальным сочетанием наиболее важных факторов. Недостатком данных методов является невозможность контролировать сам процесс автоматического распределения индикаторов и регионов по факторам и типам, осуществляемый статистическими пакетами. Поэтому в ряде статей используются методы пороговых значений, позволяющие проводить кластеризацию регионов в «ручном режиме», но включающие в себя элементы экспертной оценки.

В условиях инновационной гонки все острее стоит проблема оценки эффективности инновационной политики, что требует применения особых методов анализа, к которым относятся оболочечный анализ данных и метод анализа стохастической границы. Два метода позволяют оценить эффективность региональных инновационных систем с точки зрения соотношения их выпуска и затраченных ресурсов. Таким образом, удастся выделить наиболее и наименее эффективные регионы для последующего принятия управленческих решений. Но по-прежнему актуальным остается вопрос о достоверном индикаторе инновационного выпуска.

Результаты всех статистических расчетов должны верифицироваться иными, в том числе географическими, методами. Зачастую понимание пространственных закономерностей позволяет более качественно объяснить выявленные факты.

Обзор новых методов регионального анализа, включающий бинарные регрессии, многоуровневые модели, анализ стохастической границы эффективности и оболочечный анализ данных, которые могут применяться не только для анализа инновационной активности, должен помочь в появлении новых направлений научного поиска.

#### Библиографический список

1. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификации и снижение размерности/ Под ред. С.А. Айвазяна. М.: Финансы и статистика, 1989. 607 с.
2. Архипов Ю.Р., Блажко Н.И., Григорьев С.В., Заботин Я.И., Трофимов А.М., Хузеев Р.Г. Математические методы в географии. Казань: Изд-во КГУ, 1976. 352 с.
3. Бабурин В.Л. Инновационные циклы в российской экономике. М.: УРСС, 2010. 216 с.
4. Бабурин В.Л., Земцов С.П. География инновационных процессов в России // Вестн. Моск. ун-та. Сер. 5. География. 2013. № 5. С. 25–32.
5. Бабурин В.Л., Земцов С.П. Оценка эффективности региональных инновационных систем в России // Модернизация и инновационное развитие экономических систем: коллективная монография / Под ред. В.Л. Матюшка. М.: РУДН. 2014. С. 18–36.
6. Боровиков В.П. STATISTICA. Искусство анализа данных на компьютере. СПб.: Питер, 2001. 656 с.
7. Бортник И., Зинов В., Коцюбинский В., Сорокина А. Вопросы достоверности статистической информации об инновационной деятельности в России // Инновации. 2013. №10 (180). С. 10–17.
8. Бутс, С. Дробышевский, О. Кочеткова, Г. Мальгинов, В. Петров, Г. Федоров, А. Хехт, А. Шеховцов, А. Юдин. Типология регионов РФ. М.: СЕПРА. 2002. 230 с.

9. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. Пер. с англ. М.: Мир, 1985. 450 с.
10. Гусейн-Заде С.М., Ханин С. Е. Вычислительные методы и использование ЭВМ в экономической и социальной географии. М.: Изд-во Моск. ун-та, 1987. 126 с.
11. Гуц А.К., Фролова Ю.В. Математические методы в социологии. М.: Издательство ЛКИ, 2010. 216 с.
12. Земцов С.П., Бабурин В.Л., Баринаова В.А. Как измерить неизмеримое? Оценка инновационного потенциала регионов России // Креативная экономика. 2015. №1 (97). С. 35–53.
13. Инновационный потенциал научного центра: Методологические и методические проблемы анализа и оценки / отв. ред. В.И. Суслов. Новосибирск: ИЭОПП, 2007. 275 с.
14. Карачаровский В.В. Стратегические изменения в российской инновационной системе (элементы кросс-отраслевого и кросс-регионального анализа) // Общество и экономика. 2012. № 5. С. 5-24.
15. Мандель И.Д. Кластерный анализ. М.: Финансы и статистика, 1988. 176 с.
16. Макаров В.Л., Айвазян С.А., Афанасьев М.Ю., Бахтизин А.Р., Нанавян А.М. Оценка эффективности регионов РФ с учетом интеллектуального капитала, характеристик готовности к инновациям, уровня благосостояния и качества жизни населения // Экономика региона. 2014. № 4. С. 9–30.
17. Мариев О.С., Савин И.В. Факторы инновационной активности российских регионов: моделирование и эмпирический анализ // Экономика региона. 2010. № 3. С. 235–244.
18. Петрунин, Ю. Ю. Информационные технологии анализа данных. Data Analysis. М.: КДУ. 2011. 239 с.
19. Ратникова Т.А. Введение в эконометрический анализ панельных данных // Экономический журнал ВШЭ. 2006. №10(2). С. 267-316.
20. Сердюкова Ю.С., Валиева О.В., Суслов Д.В., Старков А.В. Инновационная система в регионах России: оценка состояния и развития // Регион: экономика и социология. 2010. № 1. С. 179–197.
21. Синергия пространства: региональные инновационные системы, кластеры и перетоки знания / Отв. ред. А. Н. Пилясов. Смоленск: Ойкумена, 2012. 760 с.
22. Халафян А.А. Statistica 6. Статистический анализ данных. М.: ООО «Бином-Пресс», 2010. 528 с.
23. Чертко, Н.К., Карпиченко А.А. Математические методы в географии. Минск: БГУ, 2008. 340 с.
24. Штерцер Т.А. Эмпирический анализ факторов инновационной активности в субъектах Российской Федерации // Вестник НГУ. Серия: социально-экономические науки. 2005. Т. 5. Вып. 2. С. 23-36.
25. Acs Z., Anselin L., Varga A. Patents and innovation counts as measures of regional production of new knowledge // Research policy. 2002. №31(7). P. 1069-1085.
26. Aigner D., Lovell C., Schmidt P. Formulation and Estimation of Stochastic Frontier Production Function Models // Journal of Econometrics. 1977. Vol. 6. № 1. P. 37–49.
27. Banker R., Charnes A., Cooper W. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis // Management Science. 1984. V. 30. № 9. P. 1078-1092.
28. Brenner T., Broekel T. Methodological issues in measuring innovation performance of spatial units // Industry and Innovation. 2011. №18 (1). P. 7-37.
29. Charnes A., Cooper W., Rhodes E. Measuring the efficiency of decision-making units // European Journal of Operation Research. 1978. V. 2. № 6. P. 429-444.
30. Chen K., Guan J. Measuring the efficiency of China's regional innovation systems: application of network data envelopment analysis (DEA) // Regional Studies. 2012. №46(3). P. 355-377.
31. Cooper W., Seiford L., Tone K. Data Envelopment Analysis. A Comprehensive Text with Models, Applications, References and DEA-Solver Software. 2nd edition. – New York: Springer-Verlag, 2006. 528 p.
32. Fagerberg J., Sroolec M. National innovation systems, capabilities and economic development. // TIK Working Paper on Innovation Studies. – Oslo: Centre for technology, innovation and culture, 2007. 190 p.
33. Fahrmeir L., Kneib Th., Lang S., Marx B. Regression. Models, Methods and Applications. – Heidelberg: Springer, 2013. 698 p.
34. Farrell M. The Measurement of Productive Efficiency // Journal of the Royal Statistical Society. 1957. V. 120. P. 253–281.
35. Feldman M. The Geography of Innovation. Boston: Kluwer Academic Publishers, 1999. 257 p.
36. Fischer M., Getis, A. Handbook of applied spatial analysis. Heidelberg: Springer. 2009. 390 p.
37. Fischer M., Nijkamp P. (Ed.). Handbook of regional science. Heidelberg: Springer. 2014. 950 p.
38. Fritsch M. Measuring the quality of regional innovation systems: a knowledge production function approach // International Regional Science Review. 2002. №25(1). P. 86-101.
39. Furman J., Porter M., Stern S. The determinants of national innovative capacity // Research policy. 2002. №31(6). P. 899-933.
40. Goldstein H. Multilevel Statistical Models. London: Arnold, 2003. 510 p.
41. Griliches Z. R&D, patents, and productivity. Chicago: University of Chicago, 1984. 340 p.
42. Griliches Z. R&D and productivity: the econometric evidence. – Chicago: University of Chicago, 1998. 350 p.
43. Isard W. Methods of Regional Analysis. Cambridge, 1965.
44. Jaffe A., Trajtenberg M. Patents, citations, and innovations: A window on the knowledge economy. Boston: MIT press, 2002. 150 p.
45. Koopmans T.C. Analysis of production as an efficient combination of activities // Activity analysis of production and allocation. 1951. T. 13. C. 33-37.
46. Kotsemir M.N. Measuring national innovation systems efficiency—a review of DEA approach // Higher School of Economics Research Paper No. WP BRP. 2013. T. 16. P. 15 – 39.
47. Mansfield E. Patents and innovation: an empirical study // Management science. 1986. №32(2). P. 173-181.
48. Moser P. Patents and innovation: Evidence from economic history // The Journal of Economic Perspectives. 2013. №1. P. 23-44.
49. Oja E., Hyvryinen A., Karhunen J. Independent Component Analysis. – Wiley Interscience, 2001. 504 p.
50. Romer P. Increasing returns and long run growth // Journal of Political Economy. 1986. №94. P. 1002–1038.
51. Scotchmer S. Standing on the shoulders of giants: cumulative research and the patent law // The Journal of Economic Perspectives. 1991. №1. P. 29-41.
52. Sroolec M. A multilevel approach to geography of innovation // Regional Studies. 2010. №44(9). P.1207-1220.
53. Sternberg R., Arndt O. The firm or the region: what determines the innovation behavior of European firms? // Economic Geography. 2001. №77(4). P. 364-382.
54. Typology of regions in Europe. 2010. URL: [http://www.i2sare.eu/uploadedFiles/Report/Regional%20typology/I2SARE\\_WorkReport\\_WP6\\_final.pdf46](http://www.i2sare.eu/uploadedFiles/Report/Regional%20typology/I2SARE_WorkReport_WP6_final.pdf46)