**UNITED NATIONS** **E**

# Economic and Social Council

Distr.
GENERAL

ECE/CES/2007/22
2 April 2007

Original: ENGLISH

**ECONOMIC COMMISSION FOR EUROPE** **STATISTICAL COMMISSION**

CONFERENCE OF EUROPEAN STATISTICIANS

Fifty-fifth plenary session
Geneva, 11-13 June 2007
Item 5 of the provisional agenda

SEMINAR ON INCREASING THE EFFICIENCY AND PRODUCTIVITY OF STATISTICAL OFFICES
SESSION II

# Pros and Cons for using Administrative records in Statistical Bureaus

Submitted by Israel[1]

## INTRODUCTION

1. Modern technologies and the wide use of identity numbers in the different databases have created a rich and a relatively cheap source of data that can be available to National Statistical Offices (NSO). The list of files that are potentially available includes a population register, and records from governmental and private activities such as income and value added tax, social welfare and social security, schooling and universities, health, municipalities, police, in addition to the traditionally used administrative records like import and export data. The growing storage capacity of data and the improved record linkage methodology have opened the way to new possibilities as exemplified in matched employer-employee records that enables the analysis of firms' performances together with the characteristics of the employees. (See Haltiwanger, Lane and Spletzer (2000) and Hamermesh (2007)). Future and potential sources, which are by-products of the use of modern technology, include non-governmental administrative records of cell-phone companies that can be used to track the origin, destination and length of phone calls,

[1] This paper has been prepared at the invitation of the secretariat.

internet surfing, electricity and water usage, car-movements generated by a GIS system and taxes paid on crossing road-junctions etc.

2. The potential use of administrative files is affected by the legislation, the technological progress, and the tradition in the country. The illustration in this paper considers mainly the environment in the country we are most familiar with – Israel. In order to make it relevant for a wide range of countries, it will be restricted to a specific administrative file that exists in many countries and includes data on wages and employment reported by the employers to the tax authorities. These records exist almost in all countries that withhold income tax on wages at source. A comparison between the total wages bills derived from different sources will illustrate the potential confusion that may occur when using administrative and survey data simultaneously. As will be shown, if one is not careful enough, differences up to a range of 20 percent can be found between the two sources of wage bill estimators. Clearly, if one restricts the comparison to sub-populations, difference can be of a larger magnitude. As far as we can see, those conclusions are relevant for other countries too.

3. The aim of this paper is to throw light on the properties of the potential use of administrative data, and to offer some remedies to the downside of using them.

## II. THE BENEFITS OF USING ADMINISTRATIVE RECORDS

4. Administrative records present a potential source of rich and relatively cheap data that can be considered as a gold mine for statistical bureaus as well as for anyone who is interested in researching and documenting social behavior.

5. In general, administrative files are of census type. That is, they cover the whole population that is covered by them. (They do not necessarily cover the whole population of interest). As a result, the statistical bureau can extend its reporting to cover small populations. Moreover, a careful design of samples intended to reduce the costs becomes less essential and consequently, the need for scarce skills of sample design, evaluation of standard errors etc. is reduced.

6. Administrative records can substitute for survey data and by doing so, reduce response burden on the citizens, and relief the statistical offices' workload of direct data collection. At times, administrative records can be complementary to the traditional survey-based data collection. Whether they can substitute or be complementary is an issue to be examined in each field, for each file and for each country.

## III. THE DOWNSIDE OF USING ADMINISTRATIVE RECORDS

7. The extensive use of administrative records may present a dangerous path to follow for a democracy, because it raises serious questions concerning privacy and human rights issues. The data are produced as a by-product of other activities, some of them imposed on the citizen by the government, without the person being asked about his/her consent to use and store data for statistical purposes. In short, the cheap data source creates two kinds of risk. On one hand, there is a danger of creating a base for Orwell's "big brother" reality, dominated by the statistical bureaus, having the ability to trace the activity of each individual in the society. On the other hand, the mere existence of a cheap data source may tempt statistical bureaus to collect data that

are not really needed because the value added of those sources may be small. For example, many agencies adjust or derive the addresses of citizens from the population registrer. Collecting the addresses from those agencies does not add "new" information. This means a waste of public resources and calls for a cost-benefit analysis of utilizing each data source.

8. Administrative records may be of little value from a statistical bureau's point of view, since they tend to be biased according to the priorities and policies of the agency in charge. Moreover, changes of definitions and content can occur without prior notice and without having a grace period in which the new and the old definitions are reported simultaneously. As a result, there is no linkage between the data before and after the changes and it is impossible to evaluate the effect of the real change that occurred. It is especially problematic when using continuously updated file, like a register. It should be stressed that the different targets of the agency producing the data and of the statistical bureaus cause these kinds of problems. While the agency is interested in the present use of the file, the statistical bureau is also interested in the history of the file. This clash in interests is one of the sources of such problems.

9. In some sense, one gets what he has paid for. Our assumption is that administrative records include only variables of interest that the collecting agency needs in order to carry on its business. In most cases, statistical bureaus are considered as "residual users" of administrative records, their needs are not taken into account in the designing phase and the flexibility to accommodate a dataset to their interests is lost. In some sense, the use of administrative data is a "take it or leave it" decision.

10. Since the administrative data consists of reports on actual transactions, it is not expected to include answers to subjective and hypothetical questions that may be of interest to the statistical bureau. Answers to "what if" questions are important whenever one wants to model the reaction of the population to proposed changes.

## IV. OTHER PROPERTIES

11. Some properties can be listed on both sides of the balance sheet depending on the relative merits and costs associated with the property. The coverage advantage of the administrative records is that they include all the population or at least all the population under a given classification, compared with most samples used by statistical bureaus that do not cover more than one percent of the population. However, this very property increases the risk that someone, interested in obtaining data on a specific person or entity will try to illegally penetrate into the system of a statistical bureau to obtain access to the data. Hence, special security measures have to be used in protecting the system from hackers and unauthorized employees.

12. As of today, administrative records tend to be available to the NSO only after the collecting agency has completed its task. It implies a delay of a year or two in the availability of the data. Therefore, this timing issue may discourage the substitution of survey data by administrative records. However, the spread of on-line and Internet collection of administrative data may reduce the effect of this consideration.

13. **Conclusion**: Comparing the list of costs and benefits, the progress of data collection and transmission, leads to the inevitable conclusion that administrative data should be utilized. However, it is also clear that they do not substitute for all the data that a modern society needs, especially because they cannot be expected to answer the subjective and hypothetical questions that are needed to evaluate the population response to different policy measures. Therefore, they should coexist with the survey data. Having two sources of data raises an issue of comparability, dealt in the rest of this paper.

## V. INTEGRATION OF ADMINISTRATIVE AND SURVEY DATA

14. The prevailing opinion concerning the comparison between administrative and survey data is represented in an excellent paper by Kapteyn and YPMA (2005). They forcefully argue that whenever one compares survey data with administrative records, the general tendency is to view the latter as representing the "truth" (Kapteyn and YPMA, 2005; Huynh et. al., 2002; Hotz and Schols, 2004). The reasons for such a belief can be attributed to the fact that administrative data serve the actions of the administration collecting them and therefore are scrutinized for errors. However, Kapteyn and YPMA also argue that this assumption should not be interpreted as the ultimate truth. Biases originated in the agency collecting them can influence the quality of such data. For example, one can expect an income tax administration to pay little attention to non-taxed income sources, even if they have to be reported.

15. The use of administrative records, along with the survey data, can be confusing because they rely on different sources that may be biased. There is a difference in definitions, time structure and time coverage, and since the source of information, data collection and processing are different, they carry different types of errors.

16. Administrative records are collected by an agency (governmental or private establishment) for its own use. As such, the agency is the sole decision-maker concerning what to collect, how to collect, etc. As a result, collection of the data can be abruptly stopped – because of a change in the law, or in the priorities at the collecting agency. Also, the agency does not have an obligation to report the changes to the NSO in advance, or to allow for a period in which two sets of data are collected simultaneously. This may have several implications:

17. Unless there is an agreement between the agency and the statistical bureau – the administrative records may be of a lesser value for the use in time-series. The statistical bureau has to be informed about changes in the priorities or in the law that may affect the nature of the records delivered to the statistical bureau. A case that illustrates this point is of the National Insurance Institute in Israel that submits its records on employees whose posts earn a salary, which is lower than 50 percent of the average wage. In the beginning of 2005 the law that entitles a reduced tax rate to income below 50 percent of the average income has changed to allow a reduced rate to income below 60 percent of the mean wage. The Israeli Central Bureau of Statistics (ICBS), not aware of the change, has continued to publish the data without modifying the definitions – raising concerns about a new "trend" in the economy – of an increase in low wage employees' posts. Several months passed till the reason for the change was detected. However, even when detected, there was no way to prevent a "break" in the series.

## VI. A CASE STUDY – THE TOTAL WAGE BILL

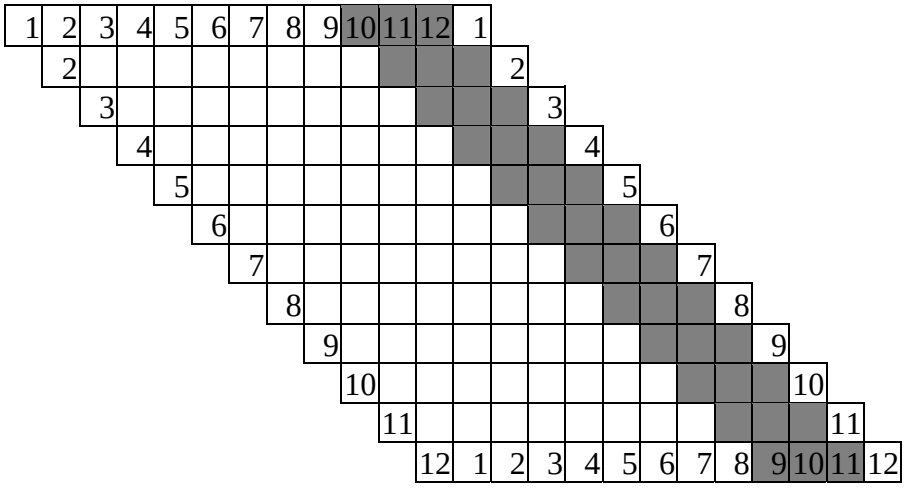
18. Whenever two sources are used to derive similar data, the question of comparability arises. In this case the two sources are administrative and survey data. Our aim is to illustrate how differences in definitions and in collection processes can create totally different figures of the wage bill, a difference that can unjustifiably reduce the trust in the survey data and tarnish the reputation of the NSO.

19. For over 30 years ICBS has been using two sources of data that enable the estimation of the wage bill: A survey based on the reports of employers to the National Insurance Office (**A** - hereafter to denote Administrative source) on posts held by the employees, and an income survey that is based on one wave of the Labor Force Survey (LFS hereafter or **S** to denote Survey data).

20. A comparison between the two sources is difficult due to the following reasons:

(a) **Population and coverage**: Employers report on employees' posts, LFS on a person. There are one-to-many relations between employees and posts; some employees may hold several posts. Also, the report on an employee post occurs whenever a payment is made. This does not necessarily mean that the actual work was done in that month, or even in that fiscal year. Whenever one is also paid according to one's past performance, an additional payment can be made six or seven months after the cessation of the employment relationship. Finally, S source does not cover all the population in the country. Small villages are left out. The group that is not covered is about 7 percent of the overall population.

(b) **Timing and reference time**: Employers report monthly throughout the calendar year, referring to the previous month, while the LFS respondents report on the period prior to the visit of the surveyor. These visits are spread uniformly over the year. Hence, participants in the LFS do not report on a fiscal year but on a moving year – each person is asked about an accounting period that ends up one month prior to the visit of the surveyor. This means that the period covered by the income survey, with a question concerning last year earnings is composed of 23 months with different weights attached to each month. In 1986 the accounting period was changed from a year to 3 months, changing the survey to cover 15 months, as illustrated below:




(c) The question asked in the LFS is whether a person worked in the last month. Then the person is investigated about working in each of the last three months. However, in the Public Use File of the survey the items reported are whether the person worked last month, and what was the income earned in the last three months.

(d) Record-Linkage: Individual records cannot be compared since personal identification numbers (PINs) are not reported in the LFS and in the administrative source. Hence, the comparison must be based on aggregated data.

21. Between 1970 and 1980, several wage bill comparisons between the two sources have been performed[2]. The findings are that the total wage bill according to A is about 15 percent higher than the wage bill according to S. From then on – the prevailing conclusion is that the S data are biased downward. As a result, whenever commentators or government officials do not like poverty figures, or other published figures, they refer to the bias as an explanation.

22. The explanation given for the source of the bias is tax evasion. However, it is worth pointing out that this explanation should not be taken seriously. The issue is not whether tax evasion exists with respect to labor income but whether the argument applies to wages. By definition, wages are reported to tax authorities and the tax is deducted at source. Hence, tax evasion does not create an incentive not to report wages to the surveyor.

## VII. A THEORETICAL EXERCISE OF COMPARISON

23. Since a comparison of wages reported on the individual record level is not feasible, one has to rely on comparisons of aggregated data. However, an additional source of administrative data is available. It is composed of the reports of the employers at the end of the year, concerning withholding of taxes. Those reports cover the whole year, disaggregated according to monthly posts held, and they have PINs, so that one can use them to imitate both, the A data source reported by the employer, and the S data source reported by the survey. Since we rely on

[2] See Special publication Series no. 593, 1973, p. 20 concerning Employees Income Survey 1971, or Monthly Statistical Report, Appendix, 29, June 1978, p.35.

the same data, differences in time periods covered, sampling variability, and different definitions of being employed do not affect our findings. It can give us an idea of the expected differences between the two sources that originate in the methodology of data collection.

24. As mentioned above, the monthly statistics published using the A source, sweeps the data all over the year while the S source surveys the households uniformly over the year, asking about employment status in the months prior to the visit of the surveyor. In order to observe the difference between the two collection methods, let us consider a person who works only one month during the year. Under the A source regime, the probability of that person to be found employed in one month during the year is one. The probability of such a person to be found employed in the S data is 1/12 because only if the surveyor visits in the month of employment, the person will be considered as employed.[3] The indication is that the S survey can be viewed as a **stratified** sample of those employed during the year with the probability of being found as employed that equals the number of months worked during the year, divided by twelve. For example, the probability of those who work all over the year to be found as working is one, while the probability of those working six months during the year is half.

25. To investigate the implication of such a difference on the workforce reported, Table 1 presents tabulation from the withholding tax file of the employees, according to the number of months of employment. This tabulation enables us to imitate the administrative and the survey data, and to examine the gap in the results caused by the seemingly innocent difference in the methods of data collection.

26. Following is a description of the columns in Table 1, which include the basic data needed for our estimates.

**Column [1]** presents the number of months worked during the year.[4] Each row represents the number of persons and their wages, according to the number of months the persons worked during the year.

**Column [2]** presents the number of persons who worked during the related number of months. If a person worked during the whole year, he is counted in the last row. About 126 thousand individuals worked for one month only. The total number in this column is the number of distinct persons who were employees for at least one month during the year. We refer to the number 2,598 thousands as of the **yearly employees**.

**Column [3]** presents the number of monthly posts held by the yearly employees [2]. The number of posts is always equal to or greater than the number of persons multiplied by the months worked.

**Column [4]** presents the number of posts held by an employee. As can be seen the number of additional posts held is about 9 percent of the posts. This number can be compared to the LFS data where 9 percent report the holding of an additional job.[5]

---

[3] To simplify the analysis the issue of asking a question on more months is ignored. The general idea is not affected, although the numbers can change.

[4] To simplify the analysis it is assumed that the employment period is a continuous one, e.g., if a person worked three months during the year, the months of employment were consecutive.

PT[5] TPThe definition of an additional post is different in the two sources. In the A file – a job is an additional post if two employers report paying a salary to the same person. In the S survey it is as defined by the interviewed. An employee who switched jobs in the middle of the month is considered as holding an additional job according to A source, while in the LFS he is not. Also, an employee can report holding to jobs, but if the salary is paid by the same unit – it will be considered as one post in the A source.

**Column [5]** presents the number of persons worked as would be reported by the S source. That is, the expected number of employees when asked whether they were employed in the previous month. The probability of receiving a positive answer from a person who was employed only one month is 1/12, for two months – 2/12 etc. Defining column [2] as the whole population of each group, we obtain an estimate of this number by multiplying the population of each group by the probability of answering positively.
**Column [6]** presents the sum of monthly wages in (New Israeli Shekels (NIS)) earned during the year by each group.
The six columns present all the data needed for the theoretical comparison between the two sources.

## VIII. RESULTS CONCERNING THE NUMBER OF EMPLOYEES

27. Comparison between column [2] and column [5] enables us to see the difference between the yearly employees (who were employed at least one month during the year) and the monthly employees (who were found as being employed in the month prior to the visit of the surveyor). The number of yearly employees is 2.60 million while the number of monthly employees is 2.06 million, making a difference of 26 percent. Dividing the total monthly posts by 12 we get (26,829/12=) 2.24 million monthly posts. The number of employees' posts is about 9 percent higher than the number of employees that should have been reported according to the LFS.

**Table 1:** Monthly positions, Persons Employed and Salaries for Fiscal year 2003

| **No. of Months** [1] | **Yearly Employees** [2] | **Monthly Posts (In** thousands) [3] | **Additional posts** [4] =[3]/[2] | **Monthly Employees** [5] =[2]*([1]/12) | **Total Wages** (In millions of NIS) [6] |
|---|---|---|---|---|---|
| **Total** | **2,597,858** | **26,829.2** | 1.09 | **2,057,388** | U187,571U |
| 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. **Comparisons of Average Wages**: It should be clear that since we are using the same data to produce the difference between the A and the S sources – any inconsistency should be

attributed to a mathematical or a logical mistake. However, those errors or logical mistakes are made or can be made if the fine distinction in the reference time is ignored; The S survey is stratified according to the time-period of being employed, and the wage per month is highly correlated with it, as seen in Table 2.

**Table 2:** Average wage per employee post and per employee (NIS), 2003

| | Months Worked [1] | Average wage per employee post [2] | Average wage per employee [3] |
|---|---|---|---|
| **Average** | | **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. Table 2 entries are derived from Table 1. Column [2] in Table 2 presents the wage per monthly employee's post (column [6]/column [3], according to Table 1), while column [3] presents the wage per monthly employee {Column [6]/ {column [1] *column [5]}. As expected, the longer the employment period, the higher the monthly salary (except for those who worked 11 months). The highest difference is between being employed during the whole year - and the rest. As can be seen in Table 2, the average salary for 12 months according to the S data source is 8,810 while the average salary per employee's post according to the A data source is 8,081 – a difference of 9 percent. Average wage per monthly post is 6,991 while the average wage per employee is 7,597 – the difference is again 9 percent.

30. The appropriate way to estimate the yearly wage bill is as follows. For the S source - multiply the average wage per employee who worked in the last month (7,597) by the average number of monthly employees (2.06 Million) *12. For the A source - multiply the total of monthly posts by the average wage per monthly post. These two alternative ways yield the same total wage bill.

31. The method described above, of using only the salary paid in the month previous to the visit of the surveyor, while ignoring the salary earned during the year, seems inefficient and one is tempted to use alternative methods that may improve the estimate. Unfortunately, it can also lead to biased results. For this purpose, we replicate the comparisons between the

two estimates of the wage bill that were done in the seventies (see footnote 2) but without the additional end-year withholding of tax file that we had, which enabled us to replicate both surveys. The comparisons were performed using the two samples, lacking the ability to compare individual records.

32. Table 3 presents the posts/persons, average wages and wage bill estimators for different types of calculation methods and actual figures reported by the A and S surveys.

33. **The first line** of Table 3 presents the estimate that imitates the A source: In order to get the average monthly posts of 2,235, the total number of monthly posts, reported over the year, is divided by 12. Total wages are divided by the average monthly posts and multiplied by 12. We refer to this line as the "true results".

34. **The second line** derives the wage bill by multiplying the monthly wage per person by the average wage per person employed in the last month (multiplied by 12), imitating the correct way to estimate the wage bill from the S survey.

35. **The third line** replicates the type of estimate that was done in the seventies. The number of employed persons was based on the number of persons who were employed in the previous month. The income is the income earned during the last year. That is, since in the LFS published figures are whether the person was employed in the previous month, and the wages earned over the last year (from 1986 – from the last three months) the logical estimate of the wage bill is to add up the wage bills of all individuals who were employed in the last month. By dividing the yearly salary by 12 we get a monthly salary of 7,257 New Shekels. This average salary per month is greater than the monthly salary per post by 4 percentage points. Multiplying it by the number of persons worked in the last month leads to an estimated yearly wage bill of 179 billions. The total wage bill one gets in this way is 5 percent lower than the true yearly wage bill. This exercise replicates the comparison that was done in the seventies, which led to the conclusion that the S data were biased downward. However, that comparison was based on actual data, and the difference found is higher than the difference we find in the theoretical comparison (14 percent).[6]

[TP6P T]In comparison of real data one also has to take into account the difference in the coverage of the different sources. This point will be discussed later.

**Table 3**: Estimated Wage Bill (NIS), 2003

| | **Posts or monthly employees** (In thousands) | **Average Wage** | **Wage bill** (Billions per year) |
|---|---|---|---|
| **(1) A source: posts** | **2,236** | **6,991** | **188** |
| **(2) S source: monthly Employees** | 2,057 | 7,597 | 188 |
| **(3) S source: yearly income** | 2,057 | 7,257 | 179 |
| **Ratio (3/1)** | 1.09 | 0.96 | 1.05 |
| **(4) Employee job (posts)** | 2,340 | 6,972 | 196 |
| **(5) Employed persons: one month on average** | 1,946* | 7,112* | 166 |
| **(6) Employed Persons-adjusted for coverage** | 2,082 | 7,112 | 178 |
| **(7) Employed persons** | 1,946* | 6,908** | 161 |
| **Ratio (4/1)** | 1.04 | 0.997 | 1.04 |
| **Ratio (6/1)** | 1.16 | 1.01 | 1.18 |

* These figures present those worked in the month prior to the visit of the surveyor, not the three months average published by the ICBS.
** Three months average, published by the ICBS.

36. Lines (4) and (5) present the actual numbers reported by the ICBS from the A and S samples. The actual number of employees' post per month reported is 2,340 while the number of persons who were employees in the last month is 1,946. However, since the survey covers only 93 percent of the population, a crude revision would be to multiply this number by 1.07. This revised figure is reported on line (6). The overall wage bill is 178 billion, which is 6 percent lower than the hypothetical wage bill. The A sample of employees' posts (line 4) reports an average of 2.34 million per month, which is 4.6 higher than our "true" estimate[7]. This brings us to an upward biased estimate of the wage bill of 4 percent.

37. Line (7) reports the kind of bias one gets if one simply takes the S survey without adjustment, replicating the exercise that was done in the seventies. The bias in the wage bill would be near 15 percent (161/188), which is identical to the percentage bias found in the seventies.

38. Additional support for the conclusion concerning the quality of the S data can be found in Romanov and Furman (2006) who compared the administrative data with census data, which include PIN, did not detect a serious downward bias.

[7] Although lines 1 and 4 are based on administrative data from tax authorities, part of the discrepancy between the estimates may be explained by different reporting methods buy the businesses for monthly and annual filings. This is in addition to the normal sampling errors.

## IX. CONCLUSION

39. The aim of this paper is to present the challenge that the use of administrative records combined with survey data can pose to National Statistical Offices. Having two sources, with seemingly related data, that can produce totally different results because of some kind of overlooked differences may damage the credibility of the survey data and tarnish the reputation of the national statistical bureaus. This problem is accentuated with the tendency to release micro data that can bring inexperienced researchers to the wrong conclusions concerning the quality of the data.

## REFERENCES

Furman, O. (2005). Comperative 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.

* * * * *