



**Экономический
и Социальный Совет**

Distr.
GENERAL

ECE/CES/2007/22
2 April 2007

RUSSIAN
Original: ENGLISH

ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ СТАТИСТИЧЕСКАЯ КОМИССИЯ
КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ СТАТИСТИКОВ

Пятьдесят пятая пленарная сессия
Женева, 11-13 июня 2007 года
Пункт 5 предварительной повестки дня

**СЕМИНАР ПО ПОВЫШЕНИЮ ЭФФЕКТИВНОСТИ И ПРОИЗВОДИТЕЛЬНОСТИ
СТАТИСТИЧЕСКИХ УПРАВЛЕНИЙ
ЗАСЕДАНИЕ I**

Плюсы и минусы использования административных данных статистическими
управлениями

Представлено Израилем¹

I. ВВЕДЕНИЕ

1. Благодаря современным технологиям и широкому применению идентификационных номеров в различных базах данных у национальных статистических управлений (НСУ) появился богатый источник сравнительно дешевых данных. К числу потенциально доступных источников данных относятся данные учета населения, а также государственная и частная документация, например документация по подоходному налогу

¹ Настоящий документ подготовлен по просьбе секретариата.

и налогу на добавленную стоимость, выплатам и отчислениям по социальному обеспечению, обучению в школах и университетах, документация учреждений здравоохранения, муниципалитетов, полиции, наряду с такими традиционно используемыми административными данными, как данные по экспорту и импорту. Растущие возможности хранения данных и повышение совместимости различных данных учета открывают новые возможности, пример которых - увязка управленческих и кадровых данных, позволяющая анализировать работу компаний в привязке к данным о персонале (см. Haltiwanger, Lane and Spletzer (2000) и Hamermesh (2007)). К числу будущих и потенциальных источников данных, являющихся побочными продуктами использования современных технологий, относятся административные материалы компаний мобильной связи, которые позволяют отслеживать источник, адресат и продолжительность телефонного звонка, данные о пользовании Интернетом, потреблении электроэнергии и воды, перемещении транспортных средств, регистрируемых ГИС, об уплате сборов за проезд по платным дорогам и т.п.

2. На возможности использования административных данных влияют законодательство, технический прогресс и традиции страны. В настоящем документе авторы иллюстрируют это на примере наиболее знакомой им страны – Израиля. Для того чтобы он соотносился с реальностью как можно большего числа стран, в нем рассматриваются только конкретные административные данные учета, которые ведутся во многих странах, включая данные о заработной плате и занятости, которые работодатели представляют налоговым органам. Такие данные имеются практически во всех тех странах, где подоходный налог взимается с заработной платы у источника. Сопоставление информации об общем фонде заработной платы, взятой из разных источников, позволит продемонстрировать те нестыковки, которые могут возникнуть при одновременном использовании административных данных и данных обследований. Будет показано, что, если не проявлять при оценке достаточную аккуратность, разница между данными о заработной плате из этих двух источников может достигать до 20%. Очевидно, что если сравнительный анализ проводить не по всей совокупности, расхождение может быть еще большим. Мы считаем, что эти выводы действительны и для других стран.

3. Цель настоящего документа в том, чтобы уточнить потенциальные возможности использования административных данных и предложить некоторые рецепты для устранения его минусов.

II. ПЛЮСЫ ИСПОЛЬЗОВАНИЯ АДМИНИСТРАТИВНЫХ ДАННЫХ

4. Административные данные способны служить быстрым и сравнительно дешевым источником данных, настоящей "золотой жилой" для статистических управлений и всех тех, кто изучает и фиксирует общественные процессы.
5. Административные данные, как правило, относятся к данным переписного типа, т.е. охватывают всю совокупность, к которой они имеют отношение (но необязательно всю совокупность, которая представляет интерес). Это позволяет статистическим управлениям охватить своей отчетностью даже небольшие контингенты. Кроме того, становится не так важно тщательно формировать выборки с целью снижения расходов и тем самым меньше становится необходимость привлечения дефицитных специалистов по составлению выборок, расчета среднеквадратических ошибок и т.д.
6. Административные данные могут заменить данные обследований и тем самым частично снять с граждан бремя участия в них, а со статистических органов - нагрузку, связанную с прямым сбором данных. Иногда административные источники могут дополнять традиционный сбор данных с помощью обследований. Вопрос о том, способны ли они заменить или лишь дополнить другие источники данных, должен отдельно решаться применительно к каждой области, каждому виду данных и каждой стране.

III. МИНУСЫ ИСПОЛЬЗОВАНИЯ АДМИНИСТРАТИВНЫХ ДАННЫХ

7. В демократическом обществе широкое использование административных данных может быть чревато опасностями, вызывая серьезные вопросы, касающиеся частной жизни и уважения прав человека. Сбор данных является побочным продуктом других видов деятельности, некоторые из которых осуществляются государством без согласия граждан на то, чтобы собранные данные использовались или хранились в статистических целях. Короче говоря, этот дешевый источник данных таит в себе две угрозы. С одной стороны, существует опасность того, что Оруэллов "старший брат" станет реальностью в лице статистических бюро, способных следить за каждым шагом в обществе каждого человека. С другой стороны, сам факт существования источника дешевых данных может побуждать статистические управления собирать данные, которые на самом деле не нужны в силу небольшой ценности этих источников. Например, многие органы составляют или получают адреса граждан из систем данных учета населения. Сбор адресов у этих органов не дает "новой" информации, а лишь является пустой тратой государственных ресурсов. Поэтому, прежде чем использовать тот или иной источник данных, необходимо сопоставлять затраты и результаты.

8. С точки зрения статистических управлений, административные источники данных могут иметь небольшую ценность, поскольку обычно они в первую очередь служат инструментом решения задач органов, в чьем ведении они находятся. Кроме того, определение и содержание данных могут меняться без предварительного уведомления и без переходного периода, в течение которого старые и новые методики использовались бы параллельно. В результате данные до и после изменений никак не связаны между собой, что не позволяет оценить реальные последствия изменений. Особенно серьезные трудности это создает при использовании постоянно обновляемых данных учета, таких, как регистры. Следует подчеркнуть, что подобные проблемы возникают в связи с тем, что органы, готовящие данные, и статистические управления преследуют разные цели. В отличие от таких органов, для которых значение имеет лишь текущее использование данных, статистические управления интересуют и их история. Такое столкновение интересов является одним из источников возникающих проблем.

9. В известном смысле странно было бы ожидать другого: мы уже говорили, что административные источники содержат лишь те данные, которые требуются собирающему их учреждению для своей деятельности. Поскольку статистические управления чаще всего рассматриваются в качестве "прочих пользователей", при планировании работы по сбору данных их потребности не учитываются, что не позволяет впоследствии скорректировать массив данных с учетом их интересов. В этом смысле у потребителей административных данных имеется небогатый выбор: использовать данные в том виде, в котором они существуют, или не использовать.

10. Поскольку административные данные фиксируют реальные события, в них не следует искать ответы на субъективные и гипотетические вопросы, которые могут интересовать статистические управления. Ответы на вопросы типа "что, если" важны для тех, кто хочет смоделировать реакцию изучаемого контингента на предлагаемые изменения.

IV. ДРУГИЕ СВОЙСТВА

11. Некоторые свойства можно занести как в плюсы, так и в минусы в зависимости от того, какое значение придавать их сравнительным достоинствам и недостаткам. Преимуществом административных данных является их охват: они охватывают всю генеральную совокупность или, по крайней мере, всю совокупность в рамках той или иной классификации, в то время как большинство выборок, которыми оперируют статистические органы, охватывают не более одного процента совокупности. В то же время именно это свойство повышает риск незаконного проникновения в систему

статистического управления для получения данных о конкретном физическом или юридическом лице. Поэтому для защиты системы от хакеров и несанкционированного входа в нее сотрудников требуются специальные меры безопасности.

12. На сегодняшний день административные данные, как правило, становятся доступными НСУ после того, как собиравший их орган выполнил свою задачу. Это означает, что они получают их в свое распоряжение через один-два года. Такие задержки не способствуют замене обследований административными источниками. В то же время это соображение может стать не столь важным с развитием методики сбора административных данных в интерактивном режиме и через Интернет.

13. **Вывод:** сопоставление этих достоинств и недостатков, с учетом прогресса в методах сбора и передачи данных, неизбежно приводит к выводу о том, что от использования административных данных не следует отказываться. В то же время очевидно, что они не могут удовлетворить все информационные потребности современного общества, в первую очередь потому, что они вряд ли помогут дать ответы на субъективные и гипотетические вопросы, без которых нельзя анализировать реакцию населения на различные политические меры. Поэтому они должны сосуществовать с данными обследований. В связи с наличием двух источников данных встает вопрос их совместимости, которому посвящена остальная часть документа.

V. ИНТЕГРАЦИЯ АДМИНИСТРАТИВНЫХ ДАННЫХ И ДАННЫХ ОБСЛЕДОВАНИЙ

14. Преобладающий взгляд на сопоставимость административных данных и данных обследований изложен в замечательной работе Каптейна и Ипмы (2005). Они доказывают, что при сопоставлении данных обследований с административными данными последние, как правило, принимаются за "истину" (Kapteyn and YPMA, 2005; Huynh et. al., 2002; Hotz and Schols, 2004). Эта вера основана на том, что, поскольку административные данные предназначаются для собирающих их административных органов, они тщательно проверяются на предмет возможных ошибок. В то же время Каптейн и Ипма утверждают, что этот довод еще не позволяет считать административные данные единственно верными. На качестве данных может отражаться предвзятость собирающего их органа. Например, налоговые органы вряд ли будут уделять большое внимание доходам из не облагаемых налогом источников, даже если информация о них должна сообщаться.

15. При использовании административных данных наряду с данными обследований может возникать противоречивая картина, поскольку эти данные берутся из разных источников, которые могут быть не совсем объективными. Существуют различия в

определениях, временной структуре и охвате, а использование разных источников информации, методов ее сбора и обработки и вызывает возможность разных видов ошибок.

16. Административные данные собираются органами (государственными или частными) для собственного пользования. Поэтому те сами решают, что и как им собирать. В результате сбор данных может неожиданно прекратиться в силу изменения законодательства или приоритетов собирающего данные органа. Кроме того, эти органы не обязаны заблаговременно уведомлять НСУ о предстоящих изменениях или в течение определенного периода вести параллельный сбор данных двумя методами. Все это чревато рядом последствий.

17. При отсутствии соглашения между собирающим данные органом и статистическим управлением административные данные могут не представлять большой ценности для использования во временных рядах. Статистическое управление должно уведомляться об изменениях в приоритетах или в законодательстве, которые могут повлиять на характер данных, предоставляемых статистическому управлению. Здесь можно привести следующий пример. В Израиле Национальный институт страхования представляло данные о наемных работниках, чья заработная плата не достигает 50% от среднего уровня. В начале 2005 года законодательство изменилось: если раньше льготной ставкой налога облагались доходы, не достигающие 50% среднего уровня, то теперь эта планка была поднята до 60%. Израильское центральное статистическое управление (ИЦСУ), не зная о произошедших изменениях, продолжало публиковать эти данные, не скорректировав их определений, что вызвало обеспокоенность новой "тенденцией" в экономике – ростом числа низкооплачиваемых работников. Прошло несколько месяцев, прежде чем была установлена причина. Однако даже после этого сомкнуть "разорванный" ряд оказалось невозможно.

VI. ПРАКТИЧЕСКОЕ ИССЛЕДОВАНИЕ – ОБЩИЙ ФОНД ЗАРАБОТНОЙ ПЛАТЫ

18. Когда для получения сходных данных используются два разных источника, встает вопрос о сопоставимости. В рассматриваемом примере этими двумя источниками являются административные данные и данные обследований. Цель этого примера – показать, как различия в определениях и в процедурах сбора данных могут привести к появлению совершенно разных данных о фонде заработной платы, что может совершенно неоправданно ослабить доверие к результатам обследований и запятнать репутацию НСУ.

19. На протяжении более тридцати лет ИЦСУ пользовалось двумя источниками данных для расчета общего фонда заработной платы: результатами обследований, основанных на отчетах работодателей о кадровой ситуации, представляемых в Национальное бюро страхования (далее А - административный источник), и информацией о доходах, собираемой в ходе обследований рабочей силы (или ОРС, далее обозначается буквой О).

20. Сопоставление двух источников данных осложняется следующими обстоятельствами:

а) **Генеральная совокупность и охват:** Объектом отчетности работодателей являются занимаемые работниками должности, а ОРС – работники. Соответствие между работниками и занимаемыми ими должностями может быть неоднозначным, так как отдельные работники могут занимать сразу несколько должностей. Кроме того, отчеты о кадровой ситуации составляются на момент производства оплаты. Это вовсе не означает, что реальная работа была проделана в данном месяце или даже финансовом году. Когда оплата производится за результаты работы в прошлый период, дополнительные выплаты могут быть перечислены через шесть-семь месяцев после прекращения трудовых отношений. Наконец, источник О охватывает не все население страны. Неохваченными остаются небольшие деревни, на жителей которых приходится около 7% общей численности населения.

б) **Время представления и базовый период отчетности:** работодатели сообщают данные ежемесячно за предыдущий месячный период, а респонденты ОРС – за период, предшествовавший посещению счетчика. Такие посещения проводятся в течение года через равные промежутки времени. Таким образом, участники ОРС сообщают данные не за финансовый год, а за годичный период, завершившийся за месяц до посещения счетчика. Это означает, что в ходе обследований доходов за последний год за базу берется 23-месячный период, где каждому месяцу присваиваются разные веса. В 1986 году продолжительность отчетного периода была уменьшена с одного года до трех месяцев, в результате чего обследованием оказались охвачены 15 месяцев, что показано на следующей схеме:

1	2	3	4	5	6	7	8	9	10	11	12	1											
	2											2											
		3											3										
			4											4									
				5											5								
					6											6							
						7											7						
							8											8					
								9											9				
									10											10			
										11											11		
											12	1	2	3	4	5	6	7	8	9	10	11	12

c) В ходе ОРС респонденту задается вопрос, работал ли он в течение последнего месяца. Затем его спрашивают о том, какое время он работал в каждый из последних трех месяцев. В то же время в публикуемый отчет о результатах обследования идут лишь сведения о том, работал ли респондент в последний месяц, и о его доходе за последние три месяца.

d) Совместимость источников: индивидуальные данные несопоставимы, поскольку ни в отчете о результатах ОРС, ни в административном источнике не указываются персональные идентификационные номера. Поэтому сравнительный анализ приходится проводить на основе сгруппированных данных.

21. В 1970-1980 годы сравнение данных из двух источников проводилось несколько раз². Было установлено, что по сравнению с О общий фонд заработной платы по данным А примерно на 15% выше. Большинство специалистов сошлись на том, что данные О являются заниженными. В результате, когда аналитикам или правительственным чиновникам не нравятся показатели бедности или другие опубликованные данные, они объясняют это тем, что эти данные являются заниженными.

22. В числе причин такого занижения данных называлось уклонение от уплаты налогов. Необходимо отметить, что такое объяснение нельзя воспринимать серьезно. Вопрос не в

² См. Special publication Series no. 593, 1973, p. 20 (проведенное в 1971 году обследование доходов работников) или Monthly Statistical Report, Appendix, 29, June 1978, p. 35.

том, существует ли практика уклонения от уплаты налогов на трудовые доходы, а в том, применим ли данный аргумент к заработной плате. Данные о заработной плате, облагаемой налогом у источника, по определению, сообщаются налоговым органам. Поэтому нет никакого смысла скрывать от счетчика сведения о заработной плате ради того, чтобы избежать уплаты налогов.

VII. СРАВНИТЕЛЬНЫЙ АНАЛИЗ НА ТЕОРЕТИЧЕСКОМ УРОВНЕ

23. Поскольку сравнительный анализ индивидуальных данных о заработной плате не представляется возможным, приходится сопоставлять сгруппированные данные. В то же время существует еще один источник административных данных – отчетность работодателей об удержании налогов, представляемая в конце года. Эта отчетность охватывает весь год, составляется в помесечной разбивке по должностям и содержит личные идентификационные номера, что позволяет использовать ее для имитации как А, административных данных, представляемых работодателями, так и О, данных обследований. Поскольку речь идет об одних и тех же данных, различия во временном охвате, используемых выборках и определениях никак не отражаются на выводах. Опираясь на них, можно получить представление о том, какими могут быть расхождения между двумя источниками, связанные с методикой сбора данных.

24. Как отмечалось выше, публикуемые административные данные представляются ежемесячно на протяжении всего года, в то время как обследования, в ходе которых опрашиваемые домохозяйства сообщают о своей трудовой деятельности за месяцы, предшествующие посещению счетчика, проводятся через равные интервалы времени. Лучше понять разницу между двумя методиками сбора данных можно на примере человека, работавшего один месяц в году. Вероятность того, что этот человек будет учтен в административном источнике данных (А) как работавший один месяц в году, равняется единице. Для обследования О такая вероятность составляет 1/12, поскольку человек будет отнесен к числу занятых лишь в том случае, если счетчик посетит его в том месяце, когда он работал³. Таким образом, результаты обследования О можно рассматривать как **стратифицированную** выборку работавших в течение года, вероятность отнесения которых к категории занятых равняется числу отработанных ими в течение года месяцев, деленному на 12. Например, для лица, работавшего круглый год, вероятность отнесения к числу занятых составляет единицу, а для человека, проработавшего шесть месяцев, она равняется 1/2.

³ Для упрощения анализа возможность занятия трудовой деятельностью в течение нескольких месяцев не учитывается. Общая идея останется той же, хотя цифры могут измениться.

25. Для анализа того, как такие различия отражаются на сообщаемых данных о рабочей силе, в таблице 1 приводятся данные о ежемесячных удержаниях подоходного налога работников с разбивкой по числу отработанных месяцев. Эта таблица позволяет сымитировать административные данные и данные обследований и оценить различия в результатах, вызванные на первый взгляд малозначащей разницей в методах сбора данных.

26. Ниже приводится описание столбцов таблицы 1, содержащих основные данные, необходимые для проведения сравнительного анализа:

Столбец [1] содержит данные о количестве отработанных в году месяцев⁴. В каждой строке показано число лиц и их заработная плата в зависимости от количества месяцев, отработанных за год.

Столбец [2] показывает число лиц, работавших соответствующее количество месяцев. Если человек работал круглый год, он учитывается в последней строке. Примерно 126 000 человек отработали всего лишь один месяц. Итог в этом столбце - число отдельных лиц, занимавшихся трудовой деятельностью по крайней мере один месяц в году. Число 2 598 000 означает **годовую численность занятых**.

Столбец [3] показывает число должностей, которые занимали в каждом месяце лица, учтенные в годовой численности занятых [2]. Количество должностей всегда больше или равно числу работников, умноженному на количество отработанных месяцев.

Столбец [4] показывает число должностей в расчете на одного работника. Как видно из таблицы, количество дополнительных должностей составляет примерно 9%. Эту цифру можно сопоставить с результатами ОРС, в ходе которых 9% опрошенных сообщили, что они работают на двух работах⁵.

⁴ Для упрощения анализа предполагается, что период занятости непрерывен, т.е., если лицо работало три месяца в году, оно работало три месяца подряд.

⁵ В двух источниках используются разные определения дополнительной должности. В административных источниках должность является дополнительной в том случае, когда два работодателя сообщают, что они платят заработную плату одному и тому же лицу. В ходе обследований это понятие определяет сам респондент. Работник, поменявший работу в середине месяца, по данным административного источника имеет дополнительную должность, а по данным ОРС – нет. Кроме того, работник может сообщить, что он работает на двух работах, однако, если заработную плату он получает в одном месте, в административном источнике он будет указан как занимающий одну должность.

Столбец [5] отражает число лиц, которые были бы отнесены к категории занятых по итогам обследования О, т.е. лиц, которые ответили бы, что они работали в предыдущем месяце. Вероятность получения положительного ответа от лица, работавшего один месяц, составляет 1/12, два месяца – 2/12 и т.д. Определяя данные колонки [2] как общую численность каждой группы работников, можно получить указанное в настоящей колонке число, умножая численность каждой группы на вероятность получения положительного ответа.

Столбец [6] содержит сведения об общей сумме месячной заработной платы (в новых израильских шекелях (НИШ)), заработной за год каждой группой работников. В шести колонках содержатся все данные, необходимые для теоретического сопоставления двух источников.

VIII. РЕЗУЛЬТАТЫ В ОТНОШЕНИИ ЧИСЛЕННОСТИ ЗАНЯТЫХ

27. Сопоставление данных в столбцах [2] и [5] позволяет увидеть разницу между годовой численностью занятых (лиц, работавших не менее одного месяца в году) и месячной численностью занятых лиц, работавших в месяце, предшествующем посещению счетчика). Разница между годовой (2,60 млн. человек) и месячной (2,06 млн. человек) численностью занятых составляет 26%. Деление общего числа должностей в каждом месяце на 12 ($26\ 829/12$) дает 2,24 млн. должностей ежемесячно. Таким образом, количество занимаемых должностей примерно на 9% превышает численность работников, рассчитанную по методике ОРС.

Таблица 1: Месячные данные о должностях, численности занятых и заработной плате в 2003 финансовом году

Число месяцев	Годовая численность занятых	Месячное количество должностей (тыс.)	Дополнительные должности	Месячная численность занятых	Общая сумма заработной платы (млн. новых израильских шекелей)
[1]	[2]	[3]	[4] =[3]/[2]	[5] =[2]*([1]/12)	[6]
Всего	2 597 858	26 829,2	1,09	2 057 388	<u>187 571</u>
1,00	125 769	128,4	1,02	10 481	329
2,00	104 556	216,7	1,04	17 426	584
3,00	91 095	285,8	1,05	22 774	820
4,00	93 149	392,8	1,05	31 050	1 257
5,00	81 680	432,6	1,06	34 033	1 443
6,00	86 358	551,4	1,06	43 179	1 982
7,00	83 432	625,5	1,07	48 669	2 327
8,00	94 604	817,3	1,08	63 069	3 333
9,00	90 917	886,1	1,08	68 188	3 634
10,00	108 087	1 176,7	1,09	90 073	5 067
11,00	117 163	1 418,2	1,10	107 399	5 997
12,00	1 521 048	19 897,9	1,09	1 521 048	160 799

28. **Сопоставление средней заработной платы:** Следует отметить, что поскольку для анализа различий между источниками А и О используются одни и те же данные, любая несогласованность должна быть связана с математической или логической ошибкой. В то же время ошибки или логические просчеты совершаются или могут совершаться в том случае, если игнорируются различия в базовом периоде: обследование О стратифицируется по продолжительности отработанного времени, с которой тесно коррелирует размер ежемесячной заработной платы (см. таблицу 2).

Таблица 2: Средняя заработная плата в расчете на одну должность и одного работника
(в новых израильских шекелях), 2003 год

	Число отработанных месяцев	Средняя заработная плата в расчете на одну должность	Средняя заработная плата в расчете на одного работника
	[1]	[2]	[3]
Среднее значение		6 991	7 597
	1	2 565	2 619
	2	2 697	2 794
	3	2 869	3 000
	4	3 199	3 372
	5	3 336	3 533
	6	3 594	3 825
	7	3 720	3 984
	8	4 078	4 404
	9	4 101	4 441
	10	4 306	4 688
	11	4 228	4 653
	12	8 081	8 810

29. Данные таблицы 2 рассчитаны на основе таблицы 1. В столбце [2] показана месячная заработная плата в расчете на одну должность (столбец [6]/столбец [3] в таблице 1), в то время как в столбце [3] показана месячная заработная плата одного работника { столбец [6]/ столбец [3]* столбец [5] }. Как и ожидалось, чем больше стаж работы, тем выше помесечная заработная плата (за исключением тех, кто проработал 11 месяцев). Наибольшей является разница между теми, кто проработал целый год, и остальными работниками. Как видно из таблицы 2, средний размер оплаты труда при работе в течение 12 месяцев по данным источника О составляет 8 810, а средняя заработная плата в расчете на одну должность по данным источника А – 8 081, т.е. на 9% меньше. Среднемесячная заработная плата в расчете на одну должность составляет 6 991, а в расчете на одного работника – 7 597. Разница опять же равняется 9%.

30. Ниже описывается приемлемый способ расчета годового фонда заработной платы. Для источника О следует умножить средний размер заработной платы сотрудника, работавшего в последний месяц (7 597) на среднемесячную численность работников (2,06 млн.)*12. Что касается источника А, общее месячное число должностей умножается

на среднемесячную заработную плату в расчете на одну должность. Эти две альтернативные методики расчета дают один и тот же показатель общего фонда заработной платы.

31. Вышеуказанная методика, при использовании которой во внимание принимается лишь заработная плата за месяц, предшествующий посещению счетчика, и игнорируется зарплата, полученная за год, представляется неэффективной, что заставляет искать альтернативные методы, способные повысить качество оценок. К сожалению, это не гарантирует полную объективность результатов. В этой связи мы воспроизводим сопоставления двух расчетных показателей фонда заработной платы, сделанные в 70-х годах (см. сноску 2), но без учета дополнительной налоговой отчетности по состоянию на конец года, позволившей нам воспроизвести оба обследования. В силу невозможности сопоставить индивидуальные данные сравнительный анализ был проведен с использованием двух выборок.

32. В таблице 3 приводятся данные о должностях/сотрудниках, среднем размере заработной платы и расчетном фонде заработной платы, полученные на основе двух методик, а также реальные данные из источников А и О.

33. **Первая строка** таблицы 3 содержит расчетные данные, имитирующие источник А. Для расчета среднемесячного числа должностей (2 235) общее число должностей в каждом месяце отчетного года делится на 12. Общий фонд заработной платы делится на среднемесячное число должностей и умножается на 12. Данные на этой строке считаются "достоверным" результатом.

34. **Вторая строка** показывает фонд заработной платы, рассчитываемый путем умножения месячной заработной платы одного работника на среднюю зарплату сотрудников, работавших в прошлом месяце (умножено на 12). Это является имитацией правильного способа расчета фонда заработной платы на основе данных обследования О.

35. **Третья строка** содержит данные, полученные в результате воспроизведения расчетов 70-х годов. За численность занятых принималось количество лиц, работавших в предыдущем месяце, а за доход – доход в предыдущем году. Таким образом, поскольку в основе результатов ОРС лежат сведения о лицах, работавших в предыдущем месяце, и о заработной плате за прошлый год (с 1986 года – за последние три месяца), логично было бы рассчитывать фонд заработной платы как сумму зарплат всех лиц, работавших в предыдущем месяце. Деление годовой заработной платы на 12 дает месячную заработную плату в размере 7 257 новых шекелей. Эта среднемесячная зарплата на 4% превышает размер месячной заработной платы в расчете на одну должность. Умножив ее

на количество лиц, работавших в последний месяц, получаем расчетное значение годового фонда заработной платы, составляющее 179 млрд. Рассчитанный таким образом фонд заработной платы на 5% меньше "истинного" годового фонда заработной платы. Данные расчеты построены на методике сравнительного анализа, проведенного в 70-х годах, по итогам которого был сделан вывод о том, что данные источника О являются заниженными. В то же время разница, рассчитанная в ходе того сравнительного анализа, в основе которого лежали реальные данные, превышает ту разницу, которая была получена по итогам теоретического сопоставления (14%)⁶.

Таблица 3: Расчетный фонд заработной платы (в новых израильских шекелях), 2003 год

	Должности или месячная численность работников (в тыс. единиц или человек)	Средняя заработная плата	Фонд заработной платы (в млрд. шекелей в год)
(1) Источник А: должности	236 2	991 6	188
(2) Источник О: месячная численность работников	057 2	597 7	188
(3) Источник О: годовой доход	057 2	257 7	179
Отношение (3/1)	1,09	0,96	1,05
(4) Выполняемая сотрудниками работа (должности)	2 340	6 972	196
(5) Число занятых в среднем один месяц	1 946*	1 712*	166
(6) Число занятых с поправкой на охват	2 082	7 112	178
(7) Число занятых	1 946*	6 908**	161
Отношение (4/1)	1,04	0,997	1,04
Отношение (6/1)	1,16	1,01	1,18

* Эти данные отражают численность занятых в месяце, предшествующем посещению счетчика, а не средние данные за три месяца, публикуемые ИЦСУ.

** Средние данные за три месяца, публикуемые ИЦСУ.

⁶ При сопоставлении реальных данных необходимо также учитывать различный охват отдельных источников. Этот вопрос будет рассматриваться позднее.

36. В четвертой и пятой строках приводятся фактические данные, полученные ИЦСУ из источников А и О. Фактическое месячное число должностей составляет 2 340, а количество занятых в последнем месяце - 1 946 человек. Однако, поскольку обследование охватывает лишь 93% населения, в качестве первого приближения было бы правильным умножить это число на 1,07. Такие скорректированные данные приводятся на шестой строке. Общий фонд заработной платы составляет 178 млрд., что на 6% меньше его гипотетической величины. Среднемесячная численность должностей по данным источника А (четвертая строка) составляет 2,34 млн., что на 4,6% выше "истинного" расчетного показателя⁷. В результате можно говорить о том, что расчетная величина фонда заработной платы завышена на 4%.

37. По данным в седьмой строке можно судить о тех ошибках, которые возникают в случае использования нескорректированных данных О, т.е. простой имитации методики, применявшейся в 1970-х годах. Такая ошибка в размере фонда заработной платы, приближающаяся к 15% (161/188), идентична той процентной ошибке, которая была характерна для расчетов, проводившихся в 1970-х годах.

38. Этот вывод в отношении качества данных источника О подтверждаются и в работе Романова и Фурмана (Romanov and Furman, 2006), которые при сопоставлении административных данных и данных обследований, содержащих персональные идентификационные номера, не обнаружили сколь-либо серьезного занижения результатов.

IX. ВЫВОД

39. Цель настоящего документа заключается в том, чтобы описать те проблемы, которые встают перед национальными статистическими управлениями в связи с использованием административных источников данных в сочетании с данными обследований. Получение совершенно разных результатов на основе двух источников внешне схожих данных в силу недооценки возможных различий может подорвать доверие к результатам обследований и запятнать репутацию национальных статистических ведомств. Тенденция публикации микроэкономических данных лишь усугубляет эту проблему, вводя неискушенных исследователей в заблуждение относительно качества данных.

⁷ Хотя в основе данных на первой и седьмой строках лежат административные, налоговые источники, расхождение отчасти можно объяснить различиями в используемых предприятиями методах составления ежемесячной и годовой отчетности. Помимо этого существуют и обычные ошибки, присущие выборочному методу обследования.

ЛИТЕРАТУРА

Furman, O. (2005). Comparative analysis of wages and work indicators from the income tax records of wage earners. CBS, CBS, Technical Paper no. 14.

Haltiwanger, J. C., J. I. Lane and J. R. Spletzer (2000). Wages, Productivity, and the Dynamic Interaction of Businesses and Workers, Working Paper No. 7994, National Bureau of Economic Research, Cambridge, MA.

Hamemesh, D. S. (2007). Fun with Matched Firm-Employee data: Progress and Road Maps, IZA Discussion Paper no. 2580, (January).

Hotz, V. J., and J. K. Scholz (2004). Measuring Employment and Income for Low-Income Populations with Administrative and Survey Data.
http://www.econ.wisc.edu/~scholz/Research/NRC_Income_Measurement_Paper_Final_Draft.pdf

Huynh, M, K. Rupp, J. Sears, (2002). The Assessment of Survey of Income and Program Participation (SIPP) Benefit Data using Longitudinal Administrative Records, Working Paper no. 238, Office of Research, Evaluation, and Statistics, Social Security Administration, Washington, DC: US Census Bureau, 2002

Kapteyn, A. and J. Y. YPMA (2005). Measurement Error and Misclassification (A comparison of Survey and Register Data), Working Paper, Rand, WR-283, (July).

Romanov, D. and O. Furman (2006). Analysis of wage data from the 1995 census by using wage file of the National Insurance Institute, CBS, Working Paper no.
