Schmollers Jahrbuch 124 (2004), 305 – 314
Duncker & Humblot, Berlin

European Data Watch

This section will offer descriptions as well as discussions of data sources that may be of interest to social scientists engaged in empirical research or teaching courses that include empirical investigations performed by students. The purpose is to describe the information in the data source, to give examples of questions tackled with the data and to tell how to access the data for research and teaching. We will start with data from German speaking countries that allow international comparative research. While most of the data will be at the micro level (individuals, households, or firms), more aggregate data and meta data (for regions, industries, or nations) will be included, too. Suggestions for data sources to be described in future columns (or comments on past columns) should be send to: Joachim Wagner, University of Lueneburg, Institute of Economics, Campus 4.210, 21332 Lueneburg, Germany, or e-mailed to ⟨wagner@uni-lueneburg.de⟩.

Firm Panel Data from the Swiss Innovation Survey

By Spyros Arvanitis and Heinz Hollenstein

1. Introduction

The firms' capability to innovate is one of the most important factors determining economic growth; for a small open economy, innovation is crucial for securing international competitiveness. For this reason, the Swiss Institute for Business Cycle Research (KOF) of the Swiss Federal Institute of Technology Zurich (ETHZ) launched an Innovation Survey in 1990. This project was financially supported by the Swiss National Science Foundation and the Swiss Federal Ministry for Economic Affairs. Based on this first experience, the Ministry became the permanent sponsor of the survey. The various waves of the Innovation Survey are the fundament for building a firm panel, which is unique in Switzerland.

The first and the second Swiss Innovation Survey of 1990 and 1993 respectively addressed only manufacturing firms; since 1996, the construction indus-

try and (the commercial part of) the service sector are also included in the survey. This is a postal survey which is conducted with a periodicity of three years, with the most recent survey undertaken in 2002. The panel data set thus comprises information from five surveys in manufacturing and three surveys in the construction and service sector respectively. The questions concerning innovative activities refer to a three-year period; for example, the survey 2002 covered the period 2000 – 2002. As a consequence, the available data extend across a period of fourteen years from 1988 / 90 to 2000 / 2002, that is the lenght of almost two business cycles.

The firm panel underlying the Innovation Survey is also used to collect data on some related topics. To mention are surveys on professional qualifications (1998), internationalization of Swiss firms (1998), ICT and workplace organization (2000) as well as ICT and E-Commerce (2002). The combination of the regular Innovation Survey and these more specific surveys yields a rich database which can be used for microeconometric cross-section and longitudinal investigations.

In the next section we describe the main topics addressed by means of the Innovation Survey. Section 3 is devoted to some statistical aspects such as the structure of the panel, the way "unit non-response and "item non-response" are treated, the weighting system used to aggregate data at the meso- and macro-level, etc.. In section 4 we indicate a selection of econometric studies based on the data collected by means of the Innovation Survey to illustrate the way these data can be used. Finally, we shortly discuss problems of data access and give some indication of the direction the panel is planned to be further developed in the next few years.

2. Information in the data

The data are collected by means of a questionnaire which has been considerably modified since 1990. Most modifications were related to new topics and new questions which were added to the questionnaire. However, the basic structure remained the same until today, so that time series of the most important data could be constructed. Since 1993 structure, content and wording of questions have been adapted to a large extent to the standard questionnaire of the Community Innovation Survey (CIS) and the guidelines of the Oslo Manual of the OECD. The last version of the Swiss Innovation Survey contained practically every question of CIS III and covered roughly four big blocks of variables:[1]

[1] Versions of the questionnaires in German, French and Italian can be found in www.kof.ethz.ch.

- *Structural characteristics and performance measures of the firm, characteristics of market environment:* firm age, changes of firm organization, share of employees with different levels of formal education, sales, exports, labour costs, intermediate inputs, market conditions (demand, number of competitors, intensity of price and non-price competition);
- *Innovation and R&D activities:* a wide spectrum of qualitative and quantitative innovation indicators at various stages of the innovation process (input-, output-, market-oriented measures), also variables on R&D co-operation (including modes, partners, motives and outcomes of R&D co-operation), external R&D, R&D abroad, etc.;
- *Characteristics of the innovation process:* motives / objectives of innovation, relevance of a series of external knowledge sources, technological potential, appropriabilty of innovation benefits, inward and outward licencing;
- *Obstacles to innovation:* importance of a series of possible obstacles covering all fields of activities relevant to innovation including several fields of policy making.

3. Statistical Background

Panel Sample

The survey is based on a (with respect to firm size) disproportionately stratified random sample of firms with at least 5 employees (with four exceptions in manufacturing industries with a very high share of very small firms; see table 1) covering all relevant industries of the business sector as well as firm size classes. On the whole, the sample contains 28 industries, and within each industry three industry-specific firm size classes ("small", "medium-sized" and "large") with full coverage of the class of large firms; the thresholds used to define the three size classes in an industry are determined by the method of optimal stratification (see Cochran 1977). Table 1 shows the current composition of the sample which contains 6'600 firms (manufacturing: 3000 firms; construction: 600 firms; services: 3000 firms) distributed over 84 cells.

The structure of the sample is adapted to the most recent Enterprise Census; the size classes are up dated according to the changes in the underlying population; in addition, small firms which did not participate at the panel surveys conducted since the last five years are replaced randomly by other small firms.

Table 1

Structure of the Sample (adapted to Enteprise Census 1998)

Industry	Small	Middle	Large	*N*
Food, beverage	6 – 71	71 – 479	>= 479	299
Textiles	6 – 47	47 – 171	>= 171	75
Clothing, leather	6 – 48	48 – 325	>= 325	48
Wood processing	10 – 23	23 – 73	>= 73	76
Paper	6 – 66	66 – 229	>= 229	60
Printing	10 – 50	50 – 283	>= 283	197
Chemicals	6 – 137	137 – 1054	>= 1054	225
Plastics, rubber	6 – 48	48 – 224	>= 224	115
Glass, stone, clay	6 – 33	33 – 143	>= 143	88
Metal	6 – 91	91 – 431	>= 431	82
Metal working	6 – 30	30 – 133	>= 133	355
Machinery	6 – 77	77 – 506	>= 506	527
Electrical machinery	6 – 92	92 – 761	>= 761	153
Electronics, instruments	6 – 75	75 – 423	>= 423	297
Watches	6 – 92	92 – 656	>= 656	146
Vehicles	6 – 85	85 – 665	>= 665	66
Other manufacturing	10 – 40	40 – 164	>= 164	99
Energy, water	10 – 102	102 – 466	>= 466	92
Construction	6 – 26	26 – 125	>= 125	600
Wholesale trade	6 – 32	32 – 202	>= 202	586
Retail trade	6 – 49	49 – 687	>= 687	550
Hotels, catering	6 – 25	25 – 181	>= 181	409
Transport	6 – 117	117 – 2509	>= 2509	445
Banks, insurances	6 – 179	179 – 3652	>= 3652	366
Real estate, leasing	6 – 23	23 – 95	>= 95	40
Computer services	6 – 37	37 – 221	>= 221	114
Business services	6 – 34	34 – 334	>= 334	436
Personal services	6 – 14	14 – 53	>= 53	54
Total				6600

Non-response analysis

The response rate was increased from about 26 % in the first survey of 1990 (only manufacturing) to about 41 % in the last one of the year 2002 (manufacturing, construction and services). This rate is low from the point of view of statistical adequacy, but rather satisfactory if compared with that of similar

surveys in other countries where participation is not compulsory. Nevertheless, the low response rate leaves open the question of the existence of a selection bias with respect to innovative activities. Possible selection bias can be corrected by a "unit non-response analysis".

The non-response analysis is based on a follow-up telephone survey of a sample of about 600 non-respondents. The telephone survey is very brief containing just a few questions on the central variables of the questionnaire ("introduction of product innovations yes / no"; "introduction of process innovations yes / no"; "existence of R&D activities yes / no"). For such a follow-up survey a response rate of at least 85 % has to be attained (see Donzé 2002a for details on the methodology used).

Treatment of missing values

Another type of selection bias results from non-response of single survey questions (problem of "item non-response") which leads to missing values in the data. We use the multiple imputations technique developed by Rubin (1987) to substitute for missing values in the variables due to item non-response (see Donzé 2001 for a detailed report on this technique).

Data weighting

We apply a multiple weighting of data in order to correct for possible selection bias as well as divergences from sample structure and from total firm population respectively, particularly when the individual data are aggregated at industry or sector level. The single weights used are as follows:

– Sample structure: we define a weight w_{hi} for every observation (firm) i belonging to cell h ($h = 1, \ldots 84$):

$$w_{hi} = 1/f_h = 1/(n_h/N_h) = N_h/n_h$$

with: f_h: drawing rate of cell h

n_h: number of firms in cell h in the sample

N_h: number of firms in cell h in the total population (Federal Enterprise Census 1998)

– Non-response rate: we define for every firm i in the cell h a weight $1/r_{hi}$, where r_{hi} is the probability that firm i responds to the questionnaire. The estimation of this probability is based on a logistic regression of the non-response frequency on the main firm characteristics (industry, size class, region, etc.). Weight w_{hi} is multiplied with $1/r_{hi}$ to give the twofold weight w^*_{hi}:

$$w^*_{hi} = w_{hi} 1/r_{hi}$$

– Non-response analysis: The results of the follow-up survey are used to calculate "calibrated" weights $w^{*(c)}_{hi}$ by applying the method of the "calibration on margins" (see Deville et al. 1993 and Sautory 1993):

$$w^*_{hi} \rightarrow w^{*(c)}_{hi} \qquad \textit{(c: calibration)}$$

– Calculation of relative weights of the single cells:

$$k^{BE}_h = BESCH98_h / BESCH98_s$$

$$k^{UM}_h = UMT_h / UMT_s$$

where $BESCH98_h$ is the number of employees (in full time equivalents) of all firms in cell h and $BESCH98_s$ the number of employees (in full time equivalents) in sector s (based on the Federal Enterprise Census 1998) respectively and UMT_h and UMT_s similar weights based on sales data of the survey.

– Finally, we define a total weight which is the product of $w^{*(c)}_{hi}$ and the relative weights of the single cells k^{BE}_h and k^{UM}_1 respectively:

$$g^{BE}_{hi} = w^{*(c)}_{hi} 1/k^{BE}_h$$

$$g^{UM}_{hi} = w^{*(c)}_{hi} 1/k^{UM}_h$$

Firm data panel

The panel contains time series for about 4'900 manufacturing firms (for the years 1990, 1993, 1996, 1999, 2002), 740 construction firms and 3'300 firms of the service sector (for the years 1996, 1999, 2002) with information on about 350 variables covering measures of general firm characteristics as well as of firm performance, variables for characteristics of a firm's market environment, measures of innovation activities at various stages of the innovation process and variables for several characteristics of the innovation process.

4. Selected studies based on the micro data

The results of the descriptive analysis and of selected econometric investigations are published for each survey in a series edited by the Swiss Federal Ministry for Economic Affairs; see, for example, the report based on the data stemming from the innovation survey 1999 (Arvanitis et al. 2001).

In addition, many other studies have been conducted drawing on the data of the innovation surveys. We would like to mention a series of topics that were investigated by econometric methods since 1990:

- Development of a *composite indicator of innovation performance* of a firm by applying factor analysis on a large number of input-, output- and market-oriented innovation measures (Hollenstein 1996).
- Determination of various *"modes of innovation"* based on a cluster analysis of innovation indicators. The resulting clusters are characterized by variables which capture a firm's knowledge network, technological opportunities, appropriability, etc. as well as some structural characteristics of the firm (size, age, industry affiliation, etc.). In a next step, it is investigated whether these innovation modes significantly differ in terms of economic performance (manufacturing: Arvanitis and Hollenstein 2001a; services: Hollenstein 2003).
- *Explanation of a firm's innovation performance* based on demand conditions, market conditions (concentration, intensity of price and non-price competition), appropriability of knowledge, technological opportunities (manufacturing: Arvanitis and Hollenstein 1994 ans 1996a; services: Arvanitis 2002); estimation of an extended model taking into account innovation objectives as well as obstacles to innovation (Arvanitis and Hollenstein 1996b). The specific role of firm size is analyzed in Arvanitis (1997). Arvanitis and Marmet (2002) investigated problems of financing innovative activity. Lenz (1998) estimated a model explaining simultaneously innovative activity and R&D co-operation.
- Several studies deal with various aspects of the *impact of innovative activity on firm performance.* An overall treatment of this topic is found in Arvanitis, 1999; more specific aspects such as the impact of technological spillovers and knowledge heterogeneity on firm performance are analyzed in Arvanitis and Hollenstein (1998a, 2002). The *impact of innovation on the demand for skills* is investigated in Arvanitis and Hollenstein (1998b).
- *Other topics* analyzed by use of data from the innovation surveys are the *internationalization of R&D* of Swiss firms (Arvanitis and Hollenstein 2001b), the *adoption of new technologies* (Arvanitis and Hollenstein 2001c) and the *evaluation of programmes promoting the diffusion of new technology* (Arvanitis et al. 2002; Donzé 2002b).

Almost all above-mentioned studies are based on cross-sections of data. An important exception is the study in Arvanitis et al. (2001, Ch.6), in which a model of a firm's innovation performance of a firm is estimated based on a longitudinal data set containing the observations of the cross-sections of the years 1993, 1996 and 1999; the results are largely in line with those of the same model estimated with cross-section data.

5. Data access

A well-defined policy of access to the data of the Swiss Innovation Survey does not exist until now. One has to distinguish between access to results at an aggregate level (industries or sectors; firm size classes) and access to the micro data.

At the aggregate level, we tabulate results according to the specifications required by outside users (researchers, institutions) on an ad-hoc basis. Micro data are strictly confidential, there is no access for outside researchers at the moment.

6. Perspectives

The plans for the near future are as follows:

- The *analysis of longitudinal date* will become a focus of our work. We are going to investigate, firstly, the impact of innovative activity on firm productivity, and, secondly, the factors determining the internationalization of R&D.
- The longitudinal database, which to date covers only the results from the five waves of the innovation survey, will be extended to *data stemming from some of the other panel surveys:* "Use of ICT and E-Commerce" (2000, 2002, "Internationalization of the Swiss Economy", 1998).
- It is urgent to tackle the problem of access to the micro data for external users. To this end, we shall assess the experience made by other institutions conducting firm-level panel surveys. In addition, we shall investigate alternative methods of anonymization of the data. Based on this work, we have to look for a solution which is practicable in the Swiss institutional environment.
- Finally, we intend to investigate the problem of "item non-response" in the setting of a longitudinal database.

If we succeed in developing the panel according to these lines, the Swiss Innovation Panel will have been transformed to a comprehensive, methodologically quite sophisticated "Enterprise Panel" containing information on a large set of topics.

References

Arvanitis, S. (1997): The Impact of Firm Size on Innovative Activity. An Empirical Analysis Based on Swiss Firm Data, Small Business Economics, 9(6), 473 – 490.

– (1999): Generierung von neuem technischem Wissen, Produktivität und Arbeitsqualifikation in der schweizerischen Industrie: Eine Querschnittsanalyse auf der Basis von Unternehmensdaten, Dissertation, Universität Zürich.

– (2002): Explaining Innovative Activities in Service Industries: Micro-data Evidence for Switzerland, in: Economic Surveys and Data Analysis: CIRET Conference Proceedings, Paris 2000, SourceOECD General Economics & Future Studies, vol. 2002, no. 11, Paris.

Arvanitis, S. / *Bezzola,* M. / *Donzé,* L. / *Hollenstein,* H. / *Marmet,* D. (2001): Innovationsaktivitäten in der Schweizer Wirtschaft. Eine Analyse der Ergebnisse der Innovationserhebung 1999, Studienreihe Strukturberichterstattung Nr. 5, hrsg. vom Staatssekretariat für Wirtschaft, Bern.

Arvanitis, S. / *Hollenstein,* H. (1994): Demand and Supply Factors in Explaining the Innovative Activity of Swiss Manufacturing Firms, Economics of Innovation and New Technology, 3(1), 15 – 30.

– (1996a): Industrial Innovation in Switzerland: A Model-based Analysis with Survey Data, in: A. Kleinknecht (ed.): Determinants of Innovation. The Message from New Indicators, Macmillan, London.

– (1996b): Das Innovationsprofil der Schweizer Industrie. Determinanten, Zielorientierung und Hemmnisfaktoren der Innovationstätigkeit, Schweizerische Zeitschrift für Volkswirtschaft und Statistik, 132(3), 335 – 358.

– (1998a): Firm Performance, Innovation and Technological Spillovers: A Cross-section Analysis with Swiss Firm Data, in: G. Eliasson, C. Green and C. McCann (eds.), The Microfoundations of Economic Growth. A Schumpeterian Perspective, University of Michigan Press, Ann Arbor.

– (1998b): Arbeitsqualifikation, Beschäftigung und Innovationsaktivitäten: Erste empirische Ergebnisse einer ökonomischen Analyse anhand von Unternehmensdaten, Beitrag für den Kongress "Bildung und Arbeit", Universität Zürich, 24. – 26. September.

– (2001a): Innovative Activity and Firm Characteristics. A Cluster Analysis of Swiss Manufacturing Using Firm-level Data, in: OECD (ed.): Innovative Networks: Co-operation in National Innovation Systems, Paris, 48–76.

– (2001b): Technologiestandort Schweiz im Zuge der Globalisierung: Eine explorative Analyse der F&E-Aktivitäten schweizerischer Industrieunternehmen im Ausland, Schweizerische Zeitschrift für Volkswirtschaft und Statistik, 137(2), 129 – 148.

– (2001c): The Determinants of Adoption of Advanced Manufacturing Technologies – An Empirical Analysis Based on Firm-level Data for Swiss Manufacturing, Economics of Innovation and New Technology, 10(5), 372 – 414.

– (2002): The Impact of Technological Spillovers and Knowledge Heterogeneity on Firm Performance: Evidence from Swiss Manufacturing, in: A. Kleinknecht and P. Mohnen (eds.): Innovation and Firm Performance, Palgrave, London.

Arvanitis, S. / *Hollenstein,* H. / *Lenz,* S. (2002): The Effectiveness of Government Promotion of Advanced Manufacturing Technologies (AMT): An Economic Analysis Based on Swiss Micro Data, Small Business Economics, 19(4), 321 – 340.

Arvanitis, S. / *Marmet,* D. (2002): Finanzierung und Innovationsaktivitäten – Eine empirische Analyse anhand von Unternehmensdaten, Studienreihe Strukturberichterstattung Nr. 9, hrsg. vom Staatssekretariat für Wirtschaft, Bern.

Cochran, W. G. (1977): Sampling Techniques, Wiley, New York.

Deville, J. C. / *Särndal,* C.-E. / *Sautory,* O. (1993): Generalized Ranking Procedures in Survey Sampling, Journal of the American Statistical Association, 88, 1013 – 120.

Donzé, L. (2001): L'imputation des données manquantes, la technique de l'imputation multiple, les conséquences sur l'analyse des données: l'enquête 1999 KOF / ETHZ sur l'innovation, Schweizerische Zeitschrift für Volkswirtschaft und Statistik, 137(3), 301 – 317.

– (2002a): Methodology to Correct the Unit Non-Response Bias in the Case of the KOF ETH Zurich Survey 2000 on Organization and Information Technology, KOF-Arbeitspapiere / Working Papers, No. 67, Zurich.

– (2002b): Matched-Pairs Analysis Based on Business Survey Data to Evaluate the Policy of Supporting the Adoption of Advanced Manufacturing Technologies by Swiss Firms, KOF-Arbeitspapiere / Working Papers, No. 65, Zurich.

Hollenstein, H. (1996): A Composite Indicator of a Firm's Innovativeness. An Empirical Analysis Based on Survey Data for Swiss Manufacturing, Research Policy, 25, 633 – 45.

– (2003): Innovation Modes in the Swiss Service Sector: a Cluster Analysis Based on Firm-level Data, Research Policy, 32, 845 – 863.

Lenz, S. (1998): Bestimmungsfaktoren des Innovations- und Kooperationsverhaltens von Unternehmen. Theorie und ökonometrische Untersuchung anhand von Daten für die schweizerische Industrie, Dissertation Universität, Zürich.

Rubin, D. B. (1987): Multiple Imputation for Non-Response in Surveys, John Wiley, New York.

Sautory, O. (1993): La macro CALMAR. Redressement d'un échatillon par calage sur marges, Serie des documents de travail de la Direction des Statistiques Démographiques et Sociales 55, INSEE, Paris.