

CONFERENCE OF EUROPEAN STATISTICIANS

Joint UNECE/EUROSTAT Work Session on Electronic Data Reporting
(Geneva, Switzerland, 13-15 February 2002)

Topic (ii): Security, confidentiality and privacy issues

BUSINESS PERCEPTIONS OF CONFIDENTIALITY

Submitted by United States¹

Invited paper

ABSTRACT

The core mandate of statistical agencies is to collect data--on a multiplicity of entities, including businesses and households. In so doing, each agency enters into an implicit pact with its respondents that the data not only will be used to fulfill its respective mission, but also will be protected from unauthorized access and use. This promise of confidentiality is not only a legal and ethical mandate, but also an important contributor to optimal data quality and response rates. However, little is known about the substance of this pact – particularly with respect to businesses. This ignorance can have serious potential consequences. If government misconstrues its role or the nature of the pact, and consequently businesses do not trust government to protect their data because they either mistrust or misunderstand the pact, it will be difficult for government not only to maintain high quality and timely response rates but also to frame new ideas on data collection, protection, and access. This study provides one of the first quantitative analyses of business perceptions of the sensitivity of different types of data and their assessment of the protection provided by different agencies – as well as their assessment of the quality of statistical work performed by the agencies. It also examines the knowledge of the business community of the financial and criminal penalties associated with breaches of confidentiality and its willingness to permit data to be shared among different types of users.

I. INTRODUCTION

1. Concern about declining response rates on the 1990 decennial Census led to the development of a very interesting literature targeted at understanding the sources of the decline. One entire thread of this literature has examined the role of public perceptions of the degree to which federal statistical agencies protect respondent data, and the link between this and respondent behavior (Singer, 2001). The results of the analysis have been extremely useful in informing statistical agency policy in the United States, and the model has served national statistical institutes abroad as well.

2. The burgeoning interest in perceptions research on individual respondents has not been matched by similar research on business respondents. This paper builds on one of the few studies of businesses (Greenia et al, 2001), and examines the links between business perceptions of statistical agency competence, and data protection and their willingness to provide data to those agencies.

¹ Prepared by Nick Greenia (Statistics of Income Division, Internal Revenue Service); J. Bradford Jensen (Center for Economic Studies, U.S. Bureau of the Census); Julia Lane (American University, The Urban Institute and LEHD program, U.S. Bureau of the Census).

3. The results of the study suggest that there is indeed a link between business perceptions and their willingness to provide data. Although firm conclusions cannot be drawn on the basis of such a preliminary study, if the results were to hold up under a more representative survey, they might well suggest that an initiative by statistical agencies to convey information to their survey population about both how the data are protected and how they are used could improve response rates and data quality.

II. BACKGROUND AND LITERATURE REVIEW

4. The lack of quantitative research does not mean that no attention has been paid to the confidentiality of business data provided for federal statistical purposes. In 1992, the U.S. Office of Management and Budget (OMB) established a working group that not only noted the differences between household and business perceptions of confidentiality, but also identified several factors that were likely to affect business trust in government statistical agencies, and consequently their willingness to provide data. One major factor was thought to be the existence and enforcement of penalties for unauthorized disclosure of information about an economic entity. Another was the use to which data were put – in other words, the benefit provided to policymakers as a result of the business responding to the survey. Finally, the OMB working group was concerned about the burden imposed on businesses by federal statistical agencies.

5. Preliminary work by Willimack and co-authors supported the idea that businesses are primarily concerned about cost associated with data provision. In a series of interviews with some 30 large businesses, they found that business respondents would be willing to trade off privacy concerns and permit data-sharing across statistical agencies, if that would mean that they had to fill out fewer surveys. However, their work did not quantitatively examine the interaction between businesses' willingness to provide data to government and their views of the governments' ability to protect the data, nor did it examine whether small businesses had different views of statistical agencies than did large.

III. DATA

6. The data are derived from a recent mail survey of 5,000 businesses administered by the Urban Institute on behalf of the U.S. Census Bureau. The data design and collection is described in much more detail in Greenia et al. (2001)

III.1 Questionnaire Design

7. The survey itself was designed to answer a number of questions about business attitudes towards confidentiality. In this study, we analyze two issues in some depth: first, we provide some preliminary evidence of businesses' perceptions of the ability of statistical agencies to do their jobs, and second, we examine their perception of the ability of statistical agencies to protect their data. In each case, we are interested in examining the relationship between their perceptions and their willingness to provide the federal government with data.

8. There are three questions in the survey that address these questions. The first one, Question 5, which is reproduced in Box 1, examines the willingness of respondents to provide data to different types of recipients. In this paper, we are particularly interested in their responses to parts a and b of the question: the concern associated with providing data to federal regulatory agencies (5a) and federal statistical agencies (5b).

Box 1: Question 5

5. In your opinion, how concerned is your company about providing business data (e.g., number of employees, payroll, sales, profits, etc.) to each of the following? *Circle one response for each line.*

	<u>Not at all Concerned</u>	<u>Somewhat Concerned</u>	<u>Very Concerned</u>	<u>Extremely Concerned</u>	
a. Federal regulatory authorities (e.g., Environmental Protection Agency, Securities and Exchange Commission and Federal Trade Commission)	1	2	3	4	DK/NA
b. Federal statistical agencies (e.g., Census Bureau, Bureau of Labor Statistics and Bureau of Economic Analysis)	1	2	3	4	DK/NA
c. Not-for-profit researchers (e.g., universities, think tanks and research organizations)	1	2	3	4	DK/NA
d. For-profit researchers (e.g., market researchers and consulting firms)	1	2	3	4	DK/NA
e. Other businesses	1	2	3	4	DK/NA
f. The general public	1	2	3	4	DK/NA

Box 1: Question 5

9. The second question of interest, Question 6 in the survey, attempts to quantify the respondent's sense of the ability of the government to collect data (part a) and provide useful information to policy makers (part b). It follows with a question about the federal government's ability to protect data (part c), which is expanded on in the subsequent question.

Box 2: Question 6

6. Business data are collected for statistical purposes by both the federal government and the private sector (e.g., universities, think tanks, research organizations, market researchers, consulting firms, etc.). For each statement below, please indicate if you strongly agree, somewhat agree, somewhat disagree or strongly disagree. *Circle one response for each line.*

	<u>Strongly Agree</u>	<u>Somewhat Agree</u>	<u>Somewhat Disagree</u>	<u>Strongly Disagree</u>	
a. The federal government is better than the private sector at collecting business data for statistical purposes.	1	2	3	4	DK/NA
b. The federal government is better than the private sector at providing useful information to government policy makers.	1	2	3	4	DK/NA
c. The federal government is better than the private sector at protecting business data from being released to those without authority to have it.	1	2	3	4	DK/NA

10. Question 7 is directly focused on examining respondents' views and understanding of the federal government's ability to protect data. In this paper, we examine the links between the responses to parts a, b and d and their responses to Question 5a and b.

Box 3: Question 7

7. For each statement below, please indicate if you strongly agree, somewhat agree, somewhat disagree or strongly disagree. *Circle one for each line.*

	Strongly <u>Agree</u>	Somewhat <u>Agree</u>	Somewhat <u>Disagree</u>	Strongly <u>Disagree</u>	
a. I believe that federal statistical agencies (e.g., Census Bureau, Bureau of Labor Statistics and Bureau of Economic Analysis) keep data provided by businesses confidential.	1	2	3	4	DK/NA
b. I believe that federal statistical agencies do not release information by which a company or its data can be identified.	1	2	3	4	DK/NA
c. I believe that federal statistical agencies do not share data provided by businesses with other government agencies.	1	2	3	4	DK/NA
d. I believe that federal statistical agencies do not release data provided by businesses to people outside the government.	1	2	3	4	DK/NA
e. I believe that any federal agency, such as the Internal Revenue Service, Small Business Administration and Federal Trade Commission, can access business data my company has provided to other federal agencies whenever it wants.	1	2	3	4	DK/NA

The analytical section will examine the interactions across the responses to these different items.

III.2 Data Collection and Response Rates

11. The sample frame for the survey was obtained from Dunn and Bradstreet's commercial database, which represents 11,300,000 businesses in the U.S. From this sample frame, we selected a stratified random sample of 5000 cases; 1,250 in each of four strata defined by the number of employees – 1-49 employees; 50-249, 250 –499 and 500 plus. Of the 5,000 total businesses, 2,530 were multi-unit businesses with headquarters locations, and 2,470 were single unit companies with only one business. While the geographic coverage was national, and the sample was representative of Dunn and Bradstreet's database, we do not attempt to weight the sample to assign any kind of representativeness to the sample and simply view our results as preliminary evidence.

12. We followed the Dillman mail survey approach. The first survey was followed two weeks later by a reminder postcard. This was followed by a repeat survey mailout two weeks after that. The outcome was that 509 questionnaires were returned, resulting in a response rate of just over 10%. While this response rate is quite reasonable for a private sector survey of businesses, the response rates for mandatory, government surveys often exceed 80%. Since a more detailed review of the response rates is available in Greenia (2002), we simply summarize the results by noting that response rates declined with business size; and varied by industry type – manufacturing firms were less likely to respond than were service sector firms. Both the response rate, and the sample size, however, mean that our results should be viewed as preliminary rather than definitive in nature.

IV. RESULTS

13. What are respondents' views in general? A brief examination of Table 1 reveals that, by and large, there is much higher trust in the federal statistical system than other potential data collectors. However, one in 3 respondents are still quite concerned about providing their data to these agencies, suggesting that in absolute terms, there is still a large group of businesses that mistrust statistical agencies.

**Table 1:
Proportion of Respondents Who Were Very or Extremely Concerned About Providing Business Data to the Following Entities (By Employment Size) (Question 5)**

Type of recipient	1-49	50-249	250-499	500+	Total
Federal Regulatory Authority (EPA, SEC, FTC)	48.40%	43.20%	25.00%	32.00%	38.20%
Federal Statistical Agency (BLS, Census, BEA)	34.10%	33.10%	20.00%	27.80%	29.30%
Not for Profit Researchers (Universities, Think-Tanks, Research Organizations)	42.20%	44.80%	41.00%	37.50%	41.60%
For-Profit Researchers	60.90%	62.40%	68.70%	56.70%	62.10%
Other Businesses	73.60%	76.00%	77.80%	67.70%	73.90%
The General Public	70.50%	73.60%	66.00%	53.60%	66.70%

Source: Greenia et al. 2001

14. There do appear to be differences in opinions by size of firm. The smallest businesses are substantially more reluctant to provide their data to federal regulatory agencies than are large ones, and slightly more reluctant to provide their data to federal statistical agencies. All businesses, regardless of size, are most suspicious of for-profit researchers, however.

15. The perception of the statistical system's competence at doing its job seems, unfortunately, not particularly good. As Table 2 indicates, only one in three respondents believe that the "federal government" -- collects data better than the private sector, or that it provides more useful information than the private sector. The differences across size class are not marked, however.

Table 2: Proportion of Respondent Who Somewhat or Strongly Agreed with the Following Statements (Question 6)					
	1-49	50-249	250-499	500+	Total
The federal government is better than the private sector at collecting business data for statistical purposes	32.60%	23.80%	38.00%	33.30%	31.50%
The federal government is better than the private sector at providing useful information	38.80%	26.20%	31.30%	29.20%	31.60%
The federal government is better than the private sector at protecting business data from being released to those without authority to have it	35.90%	39.20%	42.00%	24.00%	35.60%

Source: Greenia et al, 2001

16. An examination of the results presented in Table 3 suggests that, by and large, more businesses believe that their data are protected than do not. Here, however, we do see a marked difference by size of firm – the smallest firms have the least confidence in federal statistical agencies. This is consistent regardless of the way in which the question is phrased: the responses to the questions in the first two rows are consistent. The apparent contradiction between the answers to the first two questions and the second two (which is that the government releases data to other agencies and to the public) may possibly be due to a confusion between the release of micro-data and aggregate level data.

	1-49	50-249	250-499	500+	Total
I believe that federal statistical agencies keep data provided by businesses confidential	48.10%	69.80%	69.70%	60.80%	61.60%
I believe that federal statistical agencies do not release information by which a company or its data can be identified	41.10%	66.40%	65.70%	57.70%	57.10%
I believe that federal statistical agencies do not share data provided by businesses with other government agencies	21.10%	21.60%	27.60%	21.60%	22.80%
I believe that federal statistical agencies do not release data provided by business to people outside the government	28.30%	41.70%	43.30%	38.50%	37.60%
I believe that any federal agency, such as the Internal Revenue Service, Small Business Administration and Federal Trade Commission can access business data my company has provided to other federal agencies whenever it wants	71.10%	68.50%	67.30%	73.20%	70.00%

Source: Greenia et al, 2001

17. In any event, possibly the most interesting result in this table is the response to the last question – businesses overwhelmingly believe that particular federal agencies can access business data at any time they want, despite the existence of very strong legal constraints on such access.

18. How do these different beliefs held by businesses correlate with their concern about providing data to federal statistical agencies? We examine this by estimating simple linear probability models which regress the response to questions 5a and b (coded 1 if the respondent was extremely or very concerned about providing data to federal regulatory or statistical agencies respectively) against their responses to question 6. These results are reported in Table 4, and the results of a similar regression against the responses to question 7 are reported in Table 5.

19. The most interesting result from Table 4 is that the data support the a priori expectations that the more the business believes the government does a good job in collecting and producing information, the less concerned it is about providing data to the government, as evidenced by the negative and significant coefficients reported in the first three rows of the table. Indeed, the size of the coefficients is remarkably similar regardless of whether the data collection entity is a federal statistical or a federal regulatory agency. However, the constant term is markedly higher when the respondents are asked about federal regulatory agencies – suggesting that the level of concern is higher for regulatory agencies.

20. One other interesting result is apparent from an examination of the table – namely, that the response varies significantly by industry. In particular, firms in the Financial, Insurance and real estate industry as well as in manufacturing are significantly less likely to be concerned about federal regulatory agencies than are other firms – but they are no different in their concerns about statistical agencies than are firms in other industries.

Table 4						
The Relationship Between Response and Business Perceptions of Government Competence						
Dependent Variable: Respondent said he/she was very or extremely concerned about providing business data to the following entities						
	Federal Regulatory Authority (EPA, SEC, FTC)			Federal Statistical Agency (BLS, Census, BEA)		
The federal government is better than the private sector at collecting business data for statistical purposes	-0.123 (2.59)**			-0.116 (2.44)*		
The federal government is better than the private sector at providing useful information		-0.125 (2.42)*			-0.123 (2.59)**	
The federal government is better than the private sector at protecting business data from being released to those without authority to have it			-0.132 (2.64)**			-0.149 (3.24)**
Manufacturing	-0.234 (2.35)*	-0.211 (2.10)*	-0.209 (2.09)*	-0.115 (1.24)	-0.103 (1.11)	-0.102 (1.11)
Wholesale	-0.168 (1.35)	-0.138 (1.11)	-0.140 (1.13)	-0.048 (0.41)	-0.035 (0.30)	-0.039 (0.34)
Retail	-0.185 (1.71)	-0.167 (1.54)	-0.151 (1.39)	0.001 (0.01)	0.013 (0.13)	0.031 (0.31)
FIRE	-0.318 (2.88)**	-0.299 (2.69)**	-0.290 (2.61)**	-0.211 (2.04)*	-0.207 (2.01)*	-0.200 (1.95)
Services	-0.266 (2.96)**	-0.232 (2.58)*	-0.219 (2.45)*	-0.174 (2.09)*	-0.158 (1.91)	-0.146 (1.78)
Constant	0.641 (7.45)**	0.595 (6.99)**	0.593 (7.03)**	0.440 (5.53)**	0.429 (5.50)**	0.435 (5.62)**
Observations	404	404	404	411	411	411
R-squared	0.05	0.03	0.04	0.04	0.04	0.05
Mean of Dependent Variable	0.36			.28		
Absolute value of t statistics in parentheses; * significant at 5%; ** significant at 1%; omitted industry is agriculture, mining, contract construction						

21. A very similar pattern is uncovered by an examination of the relationship between trust and data provision (Questions 5 and 7), the results of which are reported in Table 5. Just as in the previous table, the concern a business exhibits at providing data to federal statistical and regulatory agencies is highly correlated by its view of the agencies' competence – in this case, by the trust in the ability of the agency to protect its data.

22. As before, there are interesting industry differences in businesses' views of regulatory and statistical agencies. In this case, however, businesses in retail trade as well as in finance, insurance, real estate and manufacturing are significantly less likely to distrust federal regulatory agencies than are firms in other industries – but their views of federal statistical agencies are not different.

23. It is easy to speculate about the reasons for these industry differences – but clearly only empirical analysis will provide the answer. An obvious possibility is that federal regulatory agencies have more interaction with these industries than others. Hence, firms in these industries have actually had contact with the regulatory agencies, and have found them generally competent. An alternative possibility is that the reverse is the case!

Dependent Variable: Respondent said he/she was very or extremely concerned about providing business data to the following entities						
	Federal Regulatory Authority (EPA, SEC, FTC)			Federal Statistical Agency (BLS, Census, BEA)		
I believe that federal statistical agencies keep data provided by businesses confidential	-0.238 (4.98)**			-0.199 (4.45)**		
I believe that federal statistical agencies do not release information by which a company or its data can be identified		-0.192 (4.05)**			-0.150 (3.38)**	
I believe that federal statistical agencies do not release data provided by business to people outside the government			-0.166 (3.45)**			-0.119 (2.64)**
Manufacturing	-0.213 (2.17)*	-0.187 (1.89)	-0.197 (1.98)*	-0.106 (1.17)	-0.082 (0.89)	-0.091 (0.99)
Wholesale	-0.167 (1.37)	-0.130 (1.06)	-0.129 (1.04)	-0.057 (0.50)	-0.022 (0.19)	-0.024 (0.20)
Retail	-0.214 (2.01)*	-0.150 (1.40)	-0.172 (1.59)	-0.034 (0.34)	0.026 (0.26)	0.007 (0.07)
FIRE	-0.311 (2.86)**	-0.280 (2.56)*	-0.279 (2.53)*	-0.217 (2.14)*	-0.187 (1.83)	-0.187 (1.82)
Services	-0.243 (2.78)**	-0.225 (2.55)*	-0.230 (2.59)*	-0.169 (2.08)*	-0.150 (1.83)	-0.154 (1.87)
Constant	0.719 (8.18)**	0.656 (7.63)**	0.616 (7.29)**	0.529 (6.47)**	0.466 (5.86)**	0.431 (5.52)**
Observations	404	404	404	411	411	411
R-squared	0.08	0.06	0.05	0.07	0.05	0.04
Absolute value of t statistics in parentheses * significant at 5%; ** significant at 1% ; omitted industry is agriculture, mining, contract construction						

V. CONCLUSIONS

24. The preliminary research described in this study suggests that there is a link between businesses' willingness to provide confidential data to statistical agencies and their trust in both the competence of the agency in using the data, and the ability of the agency to protect it. The research also suggests that a worryingly high number of respondents are quite distrustful of federal statistical agencies, and this was substantiated by many of the handwritten comments on the returned surveys.

25. It is clear, however, that more representative data should be collected in order to support these preliminary results. Our rather disappointing response rates, combined with the evidence that businesses trust statistical agencies more than think tanks or outside researchers, suggest that the appropriate entity to conduct such research are federal statistical agencies themselves. In addition to the imprimatur and data collection infrastructure provided by the statistical system, the agencies stand to gain the most from such research – both in terms of developing a relationship with the respondent community and in terms of supporting policy decisions. As statistical agencies well know, good decisions are driven by good data.

REFERENCES

- Biemer, P.B. and Fecso, R.S., "Evaluating and Controlling Measurement Error in Business Surveys." Chapter 15 in B. Cox, et al. (eds.), *Business Survey Methods*. New York: John Wiley & Sons, 1995.
- Christianson, A. and Tortora, R., "Issues in Surveying Businesses: An International Survey." Chapter 14 in B. Cox, et al. (eds.), *Business Survey Methods*. New York: John Wiley & Sons, 1995.
- Cox, B.G. and Chinnappa, B. N., "Unique Features of Business Surveys." Ch. 1 in B. Cox et al. *Business Survey Methods*. New York, John Wiley & Sons, 1995.

Dillman, D. A., "Mail and Telephone Surveys: The Total Design Method." New York: John Wiley & Sons, 1978.

Edwards, W.S. and Cantor, D., "Toward a Response Model in Establishment Surveys." Chapter 12 in Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A. and Sudman, S. (eds.), *Measurement Errors in Surveys*. New York: John Wiley & Sons, 1991.

Greenia, N, J.B. Jensen and J. Lane., "Business Perceptions of Confidentiality." in *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*, eds Pat Doyle, Julia Lane, Laura Zayatz and Jules Theeuwes, North Holland, 2001.

Mackie, C and N. Bradburns (eds), *Improving Access to and Confidentiality of Research Data: Report of a Workshop* National Research Council, 2000.

Nichols, E., Willimack, D. K. and Sudman, S., "Balancing Confidentiality and Burden Concerns in Censuses and Surveys of Large Businesses". Paper presented to the Washington Statistical Society, U.S. Bureau of the Census, Washington, D.C., September, 1999.

Singer, E., "Public Perceptions of Confidentiality and Attitudes Towards Data Sharing by Federal Agencies" in *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*, eds Pat Doyle, Julia Lane, Laura Zayatz and Jules Theeuwes, North Holland, 2001.

Singer, Eleanor, S. Presser and J. Vanhoewyk, "Public Attitudes Toward Data Sharing By Federal Agencies", in *Record Linkage Techniques - 1997 Proceedings of an International Workshop and Exposition* National Academy Press, 1999.

Wallman, K and J. Coffey, "Sharing Statistical Information for Statistical Purposes" in *Record Linkage Techniques - 1997 Proceedings of an International Workshop and Exposition* National Academy Press, 1999.

Willimack, Diane K., Elizabeth Nichols and Seymour Sudman (forthcoming) "Understanding Unit and Item Nonresponse In Business Surveys," in *Survey Nonresponse*, Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J.A. Little (editors), New York: John Wiley and Sons.

Willimack, Diane K, Seymour Sudman, Elizabeth Nichols and Thomas Mesenbourg 1999 "Cognitive Research on Large Company Reporting Practices: Preliminary Findings and implication for Data Collectors and Users", unpublished paper presented to the Census Advisory Committee of Professional Associations, U.S. Bureau of the Census, Washington, DC, March 2000.