



1402711209



**SWP 3/87**

**FINDING THE NEW FIRM**

**PROFESSOR SUE BIRLEY  
Cranfield School of Management  
Cranfield Institute of Technology  
Cranfield  
Bedford MK43 OAL  
United Kingdom**

**(Tel: 0234-751122)**

**(Fax: 0234-751806)**

**Copyright: Birley 1984**



## Finding the New Firm

### **Abstract**

This paper emanates from a study of the job generation effect of new firms in St. Joseph County, Indiana. It discusses the methodological problems associated with identifying the new firm and compares the results of a search from three sources, the County Unemployment ES202 Files, the Telephone Directory, and the Dun and Bradstreet Market Identifier Files (DMI). Significant differences emerge. If DMI is used, the basis of the Birch Report, 96% of the new firms found in ES202 would fail to be recorded. The paper concludes that the wide variations found suggest that the data base chosen can significantly affect the results of any study of new firms.

### Finding the New Firm

Over the past 6 years the American economy has shown severe turbulence which has been reflected in industrial decline and, with it, a rapid increase in unemployment. In response to this, both local and national government have searched for new ways to revive industries and create new jobs. Legislation, tax incentives, loans, enterprise schemes are some of the mechanisms used, and whilst most are directed at industry in general, the small firm has received particular attention. This is due in large part to the landmark study of Birch [1979] which reported that 81.5% of new jobs were generated by firms with less than 100 employees. Moreover, of those 28% were from firms less than 5 years old. These results were largely confirmed subsequently by Armington and Odle [1982] of the Brookings Institute using the same data base but collating and inspecting the data in a different way.

For the local community, however, the question which remained was whether the national pattern was universal. If this were the case, a strategy for revival could be focussed in part on creating a healthy small firm environment. However, two recent studies by Teitz [1981] in the State of California and Shapero and Giglierano [1982] in Columbus, Ohio have focussed upon the local scene and have produced results which in part contradict those previously reported by Birch. The difficulty which arises in comparing these results is in determining the extent to which they are influenced by the various data sources used. Thus, Birch and Brookings used Dunn and Bradstreet files, Teitz used data from the California State Employment Development Department and Shapero scanned the local telephone directories. This issue is particularly relevant for those parts of the studies which looked at new small firms since the data capture problems for this group are particularly acute, and the probability of both under-reporting and of sector bias in the sample is high.

This paper describes the results of a study which tested the variation between the population of new firms identified by different data sources. It emanates from a two-stage research project into firms which started in St. Joseph County, Indiana between 1977 and 1982. The two stages are:

- Stage 1: Identification of all new firms started each year since 1977 and all those which ceased trading in the area each year. Analysis of their annual birth and death patterns and job generation characteristics, both gross and net.
- Stage 2: Questionnaire survey to all those firms born since 1977 which have survived.

The paper first discusses the methodological problems associated with identifying the new firm and with collecting the necessary data, and second compares the magnitude and characteristics of the samples drawn from the three sources of the Dun and Bradstreet files, the County Unemployment (ES202) files and the telephone directory. It concludes that the ES202 files are significantly more efficient in identifying new firms than the other two sources. Moreover, it is also a superior source when data on job generation is sought.

## METHODOLOGY

### Data Required

As discussed by Birch, the job generation process has various components, which are below:

Net New Jobs = New Jobs due to Expansion of large firms

+ Expansion of small firms

+ Birth of firms

Less Jobs lost due to Contraction of large firms

+ Contraction of small firms

+ Closure of firms

+ Death of firms

Previous studies which have contributed to the understanding of this equation have fallen into three types.

1. The calculation of company birth and death rates (Siropolis [1977], Begin, Cesta, Apilado [1979], Scott Fain [1980], Pennings [1982], Altman [1983]).
2. The contribution of small firms to the job generation process (Fothergill & Gudgin [1979], Scott [1982]).
3. The regional and national patterns of job generation for all firms (Storey [1982], Whittington [1983]).

Although the latter two have included discussion of the contribution of new firms, only British studies have addressed this issue directly (Cross [1981], Binks & Jennings [1983]). No equivalent study has been done in the U.S.A.

On the surface, the data required is very simple, being the capture of the firm at birth and the recording of the number of new jobs created. Further, a number of studies (for example, Birch [1979]) have suggested a shift in birth patterns from the manufacturing to the service sector, and since new firms are likely to lead this trend, it was important in this study to relate the job generation pattern to industry sector. Thus, the data required was:

- For Stage 1:
1. Name of new firm
  2. Date of birth
  3. Jobs generated at birth
  4. SIC classification
- For Stage 2:
5. A current address.

### Defining the Data

#### **a. Identifying the start point:**

The literature abounds with discussion of the characteristics of the entrepreneur and the problems which he encounters during the start-up phase.

What is particularly important to this research however, is the recognition that start-up is a continuous process rather than a discrete event. Many authors have described this process as a series of stages involving the developing of skills, the recognising of opportunities and the assembling of resources (Stanworth and Curran [1976], Birley [1982]). Moreover, some of these stages, which have been described by Cooper [1981] as the incubation period, are passive and can take place over many years. Indeed, even the active period after the final decision to start has been made can take an extended period of time. A previous study by Birley [1983] concluded that, even for those entrepreneurs who finally begin trading, the active pre-trading period can vary from 1 month to 3 years with 6 months being the norm. From this it can be seen that a fundamental issue to be resolved in studying new firms is the definition of the point during the process at which the firm can be said to exist. There are a number of possibilities, all of which have particular data capture implications which are seen in Table 1 below.

Table 1

Start-Point	Problem
Owner makes decision to start	The business is still undefined
First date when owner becomes self-employed	The shape of the business may still be undetermined
Incorporation date	Excludes all partnerships and proprietorships
Bank account established	May still be in pre-start-up phase
Premises and equipment acquired	Orders not received--the business has not established itself in the market place
First order received	But no cash inflow/business may be part-time
Tax first paid	Will "lag" the trading start-up point
First full-time employee	Many firms employ part-time or freelance people

It is clear from Table 1 that any discussion of start-up requires a definition of when the new firm finally comes into existence. Moreover, this will be tempered by the type of survey to be conducted, and whether the data is likely to be available external or internal to the firm. In other words, a study which has already identified a group of firms in existence can ask when the first order was received. Alternatively, one which seeks only new firms must rely upon points in time when the owner chooses to register its existence with an external body such as the Internal Revenue Service.

**b. Refining the data:**

In a sample of firms where the start date is obtained by asking the owner, it is extremely important that the question is very specific, for a loose one can elicit many responses. "When did you start" requires the respondent to define the start point whereas "when did you receive your first order" relies upon memory for an event which may well be undocumented. Multiple checking of these responses is necessary. When the data is drawn external to the firm, a further set of problems arise. A firm may appear to be new because:

- it is a new subsidiary or branch of an existing firm
- it has changed legal status, for example, from proprietorship to incorporation.
- it has changed its name
- it is incorporated and registered but there is no evidence of, or intention to, trade.

This issue follows through to the discussion of company failure (Massel [1978], Scott [1982]). Merely to note that a firm has ceased trading does not automatically imply failure. After all, it may simply have

- moved premises locally
- moved to another region

--closed down one branch

--changed name.

So any study which wishes to isolate entirely new trading firms and those which have failed must include some methodology for refining the data.

Indeed, much of the discussion of the validity of Birch versus Brookings centres on this particular point (see Harris [1983]).

**c. Identifying short-lived companies:**

Most of the studies which compare birth and death rates use published sources of data and this implies that the firm has at some stage registered its existence. However, there can often be a substantial time lag in this process and it is possible that many firms do not survive long enough to be identified. Whilst this is probably of little interest to, say, the Internal Revenue Service, a study of births and deaths which excludes those short-lived ones is likely to produce biased results. So it may be that the turnover rate, or volatility, in some SIC sectors is much higher than current studies suggest.

Possible Data Sources and Their Limitations

The type of data required in this study, viz. individual company births by number of jobs generated and SIC category, is not available in any of the published macro data such as County Business Patterns, the U.S. Department of Commerce Bureau of Census, Enterprise Statistics or the Social Security Administrations Continuous Work History Sample. It is thus necessary to derive the raw data from other sources, all of which have been set up for purposes other than this. Four were potentially available to the researcher, three of which have been used elsewhere. They are:

1. Dun and Bradstreet Market Identifier File (DMI)(Birch [1979]).

2. Unemployment Insurance ES202 (U.I.) data (Teitz, Glasmeier, Svensson [1981]).
3. The telephone directory (Shapero and Giglierano [1982]).
4. Listings of the local Chamber of Commerce.

However, all have a number of limitations.

- a. **Classifying a firm:** Not only can the legal status of an organisation determine whether or not it is included, but also the type of unit reporting. This latter can be either an enterprise (U.S. Department of Commerce Enterprise Statistics) an establishment (Dun and Bradstreet Market Identifier File) or a group of establishments within a county (Unemployment Insurance ES202 (U.I.) data). For a full discussion of these comparisons, see Harris [1981].
- b. **Sector bias:** The nature of the data capture process may include an inbuilt sector bias. Thus the DMI file has, until recently, under-reported the service sector; the ES202 data excludes farmers; not all companies choose to record themselves in the Yellow Pages; the local Chamber of Commerce may show a bias towards the professional and service sector where the requirement for a contact network is prime.
- c. **Timeliness:** Although a firm may have been trading for some years, it may not appear in the DMI file until it seeks a credit rating. Moreover, whilst the file is constantly updated, each individual entry is checked infrequently. Therefore, there may be a serious time lag in picking up new firms. This is less of an issue when using the telephone directory which is updated annually or ES202, which is updated quarterly.
- d. **Job data:** All but one of the sources, the telephone director, include data on employment levels. However, these should be interpreted with care as they mostly cover only full-time, direct jobs whereas the new

small firm may create only indirect or part-time jobs. Consequently, the data may under-state the job generation effect of this group of firms.

**e. Availability:** The major attractiveness of both the DMI file and any local Chamber listing is the fact that the data is readily available, and, in the former case, on computer tape. Both the ES202 and the telephone directory require a physical scan in many instances and to pick out new firms on a national level is a physical impossibility but possible locally.

**f. Completeness:** The source must be as complete as possible within the boundaries defined. Thus, the Chamber listings and the Yellow Pages may be unsatisfactory since they are entirely voluntary. DMI files require the trigger of a credit rating. Because of the benefits of unemployment insurance, the ES202 are the most likely to cover the whole population. Table 2 summarises the apparent attractiveness of each data source.

Table 2

	DMI File	ES202	Telephone Directory	Chamber Listings
Reporting Unit	x	✓	~	x
Sector Bias	x	✓	✓	~
Timeliness	x	✓	✓	~
SIC	✓	✓	x	~
Job data	✓	✓	x	x
Availability	✓	x	x	✓
Completeness	x	✓	~	x

✓ Attractive, x Unattractive, ~ Indifferent

### THE DATA FOR ST. JOSEPH COUNTY

As a result of the above analysis the source chosen for the St. Joseph County study was taken from the Indiana State Employment Security Division ES202 (UC-1). This collects quarterly from all employers the number of employees who "worked or received pay for the pay period including the 12th of each month." It should be noted that this includes "employees in all types of payrolls."

The collection form used is the "reporting unit." These are firms which may have more than one establishment either in another industry or another county. However, as Teitz remarks, "the common usage of (the term) firm is more comfortable and the distinction is of major importance only for large businesses."

The data was collected by scanning the second quarter records, year on year. Firms were noted if they were:

1. present in year n and not in year n+1  
- LOSS
2. not present in year n and present in year n+1  
- GAIN

For each firm, the name, SIC category and employment level were recorded.

Several entries however showed a zero employment level for which subsequent analysis revealed four categories:

1. The firm had ceased trading but had not recorded the fact.
2. The firm had recently started but no employee had, as yet, received a regular salary. This included the owner.
3. The firm was in a seasonal business with a subsequent seasonal employment pattern. This was the reason for taking second quarter

data since in the summer months, most seasonal firms are likely to be active.

4. All the employees are free-lance outworkers.

In fact, only those in category 1, the still-born companies, would be of no future interest and therefore all zero entries were inspected six months later and those no longer showing in the files were excluded from the data. The rest were active, new, firms. The data base was also inspected for those firms which were known to be subsidiaries or branches of larger firms. Whilst the latter were of interest since they are new firms to the county, they would not subsequently be reported as entirely new firms.

#### COMPARISON WITH OTHER DATA SOURCES

To check the completeness of this data, and that of the potential other sources, the data for all new firms in the area between 1977 and 1982 was compared with that drawn from both the telephone directory and from the Dun and Bradstreet Market Identifier file (DMI). The Chamber listings were not included as they were acknowledged to be very inadequate.

Table 3 shows the data for St. Joseph County as it was 'cleaned'.

Table 3  
St. Joseph County Data--New Firms Since 1977

SIC Category	Scan 1. ES202	Scan 2. ES202	Scan 3. Telephone Directory
Agriculture	24	22	20
Manufacturing	101	94	73
Service	526	479	407
Construction	170	153	97
Finance	111	97	69
Transport	64	61	37
	996	906	703

Scan 1--Scan from ES202 comparing 1977 records with 1982 records.

Scan 2--Those remaining from scan 1 after zero entries were checked 6 months later, plus those found not to be born within the period, or those not for profit.

Scan 3--Those remaining after checking the telephone directories for 1981, 1982, and 1983.

Observation of Table 3 shows that when the data was cleaned for errors due to incorrect inclusion on the first scan, only a 9% potential error-rate was identified. Thus, scan 2 is the best estimate of the population of new firms, within the time period studied, from this data source.

However, use of the telephone directory (see Scan 3) across all the categories studied to identify the location of the potential new firms showed a severe erosion rate as shown in Table 4. In each case, the directories for 1981, 1982, and 1983 were used as well as both the Yellow Pages and Private Sections.

Table 4

	% Erosion: Scan 3/Scan 2
Agriculture	91
Manufacturing	78
Service	85
Construction	63
Finance	71
Transport	61
TOTAL	77

The results show that, overall, any study which uses this base as a source of new firms can expect to have missed more than 20% of the population. Clearly these firms exist, since they have registered with ES202 and so it is important to understand the reasons for their failing to use this medium to advertise a corporate telephone number. By the nature of their business, some will have chosen to be ex-directory; some, particularly those relying on word of mouth or subcontract work, such as in the construction industry, will consider it unnecessary; and some will merely use their private number.

The other potential data source was the DMI files. These were searched for all firms born between 1977 and 1982 in St. Joseph County. Table 5 shows the summary data supplied and compares it to Scan 2.

Table 5

	Scan 2	DMI files	% Erosion: DMI/Scan 2
Agriculture	22	1	4.5
Manufacturing	94	52	55
Service	479	31	6.5
Construction	153	14	9
Finance	97	6	6
Transport	61	6	10
Wholesale	-- <sup>1</sup>	23	--
Retail	-- <sup>1</sup>	70	--
	906	203	22

<sup>1</sup>Did not collect.

The erosion rates are startling, demonstrating clearly the caution expressed by other writers in using this as a source of data for company births. For the study in St. Joseph County, 78% of the population would have been missed, and omitting the wholesale and retail sectors to give a true comparison, increases the percentage missed to 82%.

Moreover, as expected, closer inspection of the data showed further errors emerging. Of the 203 firms identified, only 110 were comparable with our data base, wholesale and retail being excluded. Of these, 27 were subsidiaries and 24 were shown by the DMI file as being born since 1977 but were registered in the ES202 file prior to that date. Of the 59 remaining, 4 were missed by the ES202 search and 21 were not registered in the ES202 files for any year indicating that they were part of a larger reporting unit in the county. Only 34 were common, showing a final erosion rate when comparing the cleaned files of DMI with ES202 of 4% or 96% missed.

### CONCLUSION

This paper has explored the methodological problems associated with collecting data on the new firm population. It has examined the various sources

available and has shown that the major variable to affect the reliability of the data base identified is the source itself. Thus, the most satisfactory source is the Unemployment ES202 records, second the telephone directory and third, by a long way, the Duns Market Identifier files. Clearly any interpretation of the results taken from the latter two sources may be overwhelmed by the inadequacy of the source. Moreover, any subsequent refining of the data, whilst important, is likely to have minor effect.

66:1

## References

- Altman, E. I., "Why Businesses Fail." Journal of Business Strategy (Spring 1983).
- Begin, C.; Cesta, J.; Apilado, V., "Business Failures, Problematical Analysis, and the Question of Data Adequacy." Review of Business & Economic Research, 15, 1 (Fall 1979).
- Binks, M.; Jennings, A., "New Firms as a Source of Industrial Regeneration." Proceedings of UK Small Business Research Conference (August 1983).
- Birch, D., "The Job Generation Process." MIT Program on Neighborhood and Regional Change (Cambridge, Mass., 1979).
- Birley, S., New Enterprises: A Start-up Case Book, Croom Helm (1982).
- Birley, S., "An Educational Design for Entrepreneurs: The New Enterprise Programme." Working Paper. Notre Dame, 1983.
- Cooper, A., "Strategic Management: New Ventures and Small Business." Journal of Long Range Planning 14 (1981).
- Cross, M., New Firm Formation and Regional Development, Gower (1981).
- Fothergill, S. & Gudgin, G., "The Job Generation Process in Britain." Centre for Environmental Studies, Research Series No. 32 (November 1979).
- Harris, C., "A Comparison of Employment Data for Several Business Data Sources: County Business Patterns, Unemployment Insurance and Brookings U.S. Establishment and Enterprise Micro Data File." Working Paper No. 5, Small Business Micro Data Project, The Brookings Institution (October 1981).
- Harris, C., "Small Business and Job Generation: A Changing Economy or Differing Methodologies?" Working Paper, The Brookings Institution (February 1983).