

What Would an Exemplary Entrepreneurship Dataset Look Like?

Gary Paul Green

Department of Rural Sociology

Center for Community & Economic Development

University of Wisconsin-Madison/Extension

and

Greg Wise

Center for Community & Economic Development

University of Wisconsin-Extension

Paper presented at the RUPRI-ERS conference on *Exploring Rural Entrepreneurship: Imperatives and Opportunities for Research*, October 26-27, 2006, Washington, DC.

What Would an Exemplary Entrepreneurship Dataset Look Like?

Abstract

Much of the research on entrepreneurship has relied on “second best” data sets. In this paper, we outline the characteristics of an exemplary entrepreneurship dataset. The primary unit of analysis in the dataset is the firm, although it would be possible to trace the path of individual entrepreneurs as they move across firms. An exemplary data set would be sufficiently large to examine entrepreneurial activity over time and space, as well as between racial/ethnic groups and gender. There is a need to understand the effects of regional and local policies on entrepreneurship, which means that contextual data (political, economic and social characteristics) could be linked to firm data. Similarly, the data set could make important contributions to the research on the role of family and social networks in entrepreneurship.

What Would an Exemplary Entrepreneurship Dataset Look Like?

Introduction

Entrepreneurship is a rapidly growing field that cuts across several disciplines. As is often the case, the interdisciplinary nature of the concept makes it difficult to develop a precise definition. Most definitions include several key components, including the presence of innovation, risk-taking and opportunity-oriented business ventures. An even more difficult task has been to link these concepts with an empirical analysis. In most cases, research has been limited by the lack of adequate data to examine entrepreneurship. Researchers, instead, have had to rely on existing data that do not provide good indicators of entrepreneurial actors, firms or activity. Much of the empirical work on entrepreneurship has been conducted with “second best” data.

In this paper, we discuss these obstacles and identify the elements of an exemplary data set for studying entrepreneurship. Rather than beginning with the available data, we believe it is preferable to begin with a clear conceptual view of entrepreneurship and develop appropriate indicators of these concepts. In addition to defining entrepreneurship and identifying the necessary data elements to measure it, we discuss several other attributes to an ideal data set that would contribute to our understanding of entrepreneurship. In particular, we focus on political and economic factors influencing entrepreneurship and some of the consequences of these activities.

Defining Entrepreneurship

Most definitions of entrepreneurship focus on identifying market gaps, providing new services or goods and expanding employment. Although there is considerable overlap between the definitions of small business ownership, self-employment and entrepreneurship, they are not the same. Entrepreneurship focuses much more on innovation and risk-taking in markets. Most of the data that are available are much more appropriate for studying small business ownership and self-employment than for entrepreneurship.

One of the most popular data sets on entrepreneurship is the State Unemployment Insurance Administrative Files (ES202). The ES202 files offer one of the most comprehensive data sets on new business formation. Employers are required to provide the data, so it provides a population rather than a sample of firms. Yet, these data have several limitations for studying entrepreneurship. Each state has different rules about the data, which makes it very difficult to compare across states. In many states it is difficult for researchers to access the data due to confidentiality concerns. Finally, ES202 data do not include information on the self-employed.

Research has often made use of the decennial census to examine entrepreneurship and self-employment (Boyd 1990; 1991). Census data, especially the public use samples, has several advantages. It is a large data set that allows researchers to examine rates of self-employment across geographic areas, as well as among racial and ethnic groups. The data do not provide any information on the establishment, which makes it difficult to

study entrepreneurship. In addition, the data do not accurately provide indicators of entrepreneurship.

Other researchers have used the U.S. Establishment and Enterprise Microdata (USEEM) files developed by the Brookings Institution and formerly managed and maintained by the U.S. Small Business Administration. The data are useful for examining change in employment and job creation, but have a limited use for studying entrepreneurship.

There are few available longitudinal data sets to study entrepreneurship. The Panel of Study of Income Dynamics (PSID) provides some basis for studying businesses over time. It is almost impossible to link IRS data to this data set and it has a very limited basis for examining entrepreneurship across geographic areas.

Most available data sets omit or undercount firms that have no employees or do not have a business address or have negligible sales. Our strategy for defining entrepreneurship is to include several different indicators at the individual firm level. Some of the key indicators include high growth in revenues, number of patents or investments in research. This approach recognizes the potential multidimensional nature of entrepreneurship. We propose to begin with a firm survey that collects information on these attributes of the firm.

In addition, we need to identify entrepreneurial firms that are just in the startup phase, recently starting their business in the last few years. The limitations of the existing data sets can best be overcome by collecting new firm level data on entrepreneurship. We begin with the firm survey, but need to be able to track individual entrepreneurs, as well as measure entrepreneurial activity. In the following, we review

some of the issues identified in the literature that could/should be addressed in this data set.

One of the persistent problems in the literature has been the difficulty in defining entrepreneurship. The most common definitions include new business formation, new business growth and self-employment. Some more narrow definitions include technology/knowledge industry growth, patents and research grants activity and initial public offerings. These latter criteria may be too specific or narrow, but they do tap into some of the important elements of entrepreneurship we think should be considered.

Many researchers have limited the definition of entrepreneurship to “gazelles,” which are fast growing firms. Gazelles usually have a few key characteristics: 1) they spot market opportunities and move quickly to take advantage of them; 2) they are constantly improving productivity; and 3) their organizational structure adapts to the changing market situation. Birch defines a gazelle as a firm with a high rate of growth (more than 20%) over five years and a base revenue of more than \$100,000.

Goetz and Freshwater (2001) define entrepreneurial activity using two measures: initial public offerings offered in the technology sector per million population and *Inc.* 500 firms per million population. The *Inc.* definition ranks the 500 fastest growing firms.

Sarah Low (2004) distinguishes between the breadth and depth of entrepreneurship. Breadth refers to the dispersion of entrepreneurs, and depth is considered the concentration of high-value entrepreneurship in a region. Low measures entrepreneurial breadth as the ratio of self-employment to total employment. Two different measures of entrepreneurial depth are considered: the ratio of self-employment to the number of self-employed workers and the share of self-employment income to self-

employment receipts. The latter measure is an indicator of entrepreneurial value added. Low finds that entrepreneurial breadth is especially high in rural areas and entrepreneurial depth is highest in metropolitan areas.

Our approach will be to develop key measures of entrepreneurial activity, including innovation, risk-taking and opportunity-oriented business activity. It may be the case that entrepreneurship is a multidimensional concept that cannot be adequately measured with a few indicators.

Data Needs

In the following section we outline an exemplary data set for studying entrepreneurship. The data set should meet several basic research needs. First, it should enable researchers to develop valid and reliable measures of entrepreneurial activity. Preferably, the indicators would provide some evaluation of different levels and kinds of entrepreneurial activity. One of the difficulties is developing indicators that are appropriate and valid for firms in a wide variety of industries. Some studies have limited their analysis of entrepreneurship to high-growth industries, but this approach seems too limiting.

Second, the data set should be large enough to examine the breadth of entrepreneurship—the level of entrepreneurship across geographic areas. As we will discuss below, one of the important research needs in this area is to provide an understanding of how the entrepreneurial climate and public policy decisions influence the level of entrepreneurship. To conduct these analyses, we need a sufficient number of firms for each geographic unit to conduct a hierarchical analysis (approximately 25 per

unit). Thus, the sample of entrepreneurs will most likely be a cluster, rather than a random sample. Ideally, the data set would permit researchers to aggregate firm data at several different levels: place, county and state. Given the nested nature of many public policies influencing entrepreneurship, it would be advantageous to consider the effects of these different levels of policy and support.

Third, it is preferable to build a longitudinal data base that would allow us to track entrepreneurial firms over time. There are numerous issues that could be examined here. Longitudinal data would provide some insights into when technical assistance and other programs would have the greatest impact on entrepreneurs. We also need a better understanding of the life-cycle of entrepreneurs. There is a substantial body of literature that shows that small business owners often go through several unsuccessful businesses. We do not have a comparable understanding of the path entrepreneurs take over time.

Although the primary unit of analysis in this data base is the firm, the data should be flexible enough to allow us to trace individual entrepreneurs over time as well. As far as we know, there is very little literature on the paths of individual entrepreneurs. We need to be able to assess how new firms are spun-off from exiting firms. Similarly, it would be useful to track how individual entrepreneurs sell their firms to start up another firm. Of course, this type of analysis will raise important questions about what we would consider successful entrepreneurship. We often look at whether a firm is surviving and growing. It may be useful to consider the possibility that entrepreneurs may make several unsuccessful attempts before they are successful. Thus, investments in entrepreneurs may not “payoff” until much later and in a different form. Building a

flexible data base among firms and individual entrepreneurs would enable us to examine these issues in more detail.

Finally, the longitudinal data base would enable researchers to evaluate the extent to which entrepreneurship becomes routinized. There are several possible scenarios here. Once entrepreneurs have been successful in identifying a new opportunity, they may turn to improving the processes and structures within the firm. Alternatively, they may hand-off these activities and turn to identifying new opportunities. It is important to understand how entrepreneurial firms handle this tension between these competing needs.

Fourth, the literature on entrepreneurship has emphasized the racial, ethnic and gender differences among entrepreneurs (Aldrich and Waldinger 1990; Aldrich, Waldinger and Ward 1990; Brophy 1989; Light and Bonacich 1988; Tigges and Green 1994). The sample of entrepreneurs needs to be stratified by race, ethnicity and gender, so as to compare breadth and depth of entrepreneurship among these groups. Some of the qualitative research on immigrant entrepreneurs has found that many begin in neighborhoods with high levels of immigrants. Eventually, the firms move out to other areas. There is not much quantitative analyses of how immigrant entrepreneurs develop and grow.

Finally, much of the literature suggests that there are significant rural and urban differences in entrepreneurship. Ideally, it would be helpful to collect data on entrepreneurship in both urban and rural areas. Rural areas often lack the infrastructure, especially business services and venture capital, which are so crucial to entrepreneurship. It is important to understand how these constraints affect rural entrepreneurship. Comparative data would enable us to examine some of the consequences.

The fundamental building block is firm level data. We need to develop reliable and valid measures of entrepreneurship at this level. Yet, as we mentioned earlier the data set needs to be flexible enough to provide information on entrepreneurial firms, actors and activity. Although we need to begin with firm-level data, it may be necessary to track individual entrepreneurs separately from firms.

Entrepreneurial Climate

After collecting the firm level data, we need to develop measures of entrepreneurial climate that can be attached to the firm data. Depending on the situation, we might append data from local, regional and state level data sets to examine the effects of business climate on entrepreneurship. Similarly, we can append aggregate measures of the multiplier effects of entrepreneurship for communities, such as employment and income growth. In these analyses, it would be possible to consider both the breadth and depth of entrepreneurship in communities.

Can the local and regional environment influence entrepreneurship? There is a substantial body of literature suggesting that there are several aspects to the entrepreneurial climate that should be considered. Bruno and Tyebjee (1982) suggest that the key environmental factors influencing entrepreneurship are access to venture capital, technical support, skilled labor, restricted regulations, low taxes and access to new markets. We consider two broad contextual/environmental influences on entrepreneurship: political and social/cultural influences.

Among the most important political factors are taxes, regulations and incentives/subsidies. There is considerable debate over the effects of these contextual

influences on economic development in general, and entrepreneurship more specifically (Bartik 1991). An exemplary data set would include contextual data on the tax rate, regulatory structure and governmental incentives for entrepreneurship in regions or localities. A contextual analysis would require that the data set include clusters of entrepreneurs in rural areas. This sampling strategy would allow researchers to examine the contextual effect on entrepreneurship. One of the difficulties in this body of research is comparison of tax rates across regions. The data set would need to adjust these rates for businesses so they are comparable.

There also are obstacles in measuring the effects of policies on entrepreneurship. Most analyses only look at whether or not a policy, such as tax abatement, exists in a locality. We need more detailed data on the policies, such as the real budget cost for local governments. These data are often difficult to obtain because it may not actually show up as a budget item (e.g., loans or foregone revenues). In-kind incentives are often difficult to measure. An ideal data set would allow us to measure policy adoption and total effort, including off-budget incentives (Green, Fleischmann and Kwong 1996). Finally, the data set should enable researchers to consider development policies and entrepreneurship as endogenous variables. Much of the literature on this topic considers this a causal relationship. Localities adopting economic development policies may be more distressed and have less entrepreneurial activity. The data set, then, should enable researchers to develop nonrecursive models.

There has been an increasing amount of attention given to the role that social factors might play in entrepreneurship. Much of this work has focused on racial and ethnic differences in entrepreneurship. This research has examined why some racial and

ethnic groups have much higher rates of self-employment than others. They have focused especially on the important role that social networks play in providing resources to entrepreneurs. These networks may provide financial resources, consumer markets and labor. One of the limitations of this research is that the data are often collected from a single racial or ethnic group and as a result there is not much comparative research on this issue. An exemplary data set will include some information on the social resources utilized by entrepreneurs, especially in their start-up and expansion stages.

Economists and sociologists have long recognized the importance of access to finance capital for entrepreneurship. Although the cost of credit is often an obstacle, the most critical issue is whether entrepreneurs have adequate access to credit. We need to distinguish between equity, debt and venture capital needs of entrepreneurs. It would be useful to provide data on the supply of equity, debt and venture capital for localities. In addition to measures of supply, it would be possible to assess access to capital by collecting information on rejection rates among various credit institutions in these localities.

To understand the demand for credit among entrepreneurs, it is important to assess whether they rely on informal sources of credit (family and friends), retained earnings or savings to fund their business. Much of the literature suggests that entrepreneurs rely heavily on informal rather than formal sources of credit. One of the central questions to be addressed is: do entrepreneurs turn to informal sources of credit because their business is too risky for formal sources, they have a poor credit history, or informal sources are cheaper and may be obtained under more favorable conditions? Thus, we need to obtain better data on the credit history among entrepreneurs and their

experience in formal capital markets. These data would include not only access to credit, but also the conditions in which the credit is available (interest rates, collateral required, etc.).

The literature suggests that there is an important role that higher education can play in promoting entrepreneurship. Higher education can play several roles. Many researchers take their innovations and spin-off to start new businesses. Many students trained at universities take their research to start new businesses. In addition, universities provide the skilled and trained workforce that many entrepreneurial firms will need. An exemplary data set would be able to track these connections between entrepreneurship and university research. We need a better understanding of these linkages and their effects on the breadth and depth of entrepreneurship.

Finally, we need a much better understanding of the link between the formal labor markets and entrepreneurship. An exemplary data set would enable researchers to explore how formal labor market conditions influence entrepreneurship. Some of the most important measures would include formal labor force participation, unemployment and underemployment rates. Similarly, we need data on the average wages and benefits for various industries in the region. When appended to the data on entrepreneurship, these data would allow researchers to examine whether entrepreneurship is strongly related to opportunities in formal labor markets.

Impacts of Entrepreneurship

One issue that has received very little attention in the literature is the quality of jobs provided by entrepreneurs. Much of the economic development literature has

focused on the job creation potential of entrepreneurs and small businesses, but they rarely look at the wages, benefits or opportunities for mobility among workers in these firms (for an exception, see Brown et al.). An exemplary data set will provide detailed information on the types of jobs created (skill, educational and experience requirements), wages and benefits offered, training provided and anticipated vacancies.

Ideally, we need to be able to identify the effects of entrepreneurship on employment and income growth in communities. In regional economics terms, we need a better understanding of the multiplier effect of entrepreneurship. Conventional models for estimating their effects may be limited given the rapid growth and change in many entrepreneurial firms. One strategy for estimating these effects would be to obtain detailed information on where and how entrepreneurs purchase their inputs--goods and services. One potentially interesting issue is whether entrepreneurial firms are identifying market opportunities in the local or regional economy, or whether they primarily develop goods and services for markets outside the community. These differences may have some important regional economic effects.

There are some claims that entrepreneurs in the high tech sector benefit from frequent interaction and sharing of important technology. This interaction is especially strong among firms in industrial clusters. We need a better understanding of how this informal contact is structured and how it affects the performance of entrepreneurial firms.

Finally, it would be useful to estimate the effects of expansion on the tax base. This issue is an important question on its own, but is especially critical when we consider the costs of public policies (such as tax breaks and subsidies) to promote

entrepreneurship. We can look at the benefits, but we also should take into consideration some of the costs the public incurs in promoting entrepreneurship.

To adequately evaluate the effects of entrepreneurship, we need a longitudinal data base that allows us to examine how entrepreneurs change over time. There are several research questions that could be addressed: Do entrepreneurs continue to innovate or do they shift strategies? What factors influence the strategies that entrepreneurs use over time? What is the best stage of development for policies to promote and support entrepreneurs?

In many cases, there is a strong overlap between entrepreneurial businesses and families. It is important, then, to understand the importance of the life course to entrepreneurial families. The life course perspective focuses on the important interplay between the lives of family members, the life course of a family business and the links between the two (Moen 1998).

Entrepreneurship is considered a key element of the economic development strategy for rural communities. During much of the 1980s and 1990s, states and localities developed a wide variety of programs to aid small businesses and entrepreneurship. Some research has suggested that the evidence on the effects of capital and technical assistance programs on small business development and entrepreneurship is relatively weak (Chrisman, Nelson, Hoy and Robinson 1985; Mokry 1988). Rural areas may be especially disadvantaged in building an entrepreneurial climate due to the constraints related to size, density and distance. An exemplary data set for entrepreneurship would allow us to examine the effectiveness of economic development strategies for entrepreneurship.

Some Final Notes

To better understand entrepreneurship we need more valid measures of entrepreneurial activity. Existing data sets are very limited in their ability to assess the depth and breadth of entrepreneurship. Federal, state and local governments are investing an increasing amount of resources to support and promote entrepreneurship, but the social science evidence on the effects of these investments is thin and weak. One of the key elements of this exemplary data base would be examining entrepreneurial activity over time.

Establishing a longitudinal data base on entrepreneurship has several advantages beyond the substantive issues discussed here. Over time, we would be able to provide a better understanding of the factors contributing to the birth, growth, decline and death of entrepreneurial firms. There are few longitudinal data sets on organizations that would permit this type of analysis. It is an essential feature to understanding the dynamics of entrepreneurial firms.

References

- Aldrich, H.E., and R. Waldinger. 1990. "Ethnicity and entrepreneurship." *Annual Review of Sociology* 16: 111-35.
- Aldrich, H., R. Waldinger and R. Ward. 1990. *Ethnic Entrepreneurs*. Newbury Park, CA: Sage Publications.
- Bartik, T.J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute.
- Bates, T. 1993. "Theories of entrepreneurship." Pp. 248-264 in *Theories of Local Economic Development*, R.D. Bingham & R. Mier (eds.). Newbury Park, CA: Sage Publications.
- Birch, D. 1987. *Job Creation in America*. New York: Free Press.
- Boyd, R.L. 1990. "Black and Asian self-employment in large metropolitan areas: a comparative view." *Social Problems* 37: 258-74.
- _____. 1991. "A contextual analysis of black self-employment in large metropolitan areas, 1970-1980." *Social Forces* 70: 409-29.
- Brophy, D. 1989. "Financing women-owned firms." Pp. 55-76 in O. Hagan, C. Rivchun, and D. Sexton (eds.). *Women-Owned Businesses*. New York: Praeger.
- Brown, C., J. Hamilton and J. Medoff. 1990. *Employers: Large and Small*. Cambridge, MA: Harvard University Press.
- Bruno, A.V., and T.T. Tyebjee. 1982. "The environment for entrepreneurship." Pp. 288-315 in *Encyclopedia of Entrepreneurship*, C.A. Kent (ed.). Lexington, MA: D.C. Heath Company.

- Chrisman, J.J., R.R. Nelson, F. Hoy, and R.B. Robinson. 1985. "The impact of SBDC consulting activities." *Journal of Small Business Management* 23: 1-11.
- Eisinger, P.K. 1988. *The Rise of the Entrepreneurial State*. Madison: University of Wisconsin Press.
- Frederick, M. and C.A. Long. 1989. *Entrepreneurship Theories and Their Use in Rural Development*. Agriculture and Rural Economy Division, Economic Research Service, U.S. Department of Agriculture. Bibliographies and Literature of Agriculture Number 73. Washington, D.C.: U.S. Department of Agriculture.
- Goetz, S.J. and D. Freshwater. 2001. "State-level determinants of entrepreneurship and a preliminary measure of entrepreneurial climate." *Economic Development Quarterly* 15: 58-70.
- Green, G.P., A. Fleischmann, and T.M. Kwong. 1996. "The effectiveness of local economic development policies in the 1980s." *Social Science Quarterly* 77:609-625.
- Leibenstein, H. 1968. "Entrepreneurship and development." *American Economic Review* 58: 72-83.
- Light, I. and E. Bonacich. 1988. *Immigrant Entrepreneurs: Koreans in Los Angeles, 1965-1982*. Berkeley: University of California Press.
- Lin, X., T.F. Buss, and M. Popovich. 1990. "Entrepreneurship is alive and well in rural America." *Economic Development Quarterly* 4: 254-259.
- Low, S. 2004. "Regional asset indicators: entrepreneurship breadth and depth." *The Main Street Economist* (September): 1-4.

- Moen, Phyllis. 1998. "A life course approach to the entrepreneurial family." Pp. 16-28
in *The Entrepreneurial Family*, Ramona K.Z. Heck (Ed.). Needham, MA: Family
Business Resources, Inc.
- Mokry, R. 1988. *Entrepreneurship and Public Policy*. New York: Quorum Books.
- Tigges, L.M. and G.P. Green. 1994. "Small business success among men- and women-
owned firms in rural areas." *Rural Sociology* 59: 289-310.