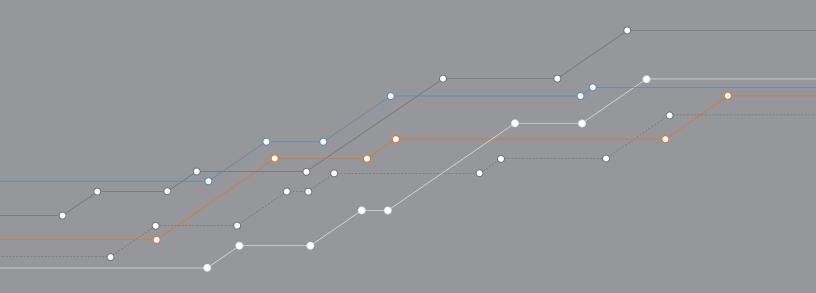# METHODOLOGY
FOR EARLY-STAGE ENTREPRENEURSHIP

Methodology

This section of the report discusses the methodology and underlying data sources for each of the Kauffman Indicators of Early-Stage Entrepreneurship and the methodology for calculating the summary KESE Index.

The underlying definitions and methodology are the same for the national and state estimates, with appropriate adjustments for geography and population size by state.

Indicator 1: Rate of New Entrepreneurs

The rate of new entrepreneurs is calculated using a special panel dataset created from the Current Population Survey (CPS). The CPS is a monthly survey of approximately 60,000 households conducted by the Bureau of Labor Statistics on behalf of the U.S. Census Bureau. The survey primarily asks questions focused on the employment status of household members, including their employment and business ownership status.[1] The CPS microdata capture all business owners, including those who own incorporated or unincorporated businesses, and those who are employers or non-employers. To create the rate of new entrepreneurs,[2] all individuals who do not own a business as their main job are identified in the first survey month. By matching monthly CPS files, it is then determined if these individuals own a business as their main job with 15 or more hours worked per usual week in the following survey month.

Changes to respondents' main jobs from month to month are measured accurately because CPS survey takers ask whether the individual has the same main job that they reported in the previous month. If the answer is yes, the interviewer carries forward job information, including business ownership, from the previous month's survey. If the answer is no, the respondent is asked the full series of job-related questions. Survey-takers ask this question at the beginning of the job section to save time during the interview process and improve consistency in reporting.

The main job is defined as the job with the most hours worked. Individuals who start side businesses will, therefore, not be counted if they are working more hours on a wage/salary job. The requirement that business owners work 15 or more hours per week in the second month is imposed to rule out part-time business owners and very small business activities.

The rate of new entrepreneurs may, therefore, underestimate or overestimate the percent of individuals creating any type of business.

The rate of new entrepreneurs excludes individuals who owned a business and worked fewer than 15 hours in the first survey month. Thus, it does not capture business owners who increased their hours from less than 15 per week in one month to 15 or more hours per week in the second month. It also does not capture when these business owners changed from being non-business owners to business owners with less than 15 hours worked. These individuals are excluded from the sample but may actually have been at the earliest stages of starting a business.

At the same time, the rate of new entrepreneurs may overstate entrepreneurship because of how individuals report their work status. Longstanding business owners who are also salaried in the business may, for example, not report that business ownership is their main job if their wage/salary jobs had more hours in that particular month. If these individuals later report having worked more hours in business ownership in a subsequent month, it would appear that a new business had been created.

For the rate of new entrepreneurs calculations presented in this report, all observations from the CPS with allocated labor force status, class of worker, and hours worked variables are excluded. The rate of new entrepreneurs is substantially higher for allocated or imputed observations.

Indicator 2: Opportunity Share of New Entrepreneurs

Building from the same data used for the rate of new entrepreneurs, the opportunity share of new entrepreneurs is defined as the share of the new business owners that are

---

1. https://www.census.gov/programs-surveys/cps.html.

2. This measure was created by Fairlie (2014), formerly known as the Kauffman Index of Entrepreneurial Activity.

coming out of wage and salary work, school, or other labor market statuses. This "opportunity entrepreneurship" can be contrasted to the "necessity entrepreneurship" that occurs when individuals start businesses coming out of unemployment. The opportunity share of new entrepreneurs considers individuals' initial labor market status in the first survey month.

The distinction between opportunity versus necessity has been discussed extensively in the entrepreneurship literature.[3] It is conceptually useful because the motivations for starting a business could influence the type, nature, and future direction of the business; it is also meaningful because it reflects to some extent the landscape of economic opportunity for entrepreneurs. Although there is some convergence about the theoretical distinction between the two motivations for business creation, a clean distinction is difficult to make with empirical data. Distinguishing between opportunity and necessity entrepreneurship using prior labor market status presents a useful approach.

UNDERLYING CURRENT POPULATION SURVEY (CPS) PANEL DATA

To calculate the rate of new entrepreneurs and the opportunity share of new entrepreneurs, a special panel dataset is created by matching the basic monthly files of the Current Population Survey (CPS) over time. These surveys, conducted monthly by the U.S. Census Bureau and the Bureau of Labor Statistics, represent the entire U.S. population and contain observations for more than 130,000 people each month. By linking the CPS files over time, longitudinal data are created, allowing for the examination of month-to-month changes in business creation. Combining the monthly files creates a sample size of roughly 700,000 adults ages 20 to 64 each year.

This method of creating panel data takes advantage of the household surveying strategies used for the CPS. Households in the CPS are interviewed each month over a four-month period. Eight months later, they are re-interviewed in each month of a second four-month period. Thus, individuals who are interviewed in January, February,

March, and April of one year are interviewed again in January, February, March, and April of the following year. The CPS rotation pattern makes it possible to match information on individuals monthly and, therefore, to create two-month panel data for up to 75 percent of all CPS respondents. To match these data, the household and individual identifiers provided by the CPS are used. False matches are removed by comparing race, sex, and age codes from the two months of data. After removing all non-unique matches, the underlying CPS data are checked extensively for coding errors and other problems.

Monthly match rates are generally between 94 percent and 96 percent. Household moves are the primary reason for non-matching. A somewhat non-random sample (mainly geographic movers) will, therefore, be lost due to the matching routine. Moves do not appear to create a serious problem for month-to-month matches, however, because the observable characteristics of the original sample and the matched sample are very similar.

The CPS sample was designed to produce national and state estimates of the unemployment rate and additional labor force characteristics of the civilian, non-institutional population ages 16 and older.[4] The total national sample size is drawn to ensure a high level of precision for the monthly national unemployment rate. For each of the 50 states and the District of Columbia, the sample also is designed to guarantee precise estimates of average annual unemployment rates, resulting in varying sample rates by state.[5] Sampling weights provided by the CPS, which also adjust for non-response and post-stratification raking, are used for all national and state-level estimates.

Indicator 3: Startup Early Job Creation

Startup early job creation uses BED data to capture early-stage job creation among startup cohorts each year. To focus on early-stage business success, a one-year window is used to measure job creation. For this measure, startups are defined as new employer establishments that are younger than one year

---

3. See Fairlie and Fossen (2017) and Desai (2017), among others.

4. The civilian non-institutional population is defined as persons 16 years of age and older residing in the 50 states and the District of Columbia, who are not inmates of institutions (e.g., penal and mental facilities, homes for the aged), and who are not on active duty in the Armed Forces. This number is reported regularly by the Federal Reserve and is available here: https://fred.stlouisfed.org/series/CNP16OV.

5. See Polivka (2000).

old in a given year. The total employment generated by these startups in their first year is divided by the population to create the per capita startup early job creation measure.

Indicator 4: Startup Early Survival Rate

The startup early survival rate uses BED data to measure the percentage of new employer establishments that survive their first year of operation.

UNDERLYING BUSINESS EMPLOYMENT DYNAMICS (BED) DATA

Startup early job creation and startup early survival rate both use the U.S. Bureau of Labor Statistics, Business Employment Dynamics (BED) series. The BED is derived from the Quarterly Census of Employment and Wages (QCEW), or ES-202, program. The data include all establishments subject to state unemployment insurance (UI) laws and federal agencies subject to the Unemployment Compensation for Federal Employees program. It covers all employer establishments in the United States (approximately seven million establishments).

The BED data include numbers of businesses tabulated by firm age, establishment age, employment size, and geography (national and state). Firm age information is used to identify and measure the number of startups, defined as employer businesses younger than one year old.

Because the BED is based on underlying administrative data that covers the universe of employer establishments in the United States, sampling concerns such as standard errors and confidence intervals are irrelevant. Nonetheless, non-sampling errors still could occur. These could be caused, for example, by data entry issues or by businesses submitting incorrect employment data.

Kauffman Early-Stage Entrepreneurship Index

The KESE is calculated from the four indicators of entrepreneurship activity. It is an equally weighted index of the four normalized indicators. Each of the measures is normalized by subtracting its mean and dividing by its standard deviation (i.e., creating a z-score for each variable). This calculation creates a comparable scale for including the four measures in the KESE. We use national annual estimates from 1996 to 2017 to calculate the mean and standard deviation for each component. The same normalization method, which is based on national data, is used for both geographical levels—national and state—for comparability and consistency over time.

KAUFFMAN
INDICATORS *of*
ENTREPRENEURSHIP

Explore the Kauffman Indicators further at: **www.kauffman.org/indicators**