An Overview of the Kauffman Firm Survey

Results from the 2004–2007 Data

Prepared By:

Alicia Robb
Janice Ballou
David DesRoches
Frank Potter
Zhanyun Zhao
E.J. Reedy

April 2009





The Kauffman Firm Survey

2004 2005 2006 2007 2008 2009 2010 2011

KAUFFMAN
The Foundation of Entrepreneurship

An Overview of the Kauffman Firm Survey

Results from the 2004–2007 Data

Alicia Robb
Janice Ballou
David DesRoches
Frank Potter
Zhanyun Zhao
E.J. Reedy

April 2009

KAUFFMAN
The Foundation of Entrepreneurship

EXECUTIVE SUMMARY

Although entrepreneurial activity is an important part of a capitalist economy, data about U.S. businesses in their early years of operation have been extremely limited.[1] As part of an effort to gather more data on new businesses in the United States, the Ewing Marion Kauffman Foundation (the Foundation) sponsored the Kauffman Firm Survey (KFS), a panel study of new businesses founded in 2004 and tracked over their early years of operation. The KFS dataset provides researchers with a unique opportunity to study a panel of new businesses from startup to sustainability, with longitudinal data centering on topics such as how businesses are financed; the products, services, and innovations these businesses possess and develop in their early years of existence; and the characteristics of those who own and operate them.[2]

Results. The current data provide an understanding of how businesses are organized and operate in their first four years of existence (2004 through 2007) and provide some indicators of survival and growth. Other measures describe the characteristics of the panel, such as the extent to which these businesses are involved in innovative activities. A series of eleven tables give a broad overview of the business and owner characteristics and firm survival over the period, as well as some new information available in the third follow-up survey. Highlights include:

- External debt markets remain critically important for most new firms.
  - In the first year of operation, external debt markets provided the single largest source of financing. The new firms injected about $80,000 on average into their new ventures during the first year of operation. Outsider debt (bank loans, credit cards, credit lines, etc.) made up more than $32,000 of that total and was the single largest funding source.
  - Three years later, in 2007, surviving firms injected another $53,000 into their businesses. This amount is much lower than in 2004, but the percentage of financial capital raised from outside credit markets increased to 62 percent. Thus, the importance of external debt markets continues to rise as firms survive and grow in their early years.
- Only about 12 percent of firms submitted new external credit applications for debt financing in 2007.
  - For the vast majority of firms, the application(s) always were approved. Nearly 18 percent of firms had mixed results, sometimes approved and sometimes denied. Just over 12 percent of firms said their loan applications always were denied.
  - For those that had some or all of their loan applications denied, almost half said that one of the main reasons given was insufficient collateral to guarantee the loan. The second most-common reason was flaws in the owner's personal credit history.
  - Seventeen percent of firms said they didn't apply for credit at some point when they needed it because they feared their loan applications would be denied. However, this group that feared denial had a higher proportion of firms that applied (20.7 percent) than the sample overall (12 percent).

1. http://www.nap.edu/catalog/11844.html
2. A comparison of the KFS dataset with other business datasets along a number of dimensions is provided in Appendix C.

- Less than 10 percent of firms owned by African Americans applied for new credit in 2007, compared with nearly 13 percent of firms owned by non-Hispanic whites.
  - Women and African Americans were more likely than men and whites to state that they didn't apply for credit at some point when they needed it in 2007 for fear of being denied.
  - Nearly 40 percent of African Americans said they feared being denied, compared with 14.5 percent of whites.
- About 90 percent of firms that began operations in 2004 survived through 2005, while about 80 percent survived through 2006 and 73.4 percent through 2007. Most of the remaining firms closed either permanently or temporarily over the period, while a small number, 3.5 percent, either merged with or were sold to another business.
- Surviving firms with employees, which are now three years old, increased average employment from 4.6 employees in 2004 to 6.7 employees in 2007.
- By 2007, about 40 percent of firms had revenues greater than $100,000, compared with just 17 percent in 2004.
- About 60 percent of surviving firms posted profits in 2007, compared with about 40 percent posting losses.
- Young firms are serving local and international markets.
  - About 12 percent of firms serve a predominantly local market, in or around the neighborhood where the business operates. Nearly a third of firms had a market area that is made up of the city or county where the business operates. More than a third had a statewide or regional market, while about 17 percent of firms had a national market. Only about 3 percent of firms stated that their main market was international. However, more than 13 percent of firms had some international sales.
  - High-tech firms were much more likely to have broader markets for their products and services. Nearly half of high-tech firms had nationwide or international markets as their main market and more than one-third of high-tech firms had some international sales. For those high-tech firms that had international sales, about 15 percent had international sales that made up at least half of their total sales.
- More than a quarter of firms sold at least some of their goods or services on the Internet. Nearly a quarter of those firms had Internet sales that were more than half of their total sales, while about a third of them said Internet sales were less than 5 percent of their total sales.
- About 60 percent of firms felt that they had a comparative advantage in the products or services that they offered.
  - About a quarter of those firms felt that the comparative advantage stemmed from teaming up with another company, while only 2 percent felt it was due to teaming up with a government lab or research center.
  - About 6 percent felt their comparative advantage was due in part to teaming up with a college or university, while 7 percent felt it was due in part to patents that the firm owned, had applied for, or had licensed.

Further analysis is available in a series of papers that are posted to the KFS section of the Ewing Marion Kauffman Foundation Web site as they are completed (http://www.kauffman.org/kfs/). Four papers are available; others will be added as they become available:

- The Kauffman Firm Survey: Results from the Baseline and First Follow-Up Surveys http://www.kauffman.org/uploadedfiles/kfs_08.pdf
- The Capital Structure Decisions of New Firms http://www.kauffman.org/uploadedFiles/Capital_Structure_Decisions_New_Firms.pdf
- Characteristics of New Firms: A Comparison by Gender http://www.kauffman.org/uploadedFiles/kfs_gender_020209.pdf
- Patterns of Financing: A Comparison between White and African American Young Firms http://www.kauffman.org/uploadedFiles/kfs_black_firms.pdf

Data Availability. The Kauffman Firm Survey is a research data set accessible to scholars around the globe. The public-use microdata file for the Kauffman Firm Survey, which contains data from the Baseline, First, Second, and Third Follow-up Surveys, is available at http://sites.kauffman.org/kfs/request_download.cfm. The dataset can be downloaded in SAS, STATA, or SPSS. Researchers wishing to access a more detailed data file and to engage with a community of researchers in analysis of the KFS should consider applying for access to the University of Chicago NORC Data Enclave. The NORC Data Enclave provides secure remote access to the KFS confidential microdata file, which contains more detail regarding industry codes, geographical codes (zip code, metropolitan statistical area, and state), firm credit scores, and many additional continuous variables (in addition to categorical variables). The KFS confidential microdata may only be accessed through the NORC Data Enclave. Details on applying can be found on the KFS Web site: http://www.kauffman.org/kfs

KFS Design. The study created the panel by using a random sample from the Dun & Bradstreet (D&B) database list of new businesses started in 2004. In response to the Foundation's interest in understanding the dynamics of high-technology businesses, the KFS oversampled these businesses based on the intensity of research and development employment in the businesses' primary industries.

Mathematica Policy Research, Inc., conducted extensive questionnaire design activities to establish consistent definitions of what constituted a new business and the start of business operations, and to investigate the most efficient methods for collecting these data. The KFS sought to create a panel that included new businesses created by a person or team of people, purchases of existing businesses by a new ownership team, and purchases of franchises. To this end, the KFS excluded D&B records for businesses that were wholly owned subsidiaries of existing businesses, businesses inherited from someone else, and not-for-profit organizations. Also, previous research on new businesses has reported variability in how business founders perceive when their businesses started operations. Therefore, a series of questions was asked about indicators of business activity and whether these were conducted for the first time in the reference year (2004). These indicators included:

- Payment of state unemployment (UI) taxes
- Payment of Federal Insurance Contributions Act (FICA) taxes
- Presence of a legal status for the business
- Use of an Employer Identification Number (EIN)
- Use of Schedule C to report business income on a personal tax return

To be "eligible" for the KFS, at least one of these activities had to have been performed in 2004 and none performed in a prior year.

The questionnaire covered a variety of topics, including business characteristics, strategy and innovation, business structure and benefits, financing, and demographics of the principals.

Data Collection Methodology. The Baseline Survey was conducted between July 2005 and July 2006. Interviews were completed with principals of 4,928 businesses that started operations in 2004, which translates to a 43 percent response rate when the sampling weights are applied. A self-administered Web survey and Computer-Assisted Telephone Interviewing (CATI) were used for the data collection, and KFS respondents were paid $50 to complete the interview. CATI completes accounted for 3,781 (77 percent) and Web completes accounted for 1,147 (23 percent) of the total interviews. The results across sampling strata show that 2,034 interviews were completed in the two high-technology strata (See Appendix A for more information about the sampling strata), and the remaining 2,894 interviews were completed among non-high-tech businesses.

The sample for the First Follow-Up Survey consisted of the 4,928 businesses that completed the Baseline Survey. The First Follow-Up was conducted between June 2006 and January 2007, and 3,998 interviews were completed, which translates to an 89 percent response rate after adjusting for the sample weights. During the First Follow-Up, a significantly larger percentage of interviews were completed through the Web survey (2,366 or 59 percent) than in the Baseline, with CATI

completes accounting for 41 percent (1,632 interviews).

Data collection on the Second Follow-Up Survey closely mirrored that of the First Follow-Up. Data collection began on May 31, 2007, and concluded on December 1, 2007. Overall, the study continued to be successful in retaining panel businesses, achieving a response rate of 84 percent (weighted). There was a slight increase in the percentage of respondents who completed the Web survey, (63 percent in the Second Follow-Up compared to 59 percent in the First Follow-Up). Because the Second Follow-Up Survey was the third annual survey in which KFS panel members were asked to participate, KFS respondents usually remembered the previous surveys and required little persuasion. Nonetheless, there were some refusals, which necessitated a refusal conversion effort. Of the 4,523 cases in the Second Follow-Up, 404 initially refused, of which 66, or 16 percent, were converted and completed the questionnaire.

The data collection for the Third Follow-Up began on June 24, 2008, and concluded on December 23, 2008. About two-thirds of the 2,915 respondents chose to answer the survey by Web, while about one-third answered by CATI. A 78 percent response rate (unweighted) was achieved. Several new questions were added on sources of comparative advantage, credit applications and loan turndowns, predominant market for the firm's products and/or services, international sales, and Internet sales. Additional details of the study design are available in the introduction as well as the appendices.

Chapter 1

INTRODUCTION TO THE KAUFFMAN FIRM SURVEY

Entrepreneurship plays an important role in the country's economic activity, and accurate information about the development and sustainability of new businesses is essential to establishing public and private programs that encourage new business development. However, obtaining accurate information on new firm dynamics is difficult. Surveys of new businesses tend to be hard to implement and typically have produced low response rates because of the difficulty of obtaining the cooperation of new business owners. Surveys of new businesses also have faced the complexities of defining what constitutes a new business and when a new business begins operations, events that lend themselves to subjectivity if not carefully defined. Further, few previous business surveys collected information about the dynamics of business development, because longitudinal surveys of new businesses faced the issue of business attrition. As part of its mission to advance entrepreneurship and the study of new business creation and development, the Ewing Marion Kauffman Foundation (the Foundation) sponsored the Kauffman Firm Survey (KFS).

A. STUDY OBJECTIVES

The main objective of the KFS is to address the informational gaps related to the study of entrepreneurship. Because of the Foundation's

commitment to providing researchers and policy decision makers with the best possible information about new business creation and sustainability, Mathematica Policy Research, Inc., (MPR) was commissioned to design and conduct a rigorous survey to understand entrepreneurial patterns by gathering information from newly formed businesses. In particular, the goal of the KFS was to learn more about the development of high-technology and women-owned businesses, the financial experiences of new businesses, and the business characteristics that are indicators of sustainability. In addition, the KFS was designed to meet the information needs of as many of the potential data users as possible. To begin the KFS development and design process, an initial meeting was held in May 2004 with a core advisory group. Based on this initial meeting and subsequent discussions, researchers agree on the following concepts to frame the development of the KFS: (1) the data collected would be relevant to a "pure" cohort of businesses that started in a single targeted year, (2) the business—not any individual owner or founder—would be the focus of the information collected, (3) financial information related to business formation would be the main analytic objective, and (4) a longitudinal survey design would be needed to inform an understanding of business development dynamics and sustainability. To achieve these objectives, researchers used a deliberate and inclusive process to address the methodological challenges related to finding and identifying businesses that qualified for the survey, develop questionnaire items that accurately measured the key concepts related to business development, and achieve survey participation of these businesses.

B. KFS DESIGN AND DEVELOPMENT

A comprehensive and collaborative process was used to design and develop the sample questionnaire, and survey operations for the KFS.

1. Literature Review and Advisory Group Consultation

Two initial actions were employed to inform the design process and to test the validity of the assumptions in the proposed research: (1) a review of business and other relevant literature and (2) consultation with an advisory group composed of probable KFS data users. The literature review included about sixty articles and related surveys that focused on business statistics and the dynamics of business formation. In particular, survey instruments from the Economic Census, the Survey of Small Business Finance, and the Panel Study of Entrepreneurial Dynamics were included in this review.

More than twenty technical advisors contributed to the development of the KFS. These advisors were selected because of their interest, expertise, and scholarship related to entrepreneurship. In addition, it was expected that the core users of the KFS data files would be among this group, so they were given an opportunity to inform the process. Also, the advisory group outlined a "wish list" of information that would best meet the needs of academic researchers, members of government agencies, and public policy decision makers who would use the KFS data.[3]

Based on a review of the literature and analysis of prior business surveys, in addition to consultation with entrepreneurial experts, multiple methodological and conceptual topics emerged that needed to be researched prior to conducting the survey. These included assumptions about the sample design, eligibility criteria for participation, incidence of eligible new businesses, and questionnaire items. A design phase was included in the KFS process to provide information to address these topics. For more detailed information about the KFS technical advisory group and other design activities, see the *Kauffman Firm Survey Baseline Methodology Report* available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1024045.

2. Pilot Tests

Critical to the KFS was the definition of a new business as envisioned by the Foundation and the technical advisory group, and how this definition matched up with the sample frame from Dun & Bradstreet (D&B). Two pilot tests were conducted

3. During the same time period as the KFS development, the Foundation also funded a major effort by the National Academies of Science (NAS) to identify the information needs related to business surveys and other information needed to study new business development. This effort resulted in the publication of *Understanding Business Dynamics: An Integrated Data System for America's Future.* Several NAS participants were also KFS advisors.

because little was known about the incidence of the proposed eligibility criteria. The August 2004 pilot test was used to identify the incidence of two criteria being considered for the definition of a new business: (1) state unemployment insurance (UI) payments and (2) Federal Insurance Contributions Act (FICA) tax payments made for the first time in the targeted year. The 20 percent incidence of businesses reporting making either UI or FICA payments for the first time during the reference year of 2003 was lower than expected. The project team also considered the potential bias related to using UI and FICA payments exclusively, since these measures are associated with having employees and would result in an underrepresentation of non-employer, single-owner businesses. For these reasons, the eligibility criteria were expanded during a second pilot test. This test assessed additional eligibility criteria, including (1) legal business status (sole proprietorship, general partnership, limited partnership, C-corporation, subchapter S-corporation, and limited liability company), (2) acquisition of an Employer Identification Number (EIN), and (3) use of an Internal Revenue Service Schedule C or C-EZ as part of the owner's income tax return. These criteria yielded a 36 percent incidence. Overall, 52 percent of the owners included in the pilot tests would have met the eligibility screening on at least one of the criteria tested at that time. Based on these results, a new business eligible for the KFS targeted year was defined as any business responding positively to any one of the five tested criteria.

3. Questionnaire Development and Pretesting

A comprehensive and iterative process was used to develop the final questionnaire. The initial draft KFS questionnaire was crafted using the matrix of topics suggested by the advisors and relevant questionnaire items from prior studies. An explicit goal suggested by the advisors during the design process was harmonization of the KFS with other business surveys. Using the initial draft questionnaire, cognitive interviews were conducted with eligible new business owners to evaluate the survey instrument. Following this developmental research, a comprehensive pretest of 400 new businesses was conducted to (1) test the questionnaire length; (2) review response distributions, missing and inappropriately skipped questions, and incomplete questionnaires; and (3) perform several methodological experiments.

C. OVERVIEW OF KFS SURVEY METHODOLOGY

The KFS Baseline Survey was conducted from July 2005 to July 2006 using both Computer-Assisted Telephone Interviewing (CATI) and self-administered Web questionnaires. Overall, 4,928 questionnaires from eligible new business owners were completed, for a response rate of 43 percent after the sampling weights were applied. Following is an overview of the survey methodology. Additional details on the data collection methodology can be found in Appendix B.

1. Sample Design

The KFS target population was all new businesses included in the D&B database and reported by D&B as having started in 2004. As described earlier, the definition of an eligible "starting" business was developed in the KFS design process. The D&B database was partitioned into six sampling strata defined by a classification of the firm's high-technology status and the gender of the firm's owner or CEO (based on the D&B data element). The high-technology strata were defined based on the categorization developed by Hadlock, et al. (1991). Overall, 32,469 businesses were sampled to achieve 4,928 completed questionnaires. Additional details on the sample design can be found in Appendix A and also in the *Kauffman Firm Survey Baseline Methodology Report.*

2. Eligibility Screening and Questionnaire Content

The KFS Baseline questionnaire was developed using the matrix of topics suggested by the advisors and refined during pretesting. The questionnaire has two main sections: (1) questionnaire items used to determine business eligibility and (2) modules to obtain information about the business. The modules included business characteristics, strategy and innovation, business organization and human resource benefits, business finances, and work behaviors and demographics of owner-operator(s).

Because there were two modes of data collection, CATI and self-administered Web, the questionnaire was customized to maximize the advantages of each mode of data collection while minimizing possible mode effects.

3. Data Collection

Data collection on the Baseline Survey involved extensive preparation and coordination to contact the 32,469 businesses that were sampled to determine eligibility. The process began with a mailed advance letter to sampled businesses inviting participation using the KFS self-administered Web questionnaire and informing them that eligible businesses would receive a $50 incentive when the questionnaire was completed. Following the invitation, business owners who did not complete the questionnaire on the Web received telephone calls from trained interviewers to determine their eligibility and to complete an interview with those that were eligible. Overall, 77 percent of the Baseline Survey questionnaires were completed using CATI, and 23 percent were completed using the self-administered Web questionnaire.

The First Follow-Up Survey was conducted among the 4,928 businesses completing the Baseline Survey. Respondents were contacted initially by e-mail and invited to complete the KFS Web survey. Those businesses without e-mail addresses or those not completing the Web survey were contacted by mail, similarly to those in the Baseline Survey. Respondents again were paid $50 after completing the survey. The Baseline Survey had established eligibility for all businesses in the panel; therefore, the only eligibility criterion for the First Follow-Up was whether the business was still in operation. Of the 4,928 completes at Baseline, 369 were identified as out of business during the First Follow-Up (7.5 percent of the total sample). Of the remaining eligible businesses, 3,998 interviews were completed, resulting in a response rate of 89 percent after the sampling weights were applied.

The Second Follow-Up was conducted among the 4,523 businesses remaining in the KFS panel. The Baseline Survey had established eligibility for all businesses in the panel; therefore, the only eligibility criterion for the Second Follow-Up was whether the business was still in operation. Specifically, the remaining businesses were those that completed the Baseline Survey (4,928), minus those identified as out of business (369) and those that had adamantly refused to participate in previous rounds (36). Panel members were contacted initially by e-mail and invited to complete the KFS Web survey. Those businesses without e-mail addresses or those not completing the Web survey after being contacted by e-mail were contacted by U.S. mail. As in prior waves, respondents were paid $50 for completing the Second Follow-Up. With the remaining eligible businesses, 3,390 interviews were completed, a response rate of 84 percent after the sampling weights were applied. As in the First Follow-Up Survey, most interviews were completed on the Web (63 percent), with the rest (37 percent) through CATI.

The Third Follow-Up was conducted among the 4,295 eligible businesses remaining in the KFS panel. Panel members were contacted initially by e-mail and invited to complete the KFS Web survey. Those businesses without e-mail addresses, or those not completing the Web survey after being contacted by e-mail, were contacted by U.S. mail. As in prior waves, respondents were paid $50 for completing the Third Follow-Up. With the remaining eligible businesses, 2,915 interviews were completed, which resulted in an unweighted response rate of 78 percent. Researchers verified 427 cases as going out of business. As in the first two Follow-Up Surveys, most interviews were completed on the Web (65 percent), with the rest (35 percent) through CATI.

More detailed information about the data collection efforts is available in Appendix B. A comparison of the KFS data and other business data is provided in Appendix C.

Chapter 2

Results of the Kauffman Firm Survey

The Kauffman Firm Survey has collected data about the selected businesses' operations during their first four years of existence (calendar years 2004, 2005, 2006, and 2007). The project is currently at the midpoint, as an additional four years are planned. At the end of the project, the KFS will contain data over the 2004–2011 period on a large cohort of firms that began operations in 2004. The Baseline sample consisted of 4,928 firms. Each year there is some loss in sample size due to sample attrition, refusals, "unlocatables," and firm closures. The tables in this chapter include only those businesses that have survived over the period or that have been verified as going out of business by 2007. The size of this sample is 3,974. The tables in this chapter can be broken out into the following groups:

A. FIRM AND OWNER CHARACTERISTICS (2004–2007)

The first set of tables shows the distribution of firms for the Baseline year (2004) and the Third Follow-Up survey (2007). Unless otherwise noted, the sample size for the 2004 distributions is 3,974, while the tables with 2007 data contain only the 2,913 businesses that survived through that year.

1. Characteristics of the Firm

Legal Form, Intellectual Property, Employment, and Location. As shown in Table One, the distribution of firms by legal form changed just slightly over the four-year period. By 2007, there was a smaller share of sole proprietorships, but a larger share of corporations. This is due to both changes in legal form of organization by businesses or a higher closure rate of sole proprietorships, compared with corporations. A slightly smaller share of businesses was home based in 2007, compared with 2004. Again, this is due to both businesses moving out of the home into leased or other owned spaces, and the higher closure rate of home-based businesses, compared with those that are not home-based. A higher share of businesses had employees in 2007, compared with 2004. The percentage of businesses having intellectual property (patents, copyrights, trademarks) increased slightly over the 2004–2007 period. In addition, the average number of these different types of intellectual property, except for the number of patents, increased slightly over the period as well. For employer firms, average employment increased from 4.6 employees to 6.7 employees.

Table One
Firm Characteristics

	All Firms 2004	Surviving Firms 2007
Legal Form		
Sole Proprietorship	36.0%	33.7%
Partnership	5.7%	4.9%
Corporation	27.7%	30.6%
Limited Liability Corporation	30.6%	30.6%
Firm Characteristics		
Home Based	50.0%	48.9%
Employer Firm	41.7%	55.9%
Percentage of Firms with Intellectual Property		
Patents	2.2%	2.8%
Copyrights	8.6%	9.5%
Trademarks	13.7%	13.9%
Average Number (for firms that had item>0)		
Patents (patents>0)	6.4	6.2
Copyrights (copyrights>0)	12.0	21.3
Trademarks (trademarks>0)	2.1	2.8
Employees (employment>0)	4.6	6.7
N	3,974	2,913

Source: Kauffman Firm Survey Microdata. Sample includes only surviving firms over the 2004–2007 period and firms that have been verified as going out of business over the same period. The original sample size in 2004 was 4,928.

Revenues and Expenses. The distributions of firms' revenues and expenses are shown in Table Two. As expected, the percentage of firms in the lower revenue and expense categories, those under $25,000, fell over the 2004–2007 period, while the percentages in the larger categories, $25,001–$100,000 and more than $100,000, increased. By 2007, about 40 percent of firms had revenues greater than $100,000, compared with just 17 percent in 2004. The changes in expenses were similar to those observed in revenues.

Profit and Loss. About 60 percent of the surviving firms posted profits in 2007, compared with about 40 percent posting losses. The distributions of firms' profits and losses are shown in Table Three. More than 15 percent of profitable firms had profits of more than $100,000, while another 30 percent posted profits between $25,000 and $100,000. For those businesses posting losses, more than 6 percent had losses of more than $100,000. In addition, more than one-third of firms had losses between $10,000 and $100,000. Comparing 2004 and 2007, a larger percentage of profitable firms were in the highest profit categories, while a smaller percentage of unprofitable firms were in the highest loss categories.

Table Two
Distribution of Revenues and Losses (2004 and 2007)

	All Firms 2004	Surviving Firms 2007
Revenues		
Zero	35.3%	23.2%
$500 or less	3.2%	1.0%
$501–$1,000	1.9%	1.0%
$1,001–$3,000	4.9%	2.9%
$3,001–$5,000	2.8%	2.1%
$5,001–$10,000	6.1%	4.5%
$10,001–$25,000	10.5%	7.0%
$25,001–$100,000	18.6%	18.7%
$100,001 or more	16.8%	39.7%
Expenses		
Zero	6.7%	4.2%
$500 or less	4.8%	3.5%
$501–$1,000	3.7%	2.1%
$1,001–$3,000	8.7%	4.7%
$3,001–$5,000	7.3%	4.9%
$5,001–$10,000	11.3%	7.4%
$10,001–$25,000	16.2%	12.3%
$25,001–$100,000	25.3%	24.4%
$100,001 or more	15.8%	36.5%
N	3,974	2,913

Source: Kauffman Firm Survey Microdata. Sample includes only surviving firms over the 2004–2007 period, and firms that have been verified as going out of business over the same period. The original sample size in 2004 was 4,928.

Table Three
Distribution of Profits and Losses (2004 and 2007)

	All Firms 2004	Surviving Firms 2007
Revenues		
$500 or less	24.9%	8.7%
$501–$1,000	4.7%	2.8%
$1,001–$3,000	9.2%	6.3%
$3,001–$5,000	7.2%	6.5%
$5,001–$10,000	12.5%	13.1%
$10,001–$25,000	17.4%	17.3%
$25,001–$100,000	20.0%	30.1%
$100,001 or more	4.1%	15.1%
Losses		
$500 or less	10.9%	11.1%
$501–$1,000	5.7%	6.4%
$1,001–$3,000	15.3%	13.6%
$3,001–$5,000	12.0%	11.0%
$5,001–$10,000	17.0%	15.0%
$10,001–$25,000	17.9%	19.0%
$25,001–$100,000	16.9%	17.3%
$100,001 or more	4.2%	6.6%
N	3,974	2,913

Source: Kauffman Firm Survey Microdata. Sample includes only surviving firms over the 2004–2007 period, and firms that have been verified as going out of business over the same period. The original sample size in 2004 was 4,928.

Asset Levels. Assets are what the business owns and may include cash, accounts receivable, equipment, machinery, product inventory, and vehicles. As expected, firms that survived through 2007 were much more likely to be in the higher asset categories, compared with 2004. As shown in Table Four, about one-third of businesses had assets of more than $100,000 by 2007, compared with just 20 percent in 2004. Almost 40 percent of businesses had assets between $10,000 and $100,000 in 2007, which is similar to firms in 2004. A smaller percentage of firms were in the lower asset categories in 2007, compared with 2004. With the longitudinal data, it is possible to track changes over each year in time and to track the magnitudes of those changes.

Table Four
Total Assets (2004 and 2007)

	Disribution of Assets For All Firms 2004	Disribution of Assets For All Surviving Firms 2007
Revenues		
Zero	11.4%	7.4%
$500 or less	3.1%	2.1%
$501–$1,000	3.1%	2.1%
$1,001–$3,000	8.5%	4.7%
$3,001–$5,000	5.8%	4.6%
$5,001–$10,000	9.4%	7.3%
$10,001–$25,000	15.1%	12.6%
$25,001–$100,000	23.2%	25.9%
$100,001 or more	20.4%	33.2%
N	3,974	2,913

Source: Kauffman Firm Survey Microdata. Sample includes only surviving firms over the 2004–2007 period, and firms that have been verified as going out of business over the same period. The original sample size in 2004 was 4,928.

Capital Injections. The next table in this section deals with new financial capital injections, shown for 2004 and 2007. The first column of each year is the average level of financial investment by that source, which includes all the firms without this type of financing source. The second column of each year is the average for just those firms that have that type of financing. The last column is the unweighted number of firms that had each source of financing. Some rows with less than ten observations have been suppressed.

As shown in Table Five, new firms injected about $80,000 on average into their new ventures during the first year of operation. Much of that, nearly $30,000, was owner equity. Outsider debt (bank loans, credit cards, credit lines, etc.) made up more than $32,000. Other debt from insiders (friends and family) and the owner(s) made up another $10,000. The remainder was external equity injected by insiders (spouse or parent) or outsiders (venture capitalists, government, etc.). While outside equity was rarely used, it was very important for the firms that did use it. The same is true about inside equity. An important observation is just how important external debt markets are for firms even in their first year of operations. It was the single largest source of financing for startups in 2004.

Three years later, surviving firms injected another $53,000 into their businesses. Although this amount is much lower than in 2004, the level of outside debt remains nearly constant. Thus, the importance of external debt markets continues to rise as firms survive and grow. The levels in most of the other categories fall quite dramatically. Again, very few firms receive equity from non-owner spouses, but these sources are very important for those firms that do access them.

Table Five
New Financial Capital Injections

	All Firms: 2004			Surviving Firms: 2007		
	All Firms	Just Firms with Source > 0	Number of Firms with Source > 0	All Firms	Just Firms with Source > 0	Number of Firms with Source > 0
Total New Financial Injections	$80,359	$89,255	3,564	$53,134	$79,878	1,905
Owner Equity	$28,541	$36,134	3,125	$8,434	$24,968	946
Insider Equity	$1,700	$36,367	177	$893	$44,157	60
Spouse Equity	$491	$30,732	62	$479	$56,649	15
Parent Equity	$1,209	$35,310	126	$870	$37,745	48
Outsider Equity	$6,901	$153,608	205	$3,238	$157,608	76
Other Informal Investors	$2,793	$107,685	110	$2,092	$120,814	45
Other Business Equity	$1,841	$162,369	56	$1,397	$131,219	23
Government Equity	$466	$85,664	27	$210	$88,994	10
Venture Capital Equity	$1,454	$352,111	26	$1,227	$360,419	10
Owner Debt	$3,487	$11,322	1,194	$3,747	$16,139	644
Personal Credit Card—Owner	$3,175	$10,587	1,159	$3,489	$15,322	629
Personal Credit Card—Other Owners	$288	$8,995	132	$255	$15,405	55
Insider Debt	$7,633	$52,048	542	$3,874	$47,544	215
Personal Family Loan	$2,670	$28,398	327	$1,544	$26,473	156
Personal Family Loan—Other Owners	$286	$34,681	29	$23	$6,379	12
Business Loan from Family	$1,350	$43,909	115	$602	$37,610	42
Business Loan from Owner	$1,887	$117,804	67	$331	$93,092	13
Other Personal Loan	$559	$29,457	73	$458	$44,297	23
Other Personal Funding	$812	$64,514	50	$873	$141,540	15
Outsider Debt	$32,097	$86,374	1,439	$32,947	$77,174	1,220
Personal Bank Loan	$10,476	$61,086	641	$7,005	$68,150	280
Business Credit Card	$1,394	$9,828	543	$4,107	$16,687	709
Other Bank Loan	$1,498	$65,010	92	$490	$43,624	32
Business Credit Card—Other Owners	$167	$9,694	62	$175	$8,711	67
Business Credit Cards	$859	$7,383	452	$2,687	$16,884	463
Bank Business Loan	$10,060	$154,043	243	$7,937	$112,723	193
Credit Line	$3,798	$71,429	210	$6,740	$62,853	315
Other Non-Bank Loan	$2,040	$120,950	72	$1,763	$100,165	51
Government Business Loan	$725	$84,303	34	$1,443	$415,773	10
N	3,974			2,913		

Source: Kauffman Firm Survey Microdata. Sample includes only surviving firms over the 2004–2007 period, and firms that have been verified as going out of business over the same period. The original sample size in 2004 was 4,928.

Survival and Outcomes. As shown in Table Six, around 90 percent of firms in the KFS survived through 2005.[4] By 2006, about 80 percent had survived and, by 2007, the survival rate had dropped to 73.4 percent. While most of the remaining firms closed down either temporarily or permanently over the period, a small fraction, 3.5 percent, by 2007 either were sold to or merged with another business. If they were sold or merged, they drop out of our frame.

Comparing outcomes in 2004 for those that closed down over the period and those that survived through 2007, those that closed had lower profits, revenues, and assets in 2004 than those that survived. Comparing the average outcomes of surviving firms from the Baseline of 2004 and the third follow-up year in 2007, one can see that the surviving firms grew substantially over the 2004–2007 period. The average profits for the surviving businesses were about $25,000 in 2004 and more than double that by 2007. Revenues nearly tripled, while average assets increased for surviving firms from about $74,000 in 2004 to more than $118,000 by 2007.

2. Characteristics of the Owners

Gender, Race, Immigrant Status, Age, and Hours Worked. The next set of tables shows the distributions of firms in 2004 and 2007 by Baseline primary owner characteristics.[5] About 65 percent of the KFS firms have just one owner, while 26 percent have two owners, and 9 percent have three or more owners. For this set of tables, a primary owner was designated for multi-owner firms by the largest equity share in 2004. In cases where two or more owners owned equal shares, hours worked and a series of other variables were used to create a rank ordering of owners in order to define a primary owner.

As shown in Table Seven, the surviving firms were slightly more likely to be male and white, although the distributions were pretty similar in 2004 and 2007. Similarly, the distributions by immigrant status and age also were quite similar. Owners of the surviving firms were slightly more likely to work

Table Six
Firm Outcomes (2004–2007)

	2005	2006	2007
Survived	89.6%	80.2%	73.4%
Closed Operations	6.0%	14.0%	21.4%
Sold to or Merged with Another Business	1.4%	2.7%	3.5%
Temporarily Ceased Operations or Other	3.0%	3.1%	1.7%
	100.0%	100.0%	100.0%

	2004 (Closed by 2007)	2004 (Survived by 2007)	2007 (Survived to 2007)
Profits	$21,197	$24,982	$54,652
Revenue	$ 51,588	$54,994	$152,063
Assets	$63,904	$73,668	$118,791

Source: Kauffman Firm Survey Microdata. Sample includes only surviving firms over the 2004–2007 period, and firms that have been verified as going out of business over the same period. The original sample size in 2004 was 4,928.

4. This is a higher survival rate than what is found in other data sources. The higher-than-average survival rates in the KFS are in part the result of the timing of the survey. Data released for the KFS survey were from June and November 2005 Dun & Bradstreet files and interviewing for the Baseline continued until July 2006. Thus, if a firm started in 2004 and closed shortly afterwards, it could be missing from the files that became the frame from which the KFS sample was drawn. See Appendix C for more details.

5. Demographic information is collected on up to ten owners for each business in the KFS.

Table Seven
Distribution of Firms by 2004 Primary Owner Demographics

	All Firms 2004	Surviving Firms 2007
Gender		
Male	69.2%	70.4%
Female	30.8%	29.6%
Race and Ethnicity		
White	79.3%	79.6%
Black	8.6%	8.3%
Asian	4.2%	4.8%
Other	2.3%	2.1%
Hispanic	5.5%	5.3%
Immigrant Status		
Native Born	89.9%	89.7%
Immigrant	10.1%	10.3%
Not a U.S. Citizen	33.9%	35.9%
U.S. Citizen	66.1%	64.1%
Age		
24 or younger	1.3%	1.2%
25–34	16.5%	16.0%
35–44	33.6%	34.0%
45–54	29.0%	28.9%
55 or older	19.6%	19.9%
Average Hours Worked (week)		
Less than 20	18.5%	17.5%
20–35	19.5%	18.4%
36–45	14.3%	14.3%
46–55	15.2%	15.8%
56 or more	32.5%	34.1%
N	3,974	2,913

Source: Kauffman Firm Survey Microdata. Sample includes only surviving firms over the 2004–2007 period, and firms that have been verified as going out of business over the same period. The original sample size in 2004 was 4,928.

more hours in an average week than those in the 2004 Baseline. There are no large shifts in distribution over the 2004–2007 period.

Human Capital. There were larger shifts in the distributions of firms by the Baseline human capital measures in the Kauffman Firm Survey. As shown in Table Eight, the surviving firms in 2007 were more likely to be in the higher human capital categories of education, previous industry work experience, and previous business starts.

Table Eight
Distribution of Firms by 2004 Primary Owner Demographics

	All Firms 2004	Surviving Firms 2007
Education		
High School Graduate and Less	13.9%	13.0%
Technical/Trade/ Vocational Degree	6.4%	6.5%
Some College, No Degree	21.8%	21.7%
Associate's Degree	8.6%	8.7%
Bachelor's Degree	25.3%	24.6%
Some Graduate School, No Degree	5.9%	6.5%
Master's Degree	13.4%	13.7%
Professional School/ Doctorate	4.7% 5	.2%
Previous Industry Experience (years)		
Zero	9.8%	8.7%
1–2	13.9%	13.2%
3–5	15.6%	13.9%
6–9	9.9%	10.4%
10–19	24.9%	25.7%
20–29	16.8%	18.6%
30+	9.3%	9.5%
Previous Start Up Experience (number)		
0	57.5%	56.8%
1	21.5%	21.8%
2	10.2%	10.5%
3	5.0%	5.4%
>=4 5.8% 5.5%		
N	3,974	2,913

Source: Kauffman Firm Survey Microdata. Sample includes only surviving firms over the 2004–2007 period, and firms that have been verified as going out of business over the same period. The original sample size in 2004 was 4,928.

B. NEW TOPICS COVERED ON ACTIVITIES IN 2007

Several additional questions were added to the Third Follow-Up questionnaire, so new information is available from this fourth year of data on activities taking place in 2007. Unfortunately, it is not possible to go back in time and collect this same information in

previous years, so this information is not available for businesses that did not survive until 2007.

Predominant Market. The top part of Table Nine gives the distribution of firms by the predominant market for the firm's products and/or services. The first column shows the distribution for all surviving firms, while the second column shows the subset of high-tech firms (see Appendix A for details on how high tech was defined). About 12 percent of firms serve a predominantly local market, in or around the neighborhood where the business operates. Nearly a third of firms have a market area that comprises the city or county where the business operates. More than a third had a statewide or regional market, while about 17 percent of firms have a national

Table Nine
New Information from Third Follow-up Data (2007) (Percentage of Surviving Firms)

	All	High Tech
Predominant Market All High Tech		
Neighborhoods Local to the Business	11.8%	2.8%
Same City or County	32.4%	17.1%
Same Region	35.6%	32.4%
Nationwide	17.1%	36.2%
International	3.2%	11.5%
Any International Sales?		
No	86.9%	65.8%
Yes	13.1%	34.2%
International Sales as a Percentage of Total Sales (for those >0)		
Less than 5%	55.2%	43.6%
5%–25%	29.5%	30.4%
26%–50%	5.4%	11.1%
51%–75%	4.1%	5.2%
76%–100%	5.9%	9.7%
Any Internet Sales?		
No	73.9%	71.0%
Yes	26.1%	29.0%
Internet Sales as a Percentage of Total Sales (for those >0)		
Less than 5%	33.1%	35.1%
5%–25%	30.9%	30.7%
26%–50%	11.3%	6.7%
51%–75%	7.8%	6.4%
76%–100%	16.9%	21.2%
Do you Feel your Firm has a Comparative Advantage?		
No	40.8%	31.0%
Yes	59.2%	69.0%
Sources of Comparative Advantage (for those >0)		
Teaming Up with a College or University	6.5%	8.1%
Teaming Up with another Company	23.9%	24.2%
Teaming Up with a Government Lab or Research Center	1.9%	5.7%
Patents that the Firm Owns, has Applied for, or Licensed	6.8%	21.7%
N	2,913	434

Source: Kauffman Firm Survey Microdata. Sample includes only surviving firms over the 2004–2007 period.

market. Only about 3 percent of firms stated that their main market was international. However, more than 13 percent of firms had some international sales. For those with international sales, more than half of firms said that the international sales made up less than 5 percent of their total sales. About 10 percent of firms with international sales generated more than half of their sales internationally. About a third of firms with international sales said that the international sales made up between 5 percent and 50 percent of their total sales.

High-tech firms were much more likely to have broader markets for their products and services. Nearly half of high-tech firms had nationwide or international markets as their main market, and more than one-third of high-tech firms had some international sales. For those high-tech firms that had international sales, about 15 percent had international sales that made up at least half of their total sales.

Internet Sales. More than a quarter of firms sold at least some of their goods or services on the Internet. For nearly a quarter of those firms, Internet sales represented more than half of their total sales; about a third of those firms said Internet sales were less than 5 percent of their total sales. Nearly another third said that Internet sales made up between 5 percent and 25 percent of their sales, while just over 11 percent of those firms said their Internet sales made up between 26 percent and 50 percent of their total sales. Compared to all firms, on average, a slightly higher percentage of high-tech firms had some Internet sales; a higher percentage of high-tech firms also had Internet sales that made up more than 75 percent of total sales.

Sources of Comparative Advantage. The survey has asked about comparative advantage over the entire survey period, but additional questions about the sources of that comparative advantage were added in 2007. About 60 percent of firms felt they had a comparative advantage in the products or services they offered. About a quarter of those firms felt that that comparative advantage stemmed from teaming up with another company, while only 2 percent felt it was due to teaming up with a government lab or research center. About 6 percent felt their comparative advantage was due in part to teaming up with a college or university, while 7 percent felt it was due in part to patents that the firm owned, had applied for, or had licensed. Nearly 70 percent of high-tech firms thought they had a comparative advantage and were more likely to cite the listed sources—especially patents—as reasons for their comparative advantage.

Borrowing Experience. While Table Seven showed the levels of financial injection by detailed source, Table Ten shows the borrowing experiences of firms in 2007. Only about 12 percent of firms submitted new credit applications in 2007. For the vast majority of firms, the application(s) always were approved. Nearly 18 percent of firms had mixed results, sometimes approved and sometimes denied. Just over 12 percent of firms said that their loan applications always were denied. For those that had some or all of their loan applications denied, almost half said that one of the main reasons given was insufficient collateral to guarantee the loan. The second most-common reason was flaws in the owner's personal credit history. Nearly a third said they were told the firm hadn't been in business long enough (recall that these firms all began operations in 2004 so they are now about three years old). About 30 percent of firms said they were told the loan size requested was too large. Nearly 29 percent said business credit history played a role in denial. Only about 7 percent said the firm had inadequate documentation. Interestingly, 17 percent of firms said that they didn't apply for credit at some point when they needed it because they feared their loan application would be denied. However, this group that feared denial had a higher proportion of firms that applied (20.7 percent) than the sample overall (12 percent).

Table Ten
2007 Borrowing Experiences

Applied for New Credit in 2007	
No	87.9%
Yes	12.1%
Credit Decision for Those that Applied	
Always Approved	70.1%
Sometimes Approved/Sometimes Denied	17.6%
Always Denied	12.2%
Reasons Given for Denial (for those sometimes or always denied)	
Insufficient Collateral	47.9%
Loan Size Requested Too Large	29.8%
Inadequate Documentation	7.1%
Business Credit History	28.6%
Personal Credit History	46.3%
Not in Business Long Enough	32.0%
Other	7.1%
Didn't Apply when Firm Needed Credit for Fear of Being Denied	
No	83.0%
Yes	17.0%

Source: Kauffman Firm Survey Microdata. Sample includes only surviving firms over the 2004–2007 period.

Borrowing Experience by Primary Owner Demographics. The last table breaks out the 2007 borrowing experiences by the demographics of the primary owner. Previous research has shown large differences in borrowing patterns, loan turndowns, and fear of borrowing by gender, race, and ethnicity. A review of that research can be found in Fairlie and Robb (2008). The KFS data confirm many of these previous findings. Firms owned by men are more likely to apply for new funding than firms owned by women (13 percent vs. 9 percent respectively). Less than 10 percent of firms owned by African Americans and less than 8 percent of firms owned by Asian Americans applied for new credit in 2007, compared with nearly 13 percent of firms owned by non-Hispanic whites. Just over 10 percent of firms owned by Hispanics applied for new credit in 2007. Women and African Americans were more likely than men and whites to state that they didn't apply for credit at some point when they needed it in 2007 for fear of being denied.

Nearly 40 percent of African Americans said they feared being denied, compared with 14.5 percent of whites. About 20 percent of Hispanics said they didn't apply when they needed credit for fear of denial. Only about 12 percent of Asians said they didn't apply for fear of denial. In looking at applications, more than 36 percent of women

applicants had their loan applications sometimes or always denied, compared with less than 30 percent of male applicants. More than three-quarters of applications by African Americans were denied, and more than 36 percent of Hispanic applications were denied. This compared with 32 percent of applications by Asians and 26 percent of applications by white business owners. Clearly, there are large racial and gender differences in the borrowing experiences by business owners. Further examination of this issue, including differences by credit history, industry, and previous financing is planned.

Table Eleven
2007 Borrowing Experiences by Primary Owner Demographics

Applied for New Credit in 2007	
Male	13.2%
Female	9.4%
White	12.9%
Black	9.7%
Asian	7.8%
Hispanic	10.2%
Credit Applications were Sometimes or Always Denied	
Male	29.6%
Female	36.2%
Black	76.7%
White	25.6%
Asian	32.3%
Hispanic	36.2%
Didn't Apply when Firm Needed Credit for Fear of Being Denied	
Male	16.6%
Female	18.1%
White	14.5%
Black	38.7%
Asian	12.2%
Hispanic	19.9%

Source: Kauffman Firm Survey Microdata. Sample includes only surviving firms over the 2004–2007 period.

Appendix A

Sampling Methods

The target population is the population on which conclusions are drawn. For the Kauffman Firm Survey, the target population was all new businesses that were started in the 2004 calendar year in the United States (the fifty states plus the District of Columbia). This population excludes any branch or subsidiary owned by an existing business or a business inherited from someone else. The issue that arose immediately with this target definition is the meaning of "started." Working with the Kauffman Foundation and the technical advisory group, Mathematica Policy Research, Inc., (MPR) used pilot studies to evaluate alternative definitions of "started" based on indicators of business operations, such as having an Employer Identification Number (EIN), Schedule C income, a legal form, or payment of state unemployment insurance or federal Social Security taxes. For the study population, a business started in 2004 was defined as a new, independent business that was created by a single person or a team of people, the purchase of an existing business, or the purchase of a franchise. Businesses were excluded if they had an EIN, Schedule C income, or a legal form or had paid state unemployment insurance or federal Social Security taxes prior to or after 2004.

The sampling frame for the KFS is based on the Dun & Bradstreet (D&B) database and restricted to businesses (or enterprises) that D&B reported started in 2004. The D&B database was partitioned into six sampling strata defined by industrial technology categories (based on industry designation) and gender of the business owner or CEO (based on the D&B data element and supplemented by including businesses whose owners had a feminine first name). The high-technology strata were defined based on categorization developed by Hadlock et al. (1991). The definition took into account the industry's percentage of R&D employment and classified the businesses into technology groups based on their Standard Industrialization Classification (SIC) codes. The industries in the technology strata are shown in Table A.1.

Because of the analytic interest of the high-technology businesses, we oversampled these businesses. Specifically, the original sampling design called for 2,000 interviews to be completed among businesses in two categories of high-technology businesses and 3,000 interviews to be completed among businesses in all other industrial classifications. Subsequently, we took all high-tech businesses into the sample. The women-owned indicator served as an explicit stratum, so that the proportion of women-owned businesses in the sample was the same as the proportion of the women-owned businesses in the frame.

A. SAMPLING FRAME

The D&B database is a compilation of data from various sources, including credit bureaus, state offices that register some new businesses, and companies (e.g., credit card and shipping companies) that are likely to be used by all businesses. However, compiling information on newly formed businesses is particularly difficult because there is no single registry of new businesses and the time between establishing the business and the business showing up in one of D&B's sources may be six or more months. To capture as complete a picture as possible of businesses starting in 2004, we arranged with D&B to provide multiple files at different time points during 2005. We obtained a file in June 2005 and then a new file in November 2005. As shown in Table A.2, in June 2005, D&B provided MPR with a file of 188,000 businesses with a reported starting year of 2004. This number was approximately 30 percent lower than a similar file received in June 2004 of businesses starting in 2003. We investigated the lower number and found no clear changes in operations by D&B and no evidence available from federal sources to verify or disprove this count.

The November D&B file included roughly 63,000 businesses with reported starting dates in 2004, resulting in a total pool of roughly 251,000

Table A.1
Technology Strata Definitions

Technology Stratum	High Tech	Industry
High Tech	28	Chemicals and allied products
	35	Industrial machinery and equipment
	36	Electrical and electronic equipment
	38	Instruments and related products
Medium Tech	131	Crude petroleum and natural gas operations
	211	Cigarettes
	291	Petroleum refining
	299	Miscellaneous petroleum and coal products
	335	Nonferrous rolling and drawing
	371	Motor vehicles and equipment
	372	Aircraft and parts
	376	Guided missiles, space vehicles, parts
	737	Computer and data processing services
	871	Engineering and architectural services
	873	Research and testing services
	874	Management and public relations
	899	Services, not elsewhere classified
	229	Miscellaneous textile goods
	261	Pulp mills
	267	Miscellaneous converted paper products
	348	Ordnance and accessories, not elsewhere classified
	379	Miscellaneous transportation equipment
Non-tech		**All other industries**

Table A.2
Sampling Frame of Businesses in D&B Database: Businesses with Start Date of 2004

Technology Stratum	Women-Owned	June File	Deceased in November[1]	New in November	Total All (June and November)	Operating Total
Total		**188,292**	**13,439**	**62,990**	**251,282**	**237,843**
High Tech		2,593	144	1,276	3,869	3,725
	Yes	361	21	166	527	506
	No	2,232	123	1,110	3,342	3,219
		22,544	926	7,117	29,661	28,735
Medium Tech	Yes	4,332	153	1,215	5,547	5,394
	No	18,212	773	5,902	24,114	23,341
		163,155	12,369	54,597	217,752	205,383
Non-Tech	Yes	32,016	2,177	9,951	41,967	39,790
	No	131,139	10,192	44,646	175,785	165,593

1. "Deceased in November" is the count of businesses in the D&B database in June 2005 that were not in the database in November 2005.

businesses from the combined June and November files (Table A.2). However, 13,000 businesses from the June file (7 percent) were not in the November file (Table A.3); the new total pool was 238,000 businesses. We presumed the 13,000 businesses were no longer in operation. Such businesses were referred to as "deceased."

B. SAMPLE ALLOCATION

Because we planned to obtain a second D&B file in November 2005, we needed to release a sufficiently large sample in June 2005 to accommodate the expected response and eligibility rates, but we also wanted to balance the sample size between the two files to reduce unequal

Table A.3
Losses of Businesses in D&B Database: June 2005 to November 2005, with 2004 Start Date

Technology Stratum	Women-Owned	June File	Deceased	Percentage Deceased
Total		**188,292**	**13,439**	**7.1**
High Tech	Total	2,593	144	5.6
	Yes	361	21	5.8
	No	2,232	123	5.5
Medium Tech	Total	22,544	926	4.1
	Yes	4,332	153	3.5
	No	18,212	773	4.2
Non-Tech	Total	163,155	12,369	7.6
	Yes	32,016	2,177	6.8
	No	131,139	10,192	7.8

sampling weights. As mentioned earlier, because the high-technology businesses numbered only 2,500 (again, fewer than expected) and because we wanted a large pool of these businesses for the longitudinal panel, we decided to include all of these businesses in the sample to obtain an adequate count of these businesses. For the other strata, we were somewhat conservative but still released relatively large samples.

When the November sample was released, we again decided to include all of the high-technology businesses in the sample to obtain an adequate count of these businesses for the longitudinal panel. For the other two strata, we attempted to balance the final sample across the two files and the sampling strata. The final sample size and sampling rates are shown in Table A.4.

To select each sample, a sequential random sample selection procedure was used, which sorts the observations in each of the sampling strata in a serpentine fashion based on a set of specified characteristics. This process, outlined by Chromy (1979), imposes implicit stratification beyond the primary strata to ensure the sample is balanced on the implicit stratification variables. For the KFS, within each sampling stratum, the records were sorted using a serpentine methodology based on the employee size category and three-digit zip code to ensure approximate proportional representation by these dimensions within each stratum.

Table A.4
Samples from D&B Database: Businesses with 2004 Start Date

Technology Stratum	Women-Owned	June Sample	June Frame Percentage	November Sample	November Frame Percentage	Total Sample
Total Sample		**23,942**	**12.7**	**8,527**	**13.5**	**32,469**
High Tech	Total	2,593		1,276		3,869
	Yes	361	100.0	166	100.0	527
	No	2,232	100.0	1,110	100.0	3,342
Medium Tech	Total	5,769		1,805		7,574
	Yes	1,029	23.8	237	19.5	1,266
	No	4,740	26.0	1,568	26.6	6,308
Non-Tech	Total	15,580		5,446		21,026
	Yes	2,090	6.5	670	6.7	2,760
	No	13,490	10.3	4,776	10.7	18,266

Appendix B

Data Collection Methods

A. BASELINE SURVEY DATA COLLECTION

The goal of the Baseline Survey was to establish the Kauffman Firm Survey panel by completing surveys with the principals of businesses that met the screening criteria for eligibility as outlined in Chapter I. MPR conducted two pilot tests to examine the eligibility criteria, the questionnaire length and structure, the use of incentives, and the collection of data through a Web survey option with Computer-Assisted Telephone Interviewing (CATI) follow-up. To minimize mode effects, MPR made significant efforts to create Web and CATI versions of the survey that were as uniform as possible. Based on the results of these pilot tests, the Baseline Survey began in July 2005 with a comprehensive screening approach to ensure a "pure" cohort of businesses that began operations in 2004. The findings from the pilot tests also led to streamlining of the questionnaire and the decision to offer eligible Baseline KFS respondents a $50 postpaid incentive.

The Baseline Survey's first contact with businesses was a letter to the business owner, which introduced the study, asked for cooperation, and provided Web login information. Accompanying the letter were instructions on how to access the KFS Web survey and a one-page Frequently Asked Questions (FAQs) document that provided answers to some common questions sample members were likely to have about the survey, MPR, and the Foundation. Included in this information was a toll-free number the business owner could call for additional information.

One week after the letter and accompanying materials were sent, we followed up with a postcard reminder to the businesses. The postcard provided the survey Web address and encouraged respondents to log on to the Web site and complete the survey. We did not include the login and password information on the postcard, as this would give potential unqualified respondents access to the Web survey. No mention was made of the telephone follow-up in either the introductory letter or the postcard. This "forced Web" approach was designed to maximize the response on the Web.

During the first two weeks between the advance mailing and the start of CATI operations, only 2 percent to 3 percent of the businesses accessed the Web survey. Most of the business that accessed the Web either completed the survey or were screened out as ineligible. This low level of response necessitated a significant effort to complete the remaining interviews by CATI.

In preparation for CATI operations, MPR project staff held comprehensive interviewer training sessions. These training sessions emphasized thorough knowledge of the study and its importance to new business owners, criteria for screening out ineligible businesses, effective ways of introducing the study, and refusal avoidance techniques. Based on the results of the pilot testing, particular emphasis was placed on refusal avoidance during the training. Interviewers practiced responding to objections, particularly when sample members cited a "lack of time" or indicated they were "not interested." Interviewers also practiced addressing issues of confidentiality and assuring business owners that information they provided would never be identified with their businesses.

One segment of the training was dedicated to dealing effectively with people who answered the phone but were not the business owner. These people, known as gatekeepers, can constitute a significant barrier to speaking with the business owner. Nonetheless, the Baseline Survey produced a high refusal rate, with 2.5 refusals on average for each completed CATI interview.

We attempted one refusal conversion effort for each business. First, all refusals were put on hold and not contacted for fourteen days, so that a refusal letter could be sent to the business. The letter acknowledged the refusal but emphasized the unique nature of the study and the importance of participation. The survey Web site was provided in

the refusal letter, along with the sample member's password and login ID. After the fourteen-day waiting period, if the sample member had not completed the survey on the Web, interviewers trained in refusal conversion techniques called the business owner. Interviewers converted 538 refusals, representing 8 percent of all completes. In addition, another 1,062 businesses that initially refused were screened out as ineligible. All businesses refusing a second time were finalized.

Efforts to locate businesses that could not be contacted using the information provided by D&B were extensive. Although the D&B database provides names, addresses, and phone numbers of the businesses, the fact that these are new businesses means that some of them will never become established. Others move or change phone numbers, especially those that are home-based businesses.

The KFS locating process used several resources to locate sampled businesses or principals, all of which provided names, addresses, and/or phone numbers of individuals and businesses or helped verify existing contact information. Through systematic use of these resources, locators sometimes could determine that a business was still operating and find updated contact information for interviewers or for mail contacts. Businesses that were confirmed as out of business were coded as ineligible.

Additional methods of interacting with Baseline respondents helped to complete surveys and identify additional businesses as ineligible, including a special e-mail sent to businesses upon request. Sample members made these requests either when contacted by telephone or by contacting MPR independently through e-mail. The e-mail included the Web login information and also a concise version of the FAQs. We also faxed advance materials upon request. The project used specially trained staff to answer questions, or provide login and password information when business owners or gatekeepers called the toll-free number. We also left answering machine messages with information about the study, the incentive, and the toll-free number.

As the Baseline Survey effort drew to a close, additional techniques were used to contact businesses and maximize the number of completes. These included sending an additional letter to all businesses that had not yet completed the survey. This letter indicated that the Baseline Survey was drawing to a close, emphasized that the project needed their participation, reminded them of the incentive, and asked them to complete the study. We also focused locating efforts on businesses that had not been worked completely, while finalizing those that had been worked thoroughly as "unlocatable." Finally, we used a special answering machine message for interviewers to use that emphasized that the study was ending and this was the last opportunity to participate.

A total of 32,469 selected businesses were released for data collection between July 2005 and July 2006. The selected businesses were released in six waves, with each wave worked with similar levels of effort. Data collection ended with 4,928 completed surveys, which translates to a 43 percent weighted response rate. Project staff, the Foundation, and the principal investigator discussed the trade-offs between reaching the original goal of 5,000 completes versus the project's budget constraints. Out of that discussion came the decision to complete at least 4,900 interviews and end data collection on July 29, 2006, making the field period exactly one year. Of the completed surveys, CATI completes accounted for 3,781 (77 percent) and Web completes accounted for 1,147 (23 percent) of the total interviews. More than 375,000 calls were required to complete the Baseline Survey.

Because these 4,928 businesses constituted our panel for future rounds of the KFS, additional efforts were made to maintain contact with panel members. We mailed a "welcome packet" about three months after their completion of the Baseline Survey. The KFS welcome packet consisted of a welcome letter, a brochure on the Kauffman Foundation, and a pen with the inscription "Kauffman Firm Survey." The welcome letter thanked respondents for completing the survey and reminded them that this is a multiyear study and that we would be contacting them again. The letter also contained contact information for MPR's survey director as an additional means to contact the researchers. The welcome packet also proved to be

an effective tool in getting updated contact information.

B. FIRST FOLLOW-UP DATA COLLECTION

While the Baseline Survey was characterized by a high refusal rate, a high rate of phone completes compared to Web completes, and a high number of phone calls per complete, the KFS First Follow-Up Survey results were significantly different. Businesses that were recruited in the Baseline Survey proved to be very cooperative in the First Follow-Up, and much more likely to complete the study via the Web.

The First Follow-Up instrument was modified to take into account Baseline Survey responses. In addition, the complex business eligibility module was eliminated in the First Follow-Up, as businesses that completed in Baseline were by definition eligible if they were still in operation. Preloaded information was added to the First Follow-Up instrument, such as the description of the business, owner names from Baseline, and contact information. Some questions asked about increases and decreases in employees, revenues, and expenses, without mentioning the actual Baseline responses.

The First Follow-Up instrument was designed to encourage the same respondent from Baseline to answer in First Follow-Up, assuming the Baseline respondent was still an owner and operator of the business. Other owner-operators could answer for the business if the Baseline respondent had left the firm, was no longer an owner-operator, or was unavailable during the field period.

In the Baseline Survey, business owners were asked for updated contact information, including e-mail addresses. Approximately 85 percent of business owners provided an e-mail address. To take advantage of that, and to continue to encourage the Web component of the study, the contact procedure was modified in the First Follow-Up. The first contact was an e-mail message that provided information similar to that contained in the Baseline Survey advance letter. It included a link to the KFS Web survey address, which was complete with the unique login and password for the business. A brief set of FAQs also was provided. One week later, an e-mail reminder was sent to all businesses that had not completed the survey. These initial e-mails were effective in getting almost 25 percent of respondents to complete by Web.

One week after that, an advance letter similar to that used with the Baseline Survey was sent to all businesses that had not completed the study. For the 15 percent of the businesses that had not given us e-mail addresses in the Baseline or whose e-mail addresses had changed or expired, this was the first contact about the First Follow-Up Survey. One week after the advance letter was mailed, a reminder postcard was sent. A week after the reminder postcard, CATI operations began. Close to 35 percent of sample members had completed by Web prior to the beginning of CATI operation.

The First Follow-Up also benefited from using experienced KFS Baseline Survey telephone interviewers. These interviewers were well versed in the study and adapted readily to the minor changes in question wording. Baseline Survey respondents generally remembered participating in the study and required little persuasion to do the second round.

Additional contact procedures and procedures used toward the end of the data collection period in Baseline also were used in the First Follow-Up. Locating procedures were also the same, although fewer businesses required locating than during the Baseline. Refusal conversion procedures also were used, although the total number of refusals was small. In contrast to the Baseline, during which all second refusals were finalized, project staff examined all second refusals during the First Follow-Up and put them into three categories: (1) refusals that might be converted on a third try, (2) businesses that should be finalized for the First Follow-Up but could be tried for the Second Follow-Up, and (3) businesses that should be finalized and not contacted again.

Data collection on the First Follow-Up Survey began in June 2006 and ended in January 2007. A total of 3,998 businesses completed the First Follow-Up, with 59 percent completing by Web, compared to 23 percent in the Baseline Survey. This transitioning of the majority of respondents from phone to Web greatly reduced the level of effort of the data collection. In contrast to the Baseline

Survey, for which more than 100 calls were made per phone complete, the First Follow-Up required only twenty-five calls per completed phone interview. The percentage of businesses verified as out of business at the time of First Follow-Up was 7.5, and the final refusal rate was slightly less than 3 percent. The response rate was 89 percent after sampling weights were applied.

To maintain the panel for the Second Follow-Up, a "cohort maintenance" packet was mailed to all First Follow-Up respondents. As in the Baseline welcome packet, this packet contained a letter thanking respondents for their participation, indicated that MPR would be contacting them for an additional survey, and included a gift of Post-it notes with the Kauffman Firm Survey name printed on them.

C. SECOND FOLLOW-UP DATA COLLECTION

Data collection on the Second Follow-Up Survey closely mirrored that of the First Follow-Up. Data collection began on May 31, 2007, and concluded on December 1, 2007. Overall, the study continued to be successful in retaining panel businesses, achieving a response rate of 84 percent (weighted). There was a slight increase in the percentage of respondents who completed the Web survey, (63 percent in the Second Follow-Up compared to 59 percent in the First Follow-Up). Because the Second Follow-Up Survey was the third annual survey in which KFS panel members were asked to participate, KFS respondents usually remembered the previous surveys and required little persuasion. Nonetheless, there were some refusals, which necessitated a refusal conversion effort. Of the 4,523 cases in the Second Follow-Up, 404 initially refused, of which 66, or 16 percent, were converted and completed the questionnaire. For the Second Follow-Up, the "panel maintenance packet" consisted of a letter of appreciation and a solar calculator with "Kauffman Firm Survey" printed on it. The letter thanked respondents for completing the survey and reminded them of the interview in the upcoming year.

D. THIRD FOLLOW-UP DATA COLLECTION

The Third Follow-Up data collection closely mirrored the collections of the first two follow-ups. Data collection began on June 24, 2008, and concluded on December 23, 2008. Some respondent fatigue was observed and a response rate dropped slightly to 78 percent (unweighted). There was a slight increase in the percentage of respondents who completed the Web survey (65 percent in the Third Follow-Up compared with 63 percent in the Second Follow-Up). For the Third Follow-Up, the "panel maintenance packet" consisted of a colorful 2009 calendar with the study name and contact information printed on it. As with previous rounds, respondents also received a $50 incentive payment for completing the survey.

Appendix C

Comparisons Between the Kauffman Firm Survey and Other Business Data Sources

A. INTRODUCTION

This appendix provides an overview of the KFS data as well as other available data sources on businesses in the United States. The KFS then is compared with these other sources along a number of dimensions.

B. THE KAUFFMAN FIRM SURVEY

The KFS data consist of a baseline sample of nearly 5,000 firms that began operations in 2004. Data are being collected about the nature of new business formation activity; characteristics of the strategy, offerings, and employment patterns of new businesses; the nature of the financial and organizational arrangements of these businesses; and the characteristics of their founders.

The KFS will undercount the number of new firms because the D&B frame is not the universe of all U.S. firms, nor does it capture all firm births. The D&B database is a compilation of data from various sources, including credit bureaus, state offices that register some new businesses, and companies (e.g., credit card and shipping companies) that are likely to be used by all businesses. However, compiling information on newly formed businesses is particularly difficult because there is no single registry of new businesses and the time between establishing the business and the business showing up in one of D&B's sources may be six or more months. Even the Census Bureau has trouble building and maintaining a frame that contains the universe of new businesses, especially if there are no employees, because of the difficulty of defining what constitutes a business start and the high churn rate of firms in the early years of operation.

The first challenge in conducting this survey was creating a consistent definition of what constituted a new business and the start of business operations. The KFS sought to create a panel that included new businesses created by a person or team of people, existing business purchases by new ownership teams, and franchise purchases. To this end, the KFS excluded D&B establishment records for businesses that were wholly owned subsidiaries of existing businesses, businesses inherited from someone else, and not-for-profit organizations. Also, previous research on new businesses has reported variability in how business founders perceive when their businesses started operations. Therefore, respondents were asked a series of questions about indicators of business activity and whether these were conducted for the first time in the reference year (2004). These indicators included:

- Payment of state unemployment (UI) taxes
- Payment of Federal Insurance Contributions Act (FICA) taxes
- Presence of a legal status for the business
- Use of an Employer Identification Number (EIN)
- Use of Schedule C to report business income on a personal tax return

To be "eligible" for the KFS, at least one of these activities had to have been performed in 2004 and none performed in a prior year. Therefore, by our definition of business start, we have created a unique population that is not exactly comparable to any other existing data sources. However, several available data sources provide statistics on new businesses that represent populations similar to the population represented by the KFS. These will be discussed next.

C. OTHER DATA SOURCES ON BUSINESSES IN THE UNITED STATES

There are several sources of data on U.S. businesses. The Panel Study of Entrepreneurial Dynamics II (PSED II) is a cohort of more than 1,200 individuals and teams in the early stages of the

business creation process, selected in 2005 and then tracked over time. The study was designed to collect detailed information on a representative sample of these individuals, sometimes referred to as nascent entrepreneurs, as they move from conception to operating as a new firm. For the purposes of the PSED II, a new firm was defined as a business activity that reports a monthly revenue stream that is greater than monthly expenses for more than three consecutive months. For more information on the PSED II, please see Reynolds and Curtin (2008).

The U.S. Census Bureau provides the Office of Advocacy in the Small Business Administration with data on employer firm size in the Statistics of U.S. Businesses (SUSB). A firm is defined as the aggregation of all establishments owned by a parent company (within a geographic location and/or industry) that have some annual payroll. A firm may be located in one or more places. SUSB's employer data contain the number of firms, number of establishments, employment, and annual payroll for employment size of firm categories by location and industry. New firms are classified by their end-of-year firm size. The employer data consist of static and dynamic. Receipts by employment size of firm are available for 1997 and 2002, and special tabulations by receipt size of firm are available for the United States. Industries are defined according to Standard Industrial Classification for 1988 to 1998; and the North American Industry Classification System (NAIC) thereafter. For more information about the SUSB data, please see http://www.sba.gov/advo/research/data.html.

The Survey of Business Owners (SBO) is conducted by the U.S. Census Bureau every five years to collect statistics that describe the composition of U.S. businesses by gender, race, and ethnicity. This survey was previously conducted as the Survey of Minority- and Women-Owned Business Enterprises (SMOBE/SWOBE). The frame for this survey is compiled from several sources: IRS business tax returns, other Economic Census reports (e.g., Annual Survey of Manufacturers; Annual Retail Trade Survey), Social Security information on race, and Hispanic or Latino origin. The data are collected from a mailout/mailback survey. The universe for the most recent survey is all firms operating during 2002 with receipts of $1,000 or more that filed tax forms as individual proprietorships, partnerships, or any type of corporation. For more information, please see http://www.census.gov/csd/sbo/.

The 2003 Survey of Small Business Finances (SSBF) is the final survey of U.S. small businesses conducted by the Board of Governors of the Federal Reserve System on small business financing. This is the fourth survey; previous surveys were done in 1987, 1993, and 1998. The survey collects information on firm and owner characteristics, a firm's use of financial services as well as its financial service suppliers, and both income and balance sheet information. The 2003 survey collected demographic characteristics for up to three individual owners. For more information on the 2003 SSBF see Mach and Wolken (2006) and http://www.federalreserve.gov/boarddocs/surveys/.

D. COMPARISONS BETWEEN THE KFS AND OTHER DATA SOURCES

The Industry Distribution of New Firms

The differences in the industry distribution of firms between the KFS new employer firms, Census new employer firms, and new firms in the Panel Study of Entrepreneurial Dynamics are examined first. The distribution of new firms by industry is available from the Census Bureau for employer births only. As such, a subset of the KFS dataset, those firms with employees, are used in this comparison. As shown in Table 1, the KFS has a higher proportion of businesses in administrative and support, and waste management and remediation services; manufacturing; finance and insurance; information; and agriculture. The KFS also has a somewhat higher proportion of businesses in professional, management, and educational services; retail trade; and wholesale trade. The KFS has a lower proportion of businesses in construction; health care and social assistance; and accommodation and food services.

Differences between the KFS and the Census data collection efforts might account for these differences in industry distribution. The Census data are administrative data collected from unemployment insurance filings, while the KFS data are survey data

collected from the Dun & Bradstreet sampling frame. It also is possible that the survey methodology may be more likely to reach firm founders in some industries, while the collection of administrative data may be more likely to reach firm founders in other industries.

The industry distribution of new businesses in the KFS population also differs from the industry distribution of new businesses in the PSED II. For instance, larger portions of the businesses in the KFS are in technical services and manufacturing than they are in the PSED II. The industry distribution might be an artifact of two factors that make the KFS different from the PSED II.

First, the PSED II data collection effort is based on a representative sampling of the adult-age population of the United States, while the KFS is based on the achievement of several screening criteria on businesses that have entered the D&B database in the year of investigation. Second, the

Table 1
Firm Distribution by Industry

	KFS New Employer Businesses	Census Employer Firm Births	PSED New Businesses
Professional, Management, and Educational Services	15.5	14.1	16.8
Retail Trade	15.6	12.0	18.6
Administrative and Support, and Waste Management and Remediation Services	11.4	6.0	2.1
Construction	9.8	15.7	10.0
Other Services (except Public Administration)	8.0	8.5	0.3
Manufacturing	7.2	3.2	3.5
Wholesale Trade	6.0	4.5	1.5
Real Estate, Rental and Leasing	3.7	5.1	5.3
Finance and Insurance	4.7	2.2	3.1
Health Care and Social Assistance	4.2	7.7	2.9
Information	2.6	1.4	4.2
Transportation and Warehousing	2.9	3.3	2.4
Arts, Entertainment, and Recreation	3.1	2.1	3.2
Accommodation and Food Services	3.9	9.1	10.9
Agriculture, Forestry, Fishing, and Hunting	1.4	0.4	2.0
Mining	0.0	0.3	0.5
Utilities	0.0	0.1	0.5
Management of Companies and Enterprises	0.0	0.1	6.7
Unclassified	0.0	2.2	5.6

Source: Kauffman Firm Survey, Baseline data; Tabulations by Mathematica Policy Research, Inc; and Office of Advocacy, U.S. Small Business Administration, from longitudinal data (established with some first-quarter payroll) provided by the U.S. Census Bureau; Reynolds, P. 2004. Entrepreneurship in the United States Assessment, Miami, Fla.: Florida International University.

PSED II samples people in the process of starting a business, while the KFS samples new firms from a business frame. If the process of transforming a startup effort into a new business differs across industries, then the industry distribution of new businesses in the KFS and PSED will be different.

Employment Size Distribution of New Firms

This section compares the employment size distributions of new employer businesses in the KFS and Census data. The employment size distribution of the new employer firms as measured by the Census Bureau and new firms as measured by the KFS are remarkably similar. For instance, 10.4 percent of the employer new firms in the KFS have ten or more employees as compared with 10.3 percent of the new establishments in the Census data (Table 2). The estimate of the proportion of new employer firms with 100–499 employees in the KFS comes from five firms, yet it is quite close to the Census estimate. The small sample size in the KFS is likely to explain the absence of new businesses with 500 or more employees. Very, very few new businesses start with 500 or more employees in their first year. It's not surprising that the KFS, which surveyed nearly 5,000 new firms, did not capture any of these rare startups in its sample.

The similarity between the KFS and the Census data on the distribution of new employer firms by employment size does raise some important questions. Given the differences in the industry distribution of new firms between the two sources and the very different methodologies used by the two research efforts, one might wonder why the employment size distributions are so similar. What accounts for the difference in the distribution across the sources in the proportion of new businesses with zero employees, but such similarity once there is at least one employee? Why is the employment size distribution so similar when Census data shows differences across industries in the employment size distribution of new employer firms?

Table 2
Distribution of Firms by Employment Size Category

	Distribution of Assets For All Firms 2004	Distribution of Assets For All Surviving Firms 2007
1–4	75.1	76.7
5–9	14.6	13.0
10–19	6.6	6.0
20–99	3.4	3.8
100–499	0.3	0.4
500+	0	0[1]

1. This number rounds to 0.0. The actual number is 0.04.
Source: Kauffman Firm Survey, Baseline data; Office of Advocacy, Small Business Administration, Statistics of U.S. Businesses, U.S. Census, 2004.

Survival Rate of New Firms

As shown in Table 3, the 91.4 percent one-year survival rate of firms in the KFS is much higher than the one-year survival rate for new businesses found in other studies. For instance, the one-year survival rate of new, single-establishment firms founded in 1997 shown in a special tabulation created by the U.S. Small Business Administration is 75 percent. Moreover, this 75 percent survival rate is very close to the same as the one-year survival rate for several other cohorts of new, single-establishment firms, different years of the Census of Business Owners, and studies using Dun & Bradstreet data (Shane, 2008).

The higher-than-average survival rates in the KFS might result in part from the Dun & Bradstreet screening for 2004 startups. Data released for the KFS survey were from June and November 2005 D&B files, and interviewing for the Baseline continued until July 2006. Thus, if a firm started in 2004 and closed shortly afterwards, it could be missing from the files that became the frame from which the KFS sample was drawn. In addition, if the firm was not in business at the time of the survey, it would not have been asked to complete the Baseline. Those firms that both started and closed

down during the calendar year 2004 would have been screened out of the full KFS survey. Overall, more than 17 percent of sampled businesses were identified as out of business throughout the data collection effort. Thus, some of the first-year survival attrition was excluded during the July 2005 to July 2006 period. Because the one-year survival rates in the KFS are different from those found in other studies, users of these data are cautioned to be careful in drawing inference to larger populations when examining business survival. As further follow-up years become available, survival rates of the KFS firms likely will converge more closely to rates from other data sources.

Location of New Firms

The KFS and Census data (Survey of Business Owners) show very similar percentages of home-based businesses (Table 4). Both show that roughly 49 percent of firms are home-based. The similarity is striking, given that the KFS measures new firms, while the SBO measures all firms that were operating in 2002. (A breakout of business location by firm age was not available in the published statistics from the U.S. Census Bureau). This suggests that neither cohort effects nor selection effects cause the KFS and Census data to differ on the proportion of home-based differences. However, users of these data should be careful to recognize that, while offsetting factors in the differences in the KFS and Census methodologies could account for the similarities in the results found using the two data sources, selection and cohort effects cannot be among those offsetting factors. Previous research has indicated that businesses that start as home-based businesses are likely to remain home-based over the life span of the business.

Table 3
One-Year Survival Rates by Business and Primary Owner Characteristics

	2004 KFS	2002 SBO Starts
All	91.4	71.3
Male	92.4	70.2
Female	89.2	68.8
Equally Owned		83.3
Agriculture, Forestry, Fishing, and Hunting	90.5	64.9
Mining	n/a	86.0
Utilities	n/a	90.3
Construction	93.3	63.0
Manufacturing	94.4	79.3
Wholesale	90.0	78.7
Retail	89.5	76.2
Transportation	86.9	67.5
Information	88.4	67.2
Finance and Insurance[1]	95.1	78.8
Real Estate, Rental, and Leasing	93.7	84.7
Professional, Scientific, and Tech Services	90.9	70.0
Administration and Support, and Waste Management	94.4	n/a
Health Care and Social Services	88.7	70.6
Arts, Entertainment	88.7	75.0
Recreation, Accommodation, and Food Services	84.1	85.1
Other Services (except public administration)	92.3	n/a

1. The SBO column includes Real Estate here instead of in Real Estate, Rental, and Leasing.

Source: Kauffman Firm Survey, Baseline and First Follow-Up data; Tabulations by Mathematica Policy Research, Inc.; and Estimation Based on Special Tabulation of the SBO Provided by the Office of Advocacy of the U.S. Small Business Administration.

Table 4
Business Location

	2004 KFS (New Firms)	2002 SBO (All Firms)
Residence—Home/Garage	49.2	49.4
Rented/Leased Space	40.5	
Other	10.3	

Source: Kauffman Firm Survey, Baseline data; U.S. Bureau of the Census, Survey of Business Owners (2002).

Intellectual Property of New Firms

Due to the absence of other data sources that measure the intellectual property ownership of new firms, it is not possible to compare the KFS to other sources on this dimension. Although comparisons cannot be made to other datasets, examination of the KFS data still provides several points of guidance to data users. First, as seen in Table 5, ownership of intellectual property by new firms is rare in their first year of operations. Only 2.2 percent of new firms have patents; only 8.7 percent of new firms have copyrights; and only 13.5 percent of new firms have trademarks. Therefore, when discussing new firms, uses of these data should assume that intellectual property ownership is the exception rather than the rule.

Table 5
Business Patents, Copyrights, and Trademarks

	Sample Count	Weighted Percentage
Percent with patents		
All	187	2.2
High Tech	137	4.1
Low Tech	50	1.8
Percent with copyrights		
All	485	8.7
High Tech	242	11.4
Low Tech	243	8.1
Percent with trademarks		
All	721	13.5
High Tech	327	13.0
Low Tech	394	13.6

Source: Kauffman Firm Survey, Baseline Data; Tabulations by Mathematica Policy Research, Inc.

Second, the technology intensity of an industry affects the ownership of patents and copyrights, but not trademarks. Therefore, when discussing the effects of the R&D intensity or technical employment of an industry on the ownership of intellectual property by new firms, users of these data should confine themselves to a discussion of patents and copyrights, rather than referring to "intellectual property" in general.

Third, operating in a high-tech industry is not a synonym for intellectual property ownership. While new firms in high-tech industries are more likely than new firms in low-tech industries to have patents and copyrights—being roughly twice as likely to have patents and 41 percent more likely to have copyrights—the vast majority of high-tech firms do not own intellectual property.

Owner Demographics

The KFS collects demographic information on up to ten owners for each business. About 65 percent of the KFS firms have just one owner, while 26 percent have two owners, and 9 percent have three or more owners. In the case of multi-owner firms, researchers may want to define a primary owner using ownership share, hours worked, or some other measure. In the first column of Table 6, a primary owner was designated for multi-owner firms by the largest equity share. In cases where two or more owners owned equal shares, hours worked and a series of other variables were used to create a rank ordering of owners to define a primary owner. The second column uses the Census Bureau's definition of 51 percent equity ownership to designate ownership by various owner demographics, such as race and gender. Some cases, in which firms had no one racial group owning 51 percent or more, are listed separately.

The demographics of the owners in the KFS are different from the demographics of the owners in the SBO and PSED II using both of the measures described above. For instance, in the PSED II, 58 percent of the people in the process of starting a business are male and 42 percent are female. In the KFS, using the Census definition, 63 percent of the businesses are male-owned, but only 28 percent are female-owned. In the SBO, 54 percent of the

businesses that started in 2002 are male-owned, while 36 percent are female-owned.

The largest differences across racial and ethnic groups were the much higher percentages of black-owned businesses in the KFS and PSED II, compared with the Census Bureau, and the lower rates of Asian-owned businesses. One difference might result from selection effects present in the SBO that are not present in the KFS. Because the SBO is conducted several years after a business was founded but the KFS is conducted in the following year after the business was founded, the SBO is less likely to include the founders of businesses that close. Because businesses founded by people of different races, ethnicities, and gender fail at different rates, the differences in the distribution of new businesses across demographic characteristics in the two studies could be the result of differences in the degree to which the samples display selection bias.

Another difference might result from how the questions about demographics are asked. For example, the differences in the proportion of female firm founders in the two samples might result from efforts by the SBO and KFS to separately classify jointly founded firms—firms founded by men and women. The PSED II does not classify jointly founded firms because it examines the demographics of the owners themselves, not the firms that were founded. Thus, the KFS gender distribution could differ from the PSED II gender distribution because joint founders that were equally male and female were not a separate allocation. It could differ from the SBO if jointly founded firms are disproportionately classified as male-founded in the KFS. Then the proportion of female-founded firms would be higher in the SBO than in the KFS.

A third difference might result from differences in the languages used to conduct the survey. The greater use of Spanish-language survey materials and surveyors in the SBO might account for a higher participation rate among Hispanics. Those Hispanics who are not comfortable responding to surveys in English might be more likely to answer the SBO than to answer the KFS.

Table 6
New Firm Owner Demographics

	Weighted Percentage			
	Primary Owner 2004 KFS	Census Definition 2004 KFS	New Firms 2002 SBO	New Startups 2005 PSED II
Sex				
Male	69.2	62.8	53.958.	
Female	30.8	27.7	35.7	41.9
Equally Owned		9.6	10.1	
Race				
White	81.2	75.9	88.5	73.6
Black	9.2	9.1	4.6	18.6
Asian	4.2	4.5	6.1	
Others	5.5	2.2	0.9	7.9
No predominant race (>50%)		8.3		
Ethnicity				
Non-Hispanic	93.4	94.4	92.5	
Hispanic	6.6	5.6	7.5	4.7

Source: Kauffman Firm Survey, Baseline data; U.S. Bureau of the Census, Survey of Business Owners (2002), Special Tabulation, Panel Study of Entrepreneurial Dynamics II.

One other important difference is that new firms in the KFS are new firm startups. In the SBO, new firms are newly owned, which includes some acquired and inherited businesses. While the KFS does not include acquired and inherited businesses, it does include some businesses that were sold to entirely new ownership. Thus, different demographics could easily be an artifact of unobserved selection, cohort effects, and different definitions of starts.

Legal Form

New businesses must take on a legal form. As seen in Table 7, the KFS, SBO, and PSED II do not show the same distribution of businesses by legal form. The SBO and PSED II show much higher proportions of sole proprietorships and a much lower percentage of corporations than the KFS does. The PSED II also shows a much higher proportion of partnerships than the SBO and KFS do. The differences in the distribution of legal form of businesses in the different data sources may stem from differences in the data collection methodology. In addition, the process of selecting only those firms that meet certain development milestones, such as applying for an employment identification number or unemployment insurance, may rule out a greater proportion of sole proprietors and partnerships, yielding fewer of these in the KFS than in the PSED II or SBO data. Furthermore, use of the D&B sampling frame may bias the KFS sample toward corporations. Because one of D&B's goals is to identify firms that have sought credit, it may have more limited liability entities among its population of new firms, since those new firms that take legal forms without limited liability may be less likely to seek credit.

Use of Financial Capital

A focus of the Kauffman Firm Survey was determining how new firms are financed. The survey collects information on new capital infusions over the life of the firm, in addition to financial capital needed to launch the firm. The KFS contains detailed information on both debt and equity financing, including the amounts and sources of each, at startup and over time.

The closest data source with this type of information is the Survey of Small Business Finances (SSBF), which is done every five years by the Board of Governors of the Federal Reserve System. Unlike the KFS, which is a panel tracked over time, the SSBF measures a cross section of firms in a given time period. The SSBF was done in 1987, 1993, 1998, and 2003. Like the KFS, the frame for the SSBF is Dun & Bradstreet. However, while the SSBF surveys firms of all ages, the KFS is a longitudinal survey of one cohort of firms that began operations in 2004, collecting information from these same

Table 7
Legal Form of Organization

	PSED New Firms	KFS New Firms	SBO 2002 Starts
Not Yet Determined/Other	1.0	0.0	0.5
Sole Proprietorship	59.8	35.8	83.5
General/Limited Partnership	21.3	5.7	5.9
Limited Liability Company	10.2	30.5	n/a
Subchapter S Corporation	3.8	20.1	10.2*
General Corporation	3.8	7.9	

* Includes both C and S corporations
Source: Kauffman Firm Survey, Baseline data; U.S. Bureau of the Census, Survey of Business Owners (2002); Reynolds, P. 2004. Entrepreneurship in the United States Assessment, Miami, Fla: Florida International University; Office of Advocacy, U.S. Small Business Administration, from special tabulations of U.S. Census Bureau, 2002 Survey of Business Owners data.

firms annually. To make the two datasets more comparable, a subset of the SSBF, which includes just those firms that were one or two years old, is compared with the KFS firms.[6] The questions asked of firms differed somewhat across the two surveys, making one-to-one comparisons impossible. However, some useful comparisons still can be made.

Table 8 contains information from the Kauffman Firm Survey on the sources of debt and equity financing by 2004 startups in their initial year of operations (2004), while Table 9 contains information on new firms from the Survey of Small Business Finances. As shown in Table 8, nearly one-half of the KFS businesses had some type of personal debt related to the business in the first year of operations and nearly one-quarter had some type of business debt. About 30 percent of businesses carried balances on personal credit cards that were used to finance business activities and about half that (15 percent) carried balances on business credit cards. About 18 percent of businesses had a personal bank loan that was used for business purposes and about 10 percent of businesses borrowed from their families to finance business activities.

New businesses used personal debt sources with much higher frequencies than they used business debt sources. Less than 7 percent of businesses had a business bank loan in their first year, while 5.5 percent of businesses had balances on business credit lines. Less than 1 percent of businesses used a government business loan in their first year of operations. Business credit cards were the most frequently used form of business financing for business startups in the KFS. Finally, about a quarter of new businesses took advantage of trade credit in their first year of operations.

On the equity side, internal equity was the most frequently used source of startup capital. Nearly 80 percent of business owners invested their own funds to launch their business ventures. Very few used external equity. Less than 10 percent of firms had any outside equity in their first year of operations. Parents were the most common source of those funds (3.4 percent); spouses were tapped less often (1.6 percent). Outside investors, such as angel investors or venture capitalists, were used infrequently (2.7 percent and 0.6 percent respectively).

Table 8
2004 Kauffman Firm Survey
Percent of Businesses with Debt and Equity by Financing Source

	Weighted Percentage
Personal Debt of Any Kind	**48.1**
Personal Credit Card Balances by Owner(s)	30.2
Personal Bank Loan by Owner(s)	18.0
Business Credit Card Balances by Owner(s)	14.6
Family Loan by Owner(s)	10.1
Other Personal Loan by Owner(s)	2.0
Other Personal Debt by Owner(s)	1.3
Business Debt of Any Kind	**24.4**
Bank Business Loan	6.6
Business Credit Line Balance	5.5
Family Business Loan	2.9
Non-bank Business Loan	1.7
Owner Business Loan	1.5
Government Business Loan	0.9
Other Industry Business Loan	0.5
Other Business Debt	0.5
Other Business Loan	0.3
Use of Trade Financing	**23.9**
Internal Equity	**78.7**
External Equity of Any Kind	**9.6**
Parents	3.4
Non-family Informal Investors	2.7
Spouses	1.6
Other Companies	1.1
Venture Capitalists	0.6
Government	0.5

Source: Kauffman Firm Survey, Baseline data.

6. Given the small sample sizes of young firms in the SSBF, using just one-year-old firms didn't yield a sufficiently large sample size to do a comparison.

As previously mentioned, the patterns of financing by one- and two-year-old firms from the 2003 Survey of Small Business Finances are shown in Table 9. While the use of business and personal credit cards for business purposes both were between 43 percent and 44 percent, businesses were more likely to carry balances on business credit cards (39.3 percent of those using business credit cards or 16.9 percent of all businesses) than on personal credit cards (30.7 percent of those using personal credit cards or 13.5 percent of all businesses). About 43 percent of businesses used trade credit, while more than a quarter of businesses used a line of credit in their first two years of operations. Nearly 40 percent of corporations used stockholder loans as a source of financing. Nearly 17 percent of one- and two-year-old businesses in the SSBF made use of "other loans," which included friends and family members. This percentage is in line with the proportions in the KFS using friends and family for personal and business loans.

The percentage of businesses with equity was slightly lower than that found in the KFS. This equity figure is predominantly internal equity, but a very small percentage is external equity. External equity sources for new firms are not listed separately in the table above because the sample size of new businesses in the SSBF with external equity was so small that it was not possible to present reliable incidence rates. The infrequent use of external equity by new firms in the SSBF is consistent with the patterns found in the KFS.

Another source of data on the use of financial capital by new firms is the U.S. Census Bureau's Survey of Business Owners, which collects data on the sources of capital used to start the business. Unfortunately, the amounts of capital were not available in 2002. These numbers are not strictly comparable to the KFS numbers because the SBO measures all firms that were operating in 2002, while the KFS measures just 2004 startups. Yet, broadly speaking, the two sources are consistent

Table 9
2003 Survey of Small Business Finances: One- and Two-Year-Old Firms

	Percent of Firms
Use of Personal Credit Cards for Business Purposes	**44.0**
Carrying Credit Card Balances (of Those that Used Credit Cards)	30.7
Carrying Credit Card Balances (of All Businesses)	13.5
Use of Business Credit Cards	**43.1**
Carrying Credit Card Balances (of Those that Used Credit Cards)	39.3
Carrying Credit Card Balances (of All Businesses)	16.9
Use of Trade Credit	**43.4**
Use of Stockholder Loans	**39.2**
Use of Lines of Credit	**26.0**
Use of Motor Vehicle Loans	**19.6**
Use of Other Loans (Includes Friends, Relatives, and Other)	**16.7**
Use of Mortgages	**13.5**
Use of Capital Leases	**11.7**
Use of Equipment Loans	**5.5**
Equity	71.8

Source: Weighted Tabulation of Survey of 2003 Small Businesses Finances microdata.

with one another. As shown in Table 10, the 2002 SBO data show that personal and family resources are by far the most common source of startup capital, while outside investors and government sources are very infrequent. The data show that, similar to the KFS and SSBF, equity financing is used more frequently than debt financing for startup.

Table 10
Sources of Captial Needed to Start or Acquire the Business (2002 Survey of Business Owners, All Firms)

	Percent of Firms
Personal/family savings	54.6
Other personal/family assets	9.0
Personal/business credit card	8.8
Business loan from government	0.9
Government-guaranteed bank loan	0.7
Business loan from bank	11.4
Outside investor	2.7
None needed	27.7

Source: U.S. Census Bureau, 2002 Survey of Business Owners

E. CONCLUSIONS

This appendix has provided an overview of the Kauffman Firm Survey data and other sources of data on businesses in the United States. The KFS data are compared to other data sources along a number of dimensions. As shown in the previous sections, the KFS has similarities and differences with other published business data by industry, employment size, survival, location, intellectual property, owner demographics, legal form, and use of financial capital. Given the different target populations underlying each data source, differences are not surprising.

The main goal of collecting data for the Kauffman Firm Survey was to provide researchers a source of data to allow a deeper examination into how businesses organize and operate in their early years of operation and to determine the main factors driving survival and growth. The KFS offers some unique advantages, such as its longitudinal nature, its oversample of high-tech firms, detailed financing information at startup and over time, and its level of detailed information on both the firm and its owners.

The KFS dataset provides researchers with a unique opportunity to study a panel of new businesses from startup to sustainability, with longitudinal data centering on topics such as how businesses are financed; the products, services, and innovations these businesses possess and develop in their early years of existence; and the characteristics of the people who own and operate them.

While there is some difficulty in directly comparing the KFS data to other business data sources due to the different target populations represented by the various data sources, the KFS provides an excellent opportunity for researchers to study a myriad of topics related to new firm dynamics.

Ballou, Janice, Tom Barton, David DesRoches, Frank Potter, E.J. Reedy, Alicia Robb, Scott Shane, and Zhanyun Zhao. (2008). Kauffman Firm Survey: Results from the Baseline and First Follow-Up Surveys. Kauffman Foundation.

Mach, Traci L., and John D. Wolken. (2006). "Financial Services Used by Small Businesses: Evidence from the 2003 Survey of Small Business Finances." *Federal Reserve Bulletin*, October 24.

Reynolds, Paul D. (2004). The Entrepreneurship in the United States Assessment, Miami, Fla: Florida International University.

Reynolds, Paul D., and Richard T. Curtin. (2008). "Business Creation in the United States: Panel Study of Entrepreneurial Dynamics II Initial Assessment." Foundations and Trends in Entrepreneurship. 4(3):155-307.

Shane, Scott. (2008). The Illusions of Entrepreneurship: The Costly Myths that Entrepreneurs, Investors, and Policy Makers Live By. New Haven: Yale University Press.

KAUFFMAN

The Foundation of Entrepreneurship

4801 ROCKHILL ROAD
KANSAS CITY, MISSOURI 64110
816-932-1000
www.kauffman.org

0309400FP