

## Introduction

Business Formation Statistics (BFS) are a product of the U.S. Census Bureau developed at the Center for Economic Studies in research collaboration with economists affiliated with Board of Governors of the Federal Reserve System, Federal Reserve Bank of Atlanta, University of Maryland, and University of Notre Dame.

Monthly Puerto Rico Business Applications provide timely and high frequency data on business applications. It measures business initiation activity (Business Applications Series) as indicated by applications for an Employer Identification Number (EIN) on the [IRS Form SS-4](#).

Monthly Puerto Rico Business Applications currently cover the period starting from July 2004 onwards at a monthly frequency. The Puerto Rico Business Applications Series are not included in U.S. or regional Business Applications Series.

## Data Sources

The data on business applications are based on applications for an Employer Identification Number (EIN) through filings of IRS Form SS-4.

## Concepts and Methodology

These series describe the business applications for tax IDs as indicated by applications for an Employer Identification Number (EIN) through filings of IRS Form SS-4. Business applications are presented in four different series reflecting different subsets of the applications for an EIN. All business applications series cover the period from 2004 JUL onwards. All Puerto Rico business applications series are not included in national and regional business applications series.

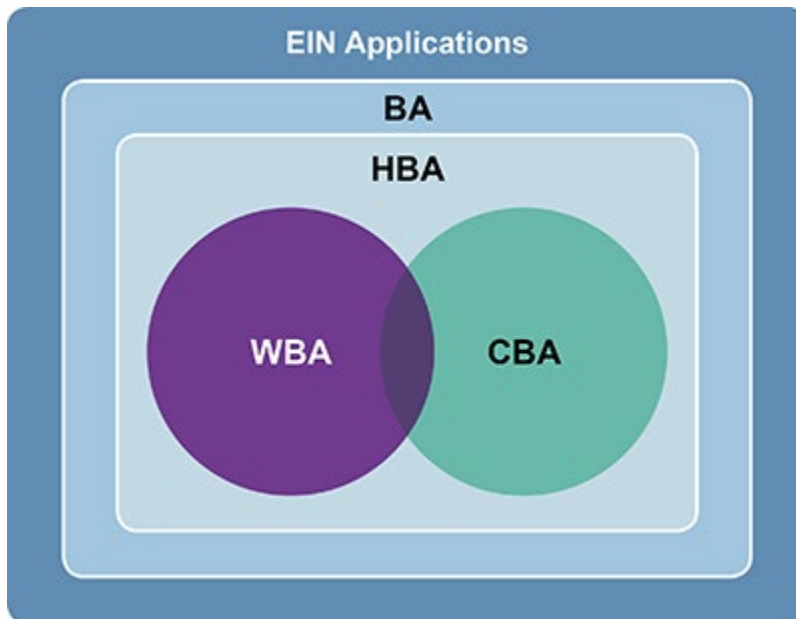
- **Business Applications (BA):** The core business applications series that correspond to a subset of all applications for an EIN. Includes all applications for an EIN, except for applications for tax liens, estates, trusts, or certain financial filings, applications outside of Puerto Rico, or with no state-county geocodes, applications with certain NAICS codes in sector 11 (agriculture, forestry, fishing and hunting) or 92 (public administration) that have low transition rates, and applications in certain industries (e.g. private households, civic and social organizations).
- **High-Propensity Business Applications (HBA):** Business Applications (BA) that have a high propensity of turning into businesses with payroll. The identification of high-propensity applications is based on the characteristics of applications revealed on the IRS Form SS-4 that are associated with a high rate of business formation. High-propensity

applications include applications: (a) for a corporate entity, (b) that indicate they are hiring employees, (c) that provide a first wages-paid date (planned wages); or (d) that have a NAICS industry code in accommodation and food services (72) or in portions of construction (237, 238), manufacturing (312, 321, 322, 332), retail (44, 452), professional, scientific, and technical services (5411, 5413), educational services (6111), and health care (621, 623).

- **Business Applications with Planned Wages (WBA):** High-Propensity Business Applications (HBA) that indicate a first wages-paid date on the IRS Form SS-4. The indication of a wages-paid date is associated with a high likelihood of transitioning into a business with a payroll.
- **Business Applications from Corporations (CBA):** High-Propensity Business Applications (HBA) from a corporation or personal service corporation, based on the legal form of organization stated in the IRS Form SS-4. Similar to the WBA series, this series is important primarily because it consists of a set of applications that have a high rate of transitioning into businesses with payroll.

The following is a Venn diagram of the relationship between the four business applications series (BA, HBA, WBA, CBA) and EIN applications.

## The Relationship Between Different Business Applications Series



## NAICS Improvement

The Census Bureau classifies BFS data by industry using the North American Industry Classification System (NAICS). NAICS codes are assigned using a variety of sources. A Census Bureau-developed automated industry-coding program first attempts to assign NAICS codes to all new EIN applications received from the IRS. The automated industry-coding program is based on established patterns in the business name and descriptions provided on the EIN applications. This auto-coding process assigned NAICS codes to over 80% of all incoming EIN applications in 2020. For applications that did not receive a NAICS code during the auto-coding process, BFS staff use a Census-developed machine learning algorithm to assign NAICS codes where possible. NAICS codes are revised each year for the previous five years as part of the BFS annual update process when more accurate and detailed NAICS codes may be available from the Social Security Administration, the Bureau of Labor Statistics, and the Census Bureau's Business Register. There are a small number of EIN applications where there is not enough information available to assign a NAICS code through any source.

Please read this [paper](#) for more information about the Census Bureau automated industry-coding program. This [presentation](#) provides more information on the machine learning algorithm.

## Reliability of the Data

Because BFS are constructed using a combination of administrative data, rather than a probability sample, sampling error does not apply to BFS. Non-sampling error, however, still exists. Non-sampling errors can occur for many reasons, such as the employer submitting corrected payroll or employment data after the end of the year as well as late filers. Other sources of error include typographical errors made by businesses when providing information on the survey or administrative forms. Such errors, however, are likely to be distributed randomly throughout the dataset.

Changes in administrative data sometimes can create complications in identifying business startups with payroll. There are also some changes in the content of the IRS Form SS-4 over time, and new information in the form is incorporated into the analysis as it becomes available.

BFS periodically evaluates the characteristics associated with high-propensity applications and their likelihood to turn to a business formation. The evaluation may result in updates to the definition of high-propensity applications. In November 2021, the definition for high-propensity business applications was updated and applied to data from 2012-current. This update was made possible in part by the NAICS improvement methodology discussed above.

**new HBA:** High-Propensity Business Applications - Business Applications (BA) that have a high-propensity of turning into businesses with payroll. The



identification of high-propensity applications is based on the characteristics of applications revealed on the IRS Form SS-4 that are associated with a high rate of business formation. High-propensity applications include applications: (a) from a corporate entity, (b) that indicate they are hiring employees, (c) that provide a first wages-paid date (planned wages); or (d) that have a NAICS industry code in accommodation and food services (72) or in portions of construction (237, 238), manufacturing (312, 321, 322, 332), retail (44, 452), professional, scientific, and technical services (5411, 5413), educational services (6111), and health care (621, 623).

The characteristics associated with high-propensity business applications were first determined during initial BFS program research. The original definition for high-propensity business applications is more representative of the older data and is applied to the data from 2004-2011.

**original HBA:** High-Propensity Business Applications - Business Applications (BA) that have a high-propensity of turning into businesses with payroll. The identification of high-propensity applications is based on the characteristics of applications revealed on the IRS Form SS-4 that are associated with a high rate of business formation. High-propensity applications include applications: (a) from a corporate entity, (b) that indicate they are hiring employees, purchasing a business or changing organizational type, (c) that provide a first wages-paid date (planned wages); or (d) that have a NAICS industry code in manufacturing (31-33), a portion of retail (44), health care (62), or accommodation and food services (72).

In order to keep the data between the definitions comparable, the original definition of high-propensity applications has an additional linking methodology applied. We did so by maintaining the growth rate of the time series but raising or lowering the level of the original definition series, increasing or decreasing its weight in the aggregation hierarchy. This method is known as retrapolation, see the [Handbook on Backcasting](#) for more information.

An adjustment factor was calculated using a simple ratio of the new HBA definition over the original HBA definition, for each month in an overlap year, 2012.

$$adj\_factor = \frac{\text{applications fulfilling the new HBA definition}}{\text{applications fulfilling the original HBA definition}}$$

The mean adjustment factor was then calculated across all months of 2012.

$$\bar{x} = \frac{\sum_{i=1}^N adj\_factor_i}{N}$$

The mean adjustment factor was then applied, via multiplication, to the data with the original HBA definition (2004-2011). For the years 2004-2011, the HBA data in our publications have reinterpolation applied.

$$\text{Published HBA data for 2004 – 2011} = \bar{x}(\text{original HBA})$$

The adjustment factor is computed at the state-level for the NSA HBA data. The region and national NSA HBA totals are tabulated from this state-level data. For the industry NSA HBA data, separate adjustment factors for each industry are computed. Seasonal adjustment methodology is applied to the updated spliced series.

## Seasonal Adjustment

Seasonal adjustment is the process of estimating and removing seasonal effects from a time series to better reveal certain nonseasonal features. Examples of seasonal effects include a July drop in automobile production as factories retool for new models and increases in heating oil production during September in anticipation of the winter heating season. When applicable, we also estimate and remove trading day effects and moving holiday effects (e.g., Easter, Labor Day, etc.) during the seasonal adjustment process. Trading day effects are recurring effects related to the weekday composition of the month. Because of strong seasonality detected in most of the business applications series, all series are provided with and without seasonal adjustment. Each month adjusted applications series are revised on for the prior two months, as well as the current and previous month in the prior year. For example, with the release of September 2021 data, the following months would be revised: August 2021, July 2021, September 2020, and August 2020. Factors for seasonal adjustments are recomputed and the seasonally adjusted applications series are revised annually.

Seasonal adjustment is performed concurrently using the X-13ARIMA-SEATS seasonal adjustment program of the U.S. Census Bureau. Concurrent seasonal factors result from re-estimating the seasonal adjustment each month or quarter when the new time series values become available. For more information on X-13ARIMA-SEATS, see the [reference manuals](#) posted on the Census Bureau's website. An assumption underlying the seasonal adjustment process is that the original series can be separated into a seasonal component, a trend-cycle component, and an irregular component, and possibly a trading day component and/or moving holiday component. The seasonally adjusted series consists of the trend-cycle and irregular components taken together. The trend-cycle component includes the long-term trend and the business cycle. The irregular component is made up of residual variations, such as the sudden impact of political events and the effects of strikes, unusual weather conditions, reporting and sampling errors, etc. Users can implement their own seasonal adjustment methods using the unadjusted data. User's results may differ from those published due to

rounding. The Census Bureau rounds in the final step after running seasonal adjustment.

## Disclosure Avoidance

Disclosure is the release of data that reveals information or permits deduction of information about a particular survey unit through the release of either tables or microdata. Disclosure avoidance is the process used to protect each unit's identity and data from disclosure. Using disclosure avoidance procedures, the Census Bureau modifies or removes the characteristics that put information at risk of disclosure. Although it may appear that a table shows information about a specific unit, the Census Bureau has taken steps to disguise or suppress a unit's data that may be "at risk" of disclosure while making sure the results are still useful.

The Census Bureau has reviewed the monthly Puerto Rico data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data (Project No. P-7515311, Disclosure Review Board (DRB) approval number: CBDRB-FY24-0098).