

**ОБ ЭКОНОМИЧЕСКИХ ИЗМЕРЕНИЯХ: ВЕРОЯТНОСТЬ И ДОСТОВЕРНОСТЬ,  
МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, БОЛЬШИЕ ДАННЫЕ,  
ЭЛЕКТРОННАЯ СТАТИСТИКА**

**В.И. Суслов**

*В статье вслед за краткой характеристикой роли И.П. Сусллова в развитии отечественной статистики XX века, 100-летие со дня рождения которого отмечалось в 2015 г., последовательно изложен авторский подход к проблеме экономических измерений в статистике. Раскрывается содержание концепции экономических измерений, показывается современная взаимосвязь между экономическими измерениями и математическим моделированием, выражается авторская позиция к интерпретации проблем больших данных и электронной статистики.*

*Применительно к предмету исследования автором подчеркивается, что неточность, неопределенность, неполнота - фундаментальные свойства экономических измерений, определяющие специфику вероятностного подхода к интерпретации статистических показателей. Использование больших данных существенно меняет понимание роли эмпирической информации для принятия практических решений и проверки теоретических гипотез при формировании научных концепций. Аргументируется тезис о том, что сбор и обработка первичной информации о состоянии и динамике социально-экономических процессов в современных условиях должны быть встроены в информационное поле электронного учета, связанного с обработкой документов, сопровождающих движение продукции и услуг, товаров, капиталов, денег.*

*Ключевые слова:* экономические измерения, достоверность, большие данные, электронная статистика.

*JEL:* C81, C82.

В 2015 г. исполнилось 100 лет со дня рождения Ивана Петровича Сусллова, известного российско-го экономиста-статистика, методолога научного познания, одного из основателей отечественной теории статистических показателей. Его именем названа одна из премий, врученных молодым статистикам на прошедшем статистическом конгрессе в г. Новосибирске.

Этому событию было посвящено два важных мероприятия: 18 марта - заседание секции социально-экономических проблем статистики Санкт-Петербургского Дома ученых им. М. Горького РАН (его вела член-корреспондент РАН И.И. Елисеева) и 14 апреля - заседание Ученого совета, совмещенного с методологическим семинаром, Института экономики и организации промышленного производства СО РАН (г. Новосибирск - в числе выступавших были академики РАН А.Г. Аганбегян, В.В. Кулешов, члены-корреспонденты РАН К.К. Вальтух, В.И. Суслов). По материалам этих заседаний издана книга [2], в которой представлены не только

презентации докладов и тексты выступлений, но и воспоминания родственников и близких Ивана Петровича, а также одна из первых его литературных работ - «Записки экономиста» - о его жизни в последний предвоенный год в городе Архангельске, где он проходил послевузовскую (после окончания ЛФЭИ, нынешнего ФИНЭКа) производственную практику.

След, оставленный И.П. Сусловым в нашей экономической науке, заметен. Значимым является и его вклад в становление вероятностного подхода к экономическим измерениям.

Приведем несколько цитат из статьи Ивана Петровича «Вероятность в системе научных категорий»:

«В нашей литературе бытует точка зрения, отрицающая вероятностные процессы в общественной жизни и возможность использования для ее познания вероятностных методов. Эта точка зрения обосновывается тем, что в общественной жизни движение индивидуальных явлений не беспоря-

*Суслов Виктор Иванович (suslov@ieie.nsc.ru) - д-р экон. наук, член-корреспондент Российской академии наук, зам. директора по науке, Институт экономики и организации промышленного производства СО РАН (г. Новосибирск, Россия).*

\* Продолжаем публикацию журнальных версий докладов, вызвавших наибольший интерес у участников I Открытого российского статистического конгресса (г. Новосибирск, 20-22 октября 2015 г.).

дочно, не случайно, а определяется сознательными действиями людей. Такую аргументацию вряд ли можно считать убедительной. Сознательное, но неопределенное поведение человека отнюдь не исключает наличие случайностей в общественных процессах, а предопределяет его» [8, с. 56].

«Общество является сложнейшей динамической системой, развивающейся под воздействием как необходимости (что определяет ее детерминированность), так и многих случайных факторов (что придает ей объективные случайные свойства)» [8, с. 57]. «Вероятность, вероятностные процессы присущи объективной действительности, категория вероятности органически и неизбежно вписывается в систему научных категорий, отражающих реальный мир. Наряду с объективной, как отмечалось выше, существует и вероятность субъективная» [8, с. 56].

Нельзя не отметить, что в 70-х годах прошлого века говорить в нашей стране о вероятности (и о, ужас - о субъективной вероятности) применительно к социалистической общественной жизни было актом гражданского мужества.

### **Концепция экономических измерений в статистике**

В 2005 г. под эгидой экономического факультета Новосибирского государственного университета (НГУ) вышла книга, скромно названная нами «Эконометрия» (не «Эконометрика») [7]. Это толстый учебник (744 с.), состоящий из трех частей, отражающих три учебных курса, созданных и ведущихся на факультете нашим небольшим творческим коллективом: введение в социально-экономическую статистику (то, что раньше называлось общей теорией статистики), базисный курс эконометрии (Эконометрия I: регрессионный анализ и анализ временных рядов) и продвинутый курс эконометрии (Эконометрия II).

Публикация «Эконометрии» - итог работы над учебными статистическими курсами разной направленности и глубины, которая длилась на факультете на протяжении четверти века, в том числе с поддержкой грантами по линии TEMPUS-TACIS. По нашему мнению, этот учебник не потерял своей актуальности и сейчас, через 10 лет после выхода в свет. Учебника такого же уровня в России за это время так и не вышло. Мы начинаем работать над вторым изданием этого труда, обновленным и дополненным.

Специфика экономических измерений, с нашей точки зрения, в первой части книги раскрыта достаточно полно.

Вот набор из нескольких цитат.

«...В строгом смысле статистика является наукой о методах количественного (численно-го) отражения фактов общественной жизни» [7, с. 18]. «...Общественные науки вынуждены опираться на неэкспериментальные данные, то есть на результаты пассивных наблюдений, в потоке которых трудно уловить, а тем более, количественно определить причинно-следственные связи. И статистика как раз и занимается методами сбора и подготовки таких данных к анализу, методами их первичного анализа, методами проверки теоретических гипотез на основе таких данных» [7, с. 19].

«Экономическая величина - суть некоторое количество определенного экономического "качества". Обычно экономические величины обозначают буквами латинского, реже - греческого алфавита». «Обозначенные таким образом экономические величины используются обычно как переменные и параметры математических моделей экономики, в которых устанавливаются зависимости между экономическими величинами» [7, с. 20].

«Статистическим (экономическим) показателем является операциональное определение экономической величины. Такое определение представляет собой исчерпывающий перечень операций, которые необходимо провести, чтобы измерить данную величину»; «...экономическая величина-признак - теоретическое понятие, статистический показатель-определение - обеспечивает практическую измеримость теоретической величины, статистический показатель-наблюдение - результат измерения величины-признака конкретного объекта в конкретный момент времени» [7, с. 21].

«Все экономические величины без исключения являются случайными с вполне определенными, часто неизвестными законами распределения вероятности. Наблюдаемые значения суть реализации соответствующих случайных величин, выборки из каких-то генеральных совокупностей». «В физике, химии генеральные совокупности очень велики, многие из них, по-видимому, можно считать бесконечными... В такой ситуации совершенно естественным кажется определение вероятности как пре-

дела относительной частоты появления нужного признака (Пуассон)». «Но и в физическом мире некоторые явления, если относиться к ним без должной фантазии, представляются единичными и уникальными, со всеми вытекающими отсюда трудностями для классического, “объективистского”, “частоточного” понимания вероятности» [7, с. 22].

«Можно сказать, что вся экономика состоит из таких нарушений. Мир людей, если к нему относиться “сильно материалистично”, без некоторой раскованности в мышлении, уникален и ограничен» [7, с. 23]. Вряд ли кто подвергнет сомнению разумность применения вероятностных характеристик, например к Саяно-Шушенской ГЭС. Но это будет совсем не «частоточная» вероятность, поскольку много раз, очень много раз, бесконечное количество раз построить такую ГЭС невозможно (хорошо, что это удалось сделать всего один раз). «В таком положении полезным и продуктивным, по крайней мере, внешне, представляется подход субъективной вероятности. Субъективная вероятность - это мера доверия исследователя к утверждению, степень уверенности в его справедливости, наконец, мера готовности действовать в ситуации, связанной с риском» (Бернулли) [7, с. 23].

«Существует подход, объединяющий в определенном смысле идеи субъективной и объективной вероятности. Он основан на понимании многовариантности развития общества вообще и экономики в частности». «Имеется множество возможных состояний экономики и путей ее развития, наблюдаемые факты в полном своем объеме являются лишь выборкой из гипотетической генеральной совокупности, образованной этим множеством». «...как можно практически работать с гипотетическими генеральными совокупностями?». «Современная экономическая наука располагает соответствующим инструментарием: это математическое моделирование». «Всякая математическая модель представляет бесконечное пространство возможных состояний экономики, расчет по модели дает точку или траекторию в этом пространстве». «Модель выступает инструментом проведения экономических экспериментов почти в таком же смысле, как и в естественных науках» [7, с. 24].

«По-видимому, “субъективист”, хотя бы в некоторых ситуациях, приписывая вероятности тем или иным событиям, пользуется неявно частот-

ным подходом применительно к некоторым гипотетическим генеральным совокупностям. При этом конструировать эти гипотетические совокупности и работать с ними помогают ему его знания, опыт и интуиция» [7, с. 24].

«Если взять Большую Советскую Энциклопедию или Математическую энциклопедию более позднего издания, то можно узнать, что измерение - это процесс сопоставления измеряемого явления с единицей измерения. Такое определение достаточно поверхностно, оно не раскрывает существа возникающих проблем» [7, с. 25]. Для экономики оно совершенно не подходит: в ней (экономике) нет таких явлений или объектов и не изобретено таких «линеек», совмещение которых позволит измерить темп роста ВВП, уровень инфляции, активы и пассивы фирмы и т. д. «В настоящее время практически всеобщим признанием пользуется репрезентативная теория, в соответствии с которой измерение есть процесс присваивания числовых выражений объекту измерения для его репрезентации (представления)» [7, с. 25].

Приведем формальное определение, которое вытекает из теории математических моделей А. Тарского, выдающегося польско-американского математика.

«Измерить, значит установить однозначное (гомоморфное) отображение эмпирической реляционной структуры в числовую реляционную структуру. Реляционная структура - это множество объектов вместе со всеми отношениями и операциями на нем... если объекты находятся в реальной действительности (в эмпирической реляционной структуре) в некоторых отношениях друг с другом (одинаковы, больше, меньше, лучше, хуже, являются суммой или разностью), то и в этих же отношениях должны находиться числа, приписанные им в результате измерения (числовая реляционная структура)» [7, с. 25].

Измерения бывают первичными и производными. Первые «сводятся к проведению эмпирических операций в непосредственном контакте с измеряемым объектом». Вторые «связаны с проведением вычислительных операций над первично измеренными величинами» [7, с. 26-27].

Экономические измерения всегда производны, являются результатом некоторой «свертки» результатов первичных измерений, не имеющих экономического характера.

«Измеряться могут только операционально определенные величины. В экономике разработка операциональных определений величин - это сложный и неоднозначный исследовательский процесс теоретического характера. Теоретики постоянно дискутируют на темы измерения общих итогов экономического развития, экономической эффективности, производительности общественного труда, экономической динамики, инфляционных процессов, структурных сдвигов и т. д.». То есть «экономические измерения, в отличие от многих физических, в очень сильной степени обусловлены теоретическими моментами» [7, с. 27].

«Неэкспериментальные данные исключают возможность анализа “при прочих равных”. В потоке наблюдений за “всеми сразу” величинами, как уже отмечалось, трудно уловить структуру взаимосвязей и измерить их интенсивность. Чисто эмпирически это, пожалуй, невозможно сделать. Это обстоятельство еще в большей степени увеличивает нагрузку на теорию, “силу абстракции” исследователя. И оно не добавляет надежности результатам измерения» [7, с. 28].

В экономических измерениях гипертрофировано влияние субъективных факторов. В результатах измерения заинтересованы как те люди, деятельность которых измеряется, так и те, которые измеряют. Сознательные искажения результатов измерения могут достигать огромных масштабов. Классический пример являют многие работы Г.И. Ханина. По его оценкам [4], реальный рост национального дохода за период с начала 1-й пятилетки (конца 20-х годов прошлого столетия) до начала 80-х годов прошлого века составлял не 90 раз, как по официальной статистике, а всего 7-8 (что тоже, кстати, очень неплохо).

«Любые измерения, а экономические в особенности, содержат ошибку. Точные величины - суть не более чем теоретические абстракции» [7, с. 29].

### **Экономические измерения и математическое моделирование**

Есть несколько моих работ, выполненных на стыке проблематики экономических измерений и математического моделирования. К ним относятся, в частности, две авторские монографии.

В *первой* из них (по результатам кандидатской диссертации) [5] изучаются методы постро-

ения (измерения показателей) региональных межотраслевых балансов и исследуется их связь с методами и моделями пространственного анализа (анализа межрегиональных экономических взаимодействий). В частности, в ней показывается, что использование цен потребителей или производителей дает возможность применять только балансовые модели (цены производителей приводят к известной балансовой модели Айзарда), а так называемые смешанные цены конечного потребления, использовавшиеся обычно в практике построения межотраслевых балансов СССР, вообще непригодны для последующего анализа. Оптимизационные модели можно корректно строить только на базе введенных в работе «цен региона», которые одновременно являются ценами производителей для производимой в регионе продукции и ценами потребителей для потребляемой в нем продукции.

Во *второй* из этих двух книг (отражающей результаты докторской диссертации) [6] показано, как следует использовать большую прикладную многорегиональную оптимизационную модель (разработанную и введенную в теорию и практику пространственного анализа А.Г. Гранбергом) для измерения эффектов межрегиональных взаимодействий.

Измерение этих эффектов основывается на использовании положений теории экономического равновесия (Вальрас) и теории кооперативных игр (ядро Нэша). В обоих случаях регионы представляются субъектами рыночной системы. В первом случае речь идет о товарно-денежном рынке, в котором регионы принимают решения о своих внешнеторговых связях, ориентируясь на сложившиеся цены обмена и свои бюджеты. Особый интерес представляют равновесия с нулевыми сальдо региональных бюджетов. Это состояния эквивалентного межрегионального обмена.

Во втором случае рынок контрактный: регионы ищут выгодных для себя партнеров и заключают между собой коалиционные контракты-соглашения. Та часть парето-множества, в которой никакой коалиции регионов не выгодно выходить из общей системы, называется ядром системы. Это состояния взаимовыгодного обмена. Их находят (точнее, очерчивают их зону) в результате весьма трудоемкого анализа (включающего десятки тысяч компьютерных реализаций большой прикладной многорегиональной модели), названного коалиционным.

Результатом такого коалиционного анализа является также шахматная таблица эффектов межрегиональных взаимодействий.

Такие измерения иллюстрируются расчетами по анализу межреспубликанских экономических взаимодействий накануне распада СССР. Этот анализ был проведен во второй половине 80-х годов прошлого века с помощью модели ОМММ (оптимизационной многорегиональной межотраслевой модели) СССР в разрезе 15 союзных республик с учетом внешней торговли и 30 отраслей материального производства. В приводимых ниже таблицах и схемах результаты агрегированы в более крупные регионы, чтобы сделать их более наглядными.

Обращает на себя внимание тот факт (см. таблицу 1), что только Россия (РСФСР) в состоянии полной автаркии могла бы сохранить значение своего целевого показателя на достаточно высоком уровне. Казахстан, Средняя Азия, Закавказье теоретически могли бы потерять после разрыва межреспубликанских связей почти три чет-

верти своего потребления. Для остальных республик последствия разрыва связей могли бы быть еще более катастрофичными (для Украины - семикратное сокращение). Как видно из данных таблицы 1, Россия прямо и косвенно обеспечивала более половины (до двух третей) потребления населения Украины, Белоруссии, Прибалтики. Для самой же России межреспубликанские связи не слишком важны (14,2% потребления). Гораздо важнее для нее внешнеэкономические связи (21,2%). Интересно, что взаимосвязи Украины и Молдавии обеспечивали более половины потребления населения последней, в то время как для Украины они имели негативные последствия. Только для РСФСР сальдо межреспубликанских взаимодействий положительно (вклад ее в общесистемное потребление превышает ее потребление, обусловленное внутрисистемными связями - последний столбец таблицы 1). Сальдо межреспубликанских взаимодействий остальных республик отрицательно.

Таблица 1

**Эффекты фактических межрегиональных взаимодействий**  
(1987 г., % к непроизводственному потреблению республик)

	РСФСР	Украина	Белоруссия	Казахстан	Средняя Азия	Молдавия	Закавказье	Прибалтика	Итого (общий вклад)
РСФСР	64,6	67,3	55,5	42,5	36,3	31,7	35,8	65,0	60,2 (+14,5)
Украина	1,2	14,8	16,5	4,9	18,0	52,1	7,4	8,1	6,3 (-9,8)
Белоруссия	2,3	4,0	3,8	3,5	2,1	4,1	3,3	3,7	2,8 (-0,8)
Казахстан	1,7	0,6	-1,4	27,1	3,8	-0,6	6,7	-0,6	3,0 (-1,4)
Средняя Азия	3,7	1,1	15,4	0,5	26,4	1,7	-0,0	2,8	4,8 (-1,5)
Молдавия	0,8	-2,7	-0,3	0,7	0,3	0,0	0,6	0,9	0,1 (0,0)
Закавказье	2,6	1,7	0,5	4,5	3,9	0,2	25,7	0,7	3,4 (0,0)
Прибалтика	1,9	1,5	4,3	3,3	2,5	1,9	2,7	8,0	2,2 (-1,0)
Внутренний эффект	78,8 (45,7)	88,3 (16,1)	94,3 (3,6)	87,0 (4,4)	93,3 (6,3)	91,1 (0,1)	82,2 (3,4)	88,6 (3,2)	82,8 (0,0) (82,8)
Внешние связи	21,2	11,7	5,7	13,0	6,7	8,9	17,8	11,4	17,2
Итого (потребление населения)	100 (58,1)	100 (18,5)	100 (3,8)	100 (5,1)	100 (6,7)	100 (0,1)	100 (4,2)	100 (3,6)	100 (100,0)

Несколько иную картину давали результаты равновесного анализа (по Вальрасу и Нэшу). Зона ядра (см. рис. 1) сильно вытянута в сторону увеличения доли РСФСР в общесистемном непроизводственном потреблении. Это означает, что

непроизводственное потребление России могло бы быть значительно увеличено за счет других республик, но межреспубликанский обмен оставался бы взаимовыгодным, то есть коалиции республик без РСФСР имели бы меньшее потребление.

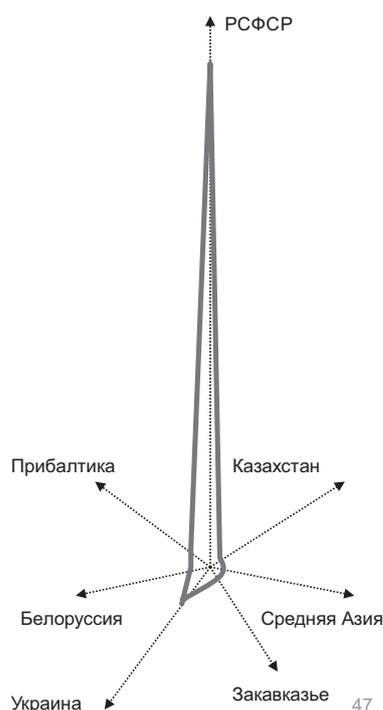


Рис. 1. Зона ядра

Таблица 2

Территориальная структура непроизводственного потребления  
(1987 г., процентные пункты)

Макрорегионы	Факт	Нижний предел ядра	Эквивалентный обмен	Верхний предел ядра
РСФСР	58,06	56,25	56,37	89,62
Украина и Молдавия	18,58	19,21	19,42	19,82
Белоруссия	3,79	4,76	5,01	5,08
Казахстан	5,07	3,91	4,36	4,41
Средняя Азия	6,71	5,05	5,54	5,59
Закавказье	4,24	4,92	5,49	5,52
Прибалтика	3,55	3,56	3,81	3,89
Итого	100	-	100	-

При этом фактическая доля непроизводственного потребления России выше ее доли в состоянии эквивалентного обмена (см. таблицу 2). То есть ее потребление преувеличено по сравнению с тем, которое имело бы место при эквивалентном межреспубликанском обмене. Такая же ситуация, но гораздо в большей степени, была характерна для Казахстана и Средней Азии. А вот потребление Украины, Закавказья, Прибалтики, и особенно Белоруссии, занижено.

## О проблеме больших данных

Большие данные (Big Data) - это полная, исчерпывающая, а не выборочная информация об объекте в режиме онлайн (см., например, [3]). Использование такой информации исключает наиболее распространенные и наиболее значимые ошибки, вызванные выборочным характером данных. Еще важнее другое: актуальность выявления причинности теряется. Найденные простые корреляции становятся надежными основаниями для принятия решений.

Сегодня безусловными лидерами в сфере больших данных являются США, Великобритания, Япония и Китай. В нашей стране разработана мощнейшая алгоритмическая и математическая база для интеллектуального анализа больших данных, но она фактически не применяется. Большие данные, генерируемые таможенными, налоговыми, транспортными службами, торговыми сетями, финансовыми структурами, лежат мертвым грузом. То, что у нас называют анализом больших данных, в подавляющей части - это уже много лет применяемая за рубежом традиционная бизнес-аналитика.

Масштабы накопленных данных невообразимо огромны.

*О единицах измерения и объемах информации.*

*Мегабайт (МБ)* -  $10^6$  (миллион) байт; 5 МБ - полное собрание сочинений Шекспира.

*Гигабайт (ГБ)* -  $10^9$  (миллиард) байт; 10 ГБ - полнометражный фильм в хорошем качестве (2 тыс. библиотек Шекспира).

*Терабайт (ТБ)* -  $10^{12}$  (триллион) байт; 10-20 ТБ - печатные тексты библиотеки Конгресса США (1-2 тыс. полнометражных фильмов, 4-5 млн библиотек Шекспира).

*Петабайт (ПБ)* -  $10^{15}$  (квотраллион) байт; 3-4 ПБ - библиотека Конгресса США с учетом аудио- и видеоформатов; несколько ПБ - объем накопленных данных в базах данных (БД) ряда российских компаний («Мегафон», «РЭЛЭКС» и др.).

*Эксабайт (ЭБ)* -  $10^{18}$  (квинтиллион) байт; несколько ЭБ - потенциальная емкость современной БД.

*Зетабайт (ЗБ)* -  $10^{21}$  (за квинтиллионом предлагают использовать термин «секстиллион»); 10 ЗБ - объем накопленных данных на планете на 2015 г. (несколько сот тыс. библиотек Конгресса США).

*Йотабайт* (ИБ) -  $10^{24}$  (следуя логике - септиллион), объем накопленных данных на 100 планетах типа «Земля-2015». По-видимому, это 7-10-летняя перспектива для нашей планеты.

Имеется только одна российская система управления базами данных (СУБД), получившая коммерческое признание. Это - ЛИНТЕР, продукт НТЦ «РЕЛЭКС» (РЕЛЯционные ЭКспертные Системы<sup>1</sup>). Эта база данных, мощностью почти эксабайт (несколько сот библиотек Конгресса США), является классической реляционной СУБД, то есть хранилищем таблиц - жестко структурированных данных. Традиционным средством доступа к данным в таких СУБД является язык SQL - *Structured Query Language* - «структурированный язык запросов».

Однако одной из основных особенностей больших данных является их неоднородность и плохая структурированность. Единицами информации в них может быть все что угодно: не только числа или таблицы, но и тексты, картинки, ролики и т. д. Для работы с такими данными используются подходы, совокупность которых получила название NoSQL - *not only SQL*, «не только SQL»<sup>2</sup>. Ключи доступа к данным в СУБД, реализующих эти подходы, могут быть самыми разнообразными, а не только как в реляционных БД: имя таблицы, строки и столбца. Большое распространение получили БД типа «ключ-значение», в которых пользовательский адрес (ключ) имеет каждая единица данных.

Другой особенностью больших данных является то, что их техническое обеспечение ориентировано не на суперкомпьютеры, нацеленные на вычисления и численное моделирование, мощность которых измеряется количеством операций с числами в секунду (флопсами), а на кластеры компьютерных устройств, состоящих из сотен тысяч узлов, мощность которых измеряется байтами информации и скоростью доступа к ним. Эти кластеры реализуют облачные и распределенные технологии. Одной из наиболее успешных таких реализаций является проект Hadoop организации Apache Software Foundation<sup>3</sup>. Реализация этого проекта была инициирована Дугом Каттингом в 2005 г., который и назвал его в честь слоненка - любимой игрушки своей дочери.

<sup>1</sup> <http://www.linter.ru/>.

<sup>2</sup> <https://ru.wikipedia.org/wiki/NoSQL>.

<sup>3</sup> <https://ru.wikipedia.org/wiki/Hadoop>.

<sup>4</sup> [http://soc-research.info/blog/index\\_files/tag-04140440044e-041a043e043d043204350439.html](http://soc-research.info/blog/index_files/tag-04140440044e-041a043e043d043204350439.html).

Однако суть феномена «большие данные» составляют технологии цепочки «данные - информация - знания» (интеллектуальный анализ данных - *data mining*). В начале ее (цепочки) огромные массивы плохо структурированных, противоречивых и несопоставимых данных; в конце - знания, полезные для принятия решений. В начале - «руда»; в конце - «золото» и «алмазы».

Теоретической разработкой таких технологий занимается формируемая «наука о данных» (теория информации Шеннона к ней не имеет никакого отношения).



Рис. 2. Положение «науки о данных»

На рис. 2 показано положение «науки о данных», предложенное крупнейшим специалистом в этой области, бывшим работником американских спецслужб Дрю Конвеем<sup>4</sup>.

### Электронная статистика

На больших данных должны (могут) основываться такие направления развития теории и практики, как агента-ориентированное моделирование, экспертные когнитивные системы, сетевая и цифровая экономика, а также электронная статистика. В данном случае под *электронной статистикой* понимается не теория и практика перевода в электронные форматы существующих форм статистической отчетности и результирующих таблиц статистической информации. Речь идет об автоматической (компьютерной) ге-

нерации полного набора статистических документов, отражающих состояние и динамику развития социально-экономической системы, на основе первично регистрируемых документов, которые жестко контролируются государством в лице, прежде всего, налоговых, таможенных и других органов. Эти первичные документы в своем большинстве (включая «платежки», таможенные декларации, магазинные чеки и др.) уже давно существуют в электронных форматах. Их надо только немного модифицировать, чтобы на их основе с помощью некоей системы интеллектуальной обработки (ее предстоит создать) в концепции больших данных сгенерировать всю требуемую статистику.

Необходимая модификация: введение в эти документы единого кода продукции-услуг.

В 1999 г. одной из консалтинговых фирм г. Новосибирска совместно с Институтом экономики и организации промышленного производства СО РАН удалось инициировать Постановление Правительства РФ (№ 1189) «О проведении экономического эксперимента в Новосибирской области», предусматривающего введение универсального кода в платежные документы. Противодействие областных налоговых органов поставило тогда на этом Постановлении «крест». Хотя и нам не удалось в то время разработать универсальный код продукции и услуг, несмотря на то, что были оцифрованы и включены в анализ практически все существовавшие в СССР классификации продукции и услуг (десятки миллионов позиций). Этот искомый код мы тогда называли расширенным - он включал 10-12 десятичных позиций.

По-видимому, с использованием современных средств интеллектуального анализа данных, распознавания образов, когнитивных систем эта проблема может найти решение.

Тогда же - в 1999-2000 гг. - был выполнен исследовательский проект по построению всей системы статистических показателей, включая ВВП и МОБ в СНС, на базе первичных платежных документов, содержащих расширенный код продукции и услуг. Результаты этой работы были опубликованы в [1]. В обозначениях данной публикации I и II квадранты МОБ в СНС по методу валового оборота имеют следующий вид:

$$\widehat{x}_i^{*\Omega} - \widehat{x}_{iP}^{\Omega*} = x_i^{n*} + x_{in}^{\Omega} + x_{ig}^G + \Delta s_i^{\Omega} + x_{iu}^{\Omega} + x_i^{\Omega}.$$

В такой системе статистики, основанной на больших данных, жестко контролируемых государством, будут легко выявляться «нестыковки», свидетельствующие о наличии «серых» и «черных» областей в экономике. Это позволит эффективно бороться с теневой экономикой, коррупцией, коренным образом увеличить достоверность статистики.

Статистические органы в их существующем формате, на наш взгляд, окажутся ненужными. А статистики во взаимодействии с экономистами, математиками, социологами, специалистами в области информационных технологий смогут сосредоточиться на решении действительно важных задач: на совершенствовании операционных определений экономических величин и их совокупностей, с целью адекватного отражения состояния и динамики развития социально-экономических систем.

### Литература

1. Дашут Е.С., Суслов В.И. Проблемы создания единого информационного пространства материально-финансовых потоков в регионе // Регион: экономика и социология. 2000. № 2. С. 37-64.
2. К 100-летию со дня рождения Ивана Петровича Сулова: ученый, учитель, человек. - Новосибирск: Изд-во ИЭОПП СО РАН, 2015. - 230 с.
3. Майер-Шенбергер В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. - М.: Манн, Иванов и Фербер, 2014. - 221 с.
4. Селюнин В.И., Ханин Г.И. Лукавая цифра // Новый мир. 1987. № 2. С. 181-201.
5. Суслов В.И. Измерение и анализ региональной материалоемкости производства. - Новосибирск: Наука, 1982. - 191 с.
6. Суслов В.И. Измерение эффектов межрегиональных взаимодействий: модели, методы, результаты. Новосибирск: Наука, 1991. - 251 с.
7. Суслов В.И., Ибрагимов Н.М., Тальшева Л.П., Цыплаков А.А. Эконометрия: учебник для вузов / отв. ред. Г.М. Мкртчян; НГУ, Нац. фонд подготовки кадров. Новосибирск: Изд-во СО РАН, 2005. - 744 с.
8. Суслов И.П. Вероятность в системе научных категорий // Динамическая и вероятностная оптимизация экономики. Новосибирск: Наука, 1978.

**ECONOMIC DIMENSIONS: PROBABILITY AND RELIABILITY, MATHEMATICAL MODELING,  
BIG DATA, E-STATISTICS**

*Victor Suslov*

*Author affiliation:* Institute of Economics and Industrial Engineering of the Siberian Branch of the RAS (Novosibirsk, Russia).  
E-mail: suslov@ieie.nsc.ru

After the brief description of the role Suslov I.P. (in 2015 was celebrated the 100<sup>th</sup> anniversary of his birth) had in the development of national statistics in XX century, the article consistently describes the author's approach to the problem of measuring economic statistics. The content of economic measurement is revealed; the author shows the modern relationship between the economic dimension and mathematical modeling; the author's position in interpreting the problems of big data and e-statistics is expressed.

With regard to the subject of the study the author emphasizes that accuracy, uncertainty, under coverage - are the fundamental features of economic dimensions that influence the probabilistic approach to the interpretation of statistical indicators. Using Big Data significantly changes the understanding of the role of empirical information for practical decision-making and testing of theoretical hypotheses in forming scientific concepts. Collection and processing of primary information on the status and dynamics of economic processes, statistical reporting of enterprises and organizations must be based on electronic bookkeeping of tax, customs and others documents, accompanying the movement of goods, services and capital.

*Keywords:* economic measurements, reliability, Big Data, e-statistics.  
*JEL:* C81, C82.

### References

1. **Dashut E.S., Suslov V.I.** Problemy sozdaniya yedinogo informatsionnogo prostranstva material'no-finansovykh potokov v regione [Problems of creating a unified information space of material and financial flows in the region]. *Region: Economics and Sociology*, 2000, no. 2, pp. 37-64. (In Russ.).
2. *K 100-letiyu so dnya rozhdeniya Ivana Petrovicha Suslova: uchenyy, uchitel', chelovek* [On the 100<sup>th</sup> anniversary of the birth of Ivan Petrovich Suslov: A scientist, a teacher, a man]. Novosibirsk, IEIE SB RAS Publ., 2015. 230 p. (In Russ.).
3. **Mayer-Schönberger V., Cukier K.** *Big Data: A revolution that will transform how we live, work, and think*. New York, Houghton Mifflin Harcourt Publishing Company, 2014 (Russ. ed.: Majer-Shenberger V., Kuk'er K. *Bol'shiye dannyye. Revolyutsiya, kotoraya izmenit to, kak my zhivem, rabotayem i myslim*. Moscow, Mann, Ivanov i Ferber Publ., 2014. 221 p.).
4. **Seljunin V.I., Hanin G.I.** Lukavaya tsifra [Deceitful figure]. *Novy mir*, 1987, no. 2, pp. 181-201. (In Russ.).
5. **Suslov V.I.** *Izmereniye i analiz regional'noy material'noy proizvodstva* [Measurement and analysis of regional material intensity of production]. Novosibirsk: Nauka Publ., 1982. 191 p. (In Russ.).
6. **Suslov V.I.** *Izmereniye effektov mezhregional'nykh vzaimodeystviy: modeli, metody, rezul'taty* [Measuring the effects of interregional interactions: models, methods, results]. Novosibirsk: Nauka Publ., 1991. 251 p. (In Russ.).
7. **Suslov V.I., Ibragimov N.M., Talysheva L.P., Tsyplakov A.A.** *Ekonometriya: Uchebnik dlya vuzov* [Econometrics: Textbook for universities]. Ed. G.M. Mkrtchyan; National Training Foundation. Novosibirsk: SO RAS Publ., 2005. 744 p. (In Russ.).
8. **Suslov I.P.** *Veroyatnost' v sisteme nauchnykh kategoriy*. In: *Dinamicheskaya i veroyatnostnaya optimizatsiya ekonomiki* [The probability in the system of scientific categories. In: Dynamic optimization and probabilistic economy]. Novosibirsk: Nauka Publ., 1978. (In Russ.).