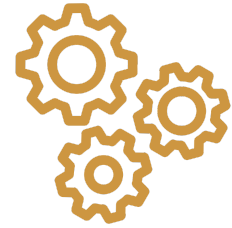


Analysis plan

Project name: Understanding rural small businesses that successfully win federal contracts

Project code: 2502

Date finalized: 6/26/2025



Project description

Rural small businesses are an integral part of local economies and communities. In March 2025, the Small Business Administration (SBA) [announced its new priorities](#), including efforts to empower rural small businesses. This initiative includes making SBA offices and services more accessible to rural communities by leveraging inter-agency, state, and local partnerships for economic and workforce development. The SBA also launched the [Rural Initiative pilot loan program](#) to offer rural small businesses greater opportunities to access capital, and other efforts to engage rural small businesses in the federal marketplace. This evaluation was developed to support efforts to better identify the sectors in which small rural businesses win federal contracts.

Objectives

The primary objective of this project is to provide actionable descriptive information to the SBA surrounding the scope of U.S. small businesses in rural areas, especially regarding opportunities for expansion of small rural businesses into the federal marketplace and for expansion of SBA programs to small rural businesses. A secondary objective is to understand how such gaps differ between small rural businesses awarded prime contracts versus those awarded subcontracts.

Anticipated utilization of results

The findings from this study can assist the SBA in targeting specific areas for growth and improvement by identifying sectors in which rural businesses are undersupplied in the federal marketplace and SBA contracting programs are underutilized among rural businesses. With these findings, the SBA can focus its outreach and programming efforts, allowing for more efficient support of Main Streets across the United States. For example, findings from this evaluation could inform future interventions to encourage rural firms that are a good fit for federal opportunities to join the marketplace and apply for SBA certification programs.

Research questions

To fulfill the objectives delineated above, this project will answer four general research questions:

1. How many rural small businesses (i.e., small businesses with physical headquarters in rural zip codes) are currently registered to do business with the federal government?
2. What sectors do these rural small businesses work in?

3. What are the rates of prime and subcontract awards to rural small businesses?
4. What proportion of rural small businesses registered to do business with the US government participate in SBA's contracting assistance programs (e.g., HUBZone)? Compared to potential eligibility, which programs are underutilized?

Target estimands

In this section, we describe the target estimands we would like to be able to estimate, in line with the four research questions described above. However, not all of these will be directly estimable, due to data constraints. We will thus use proxies where necessary. Our proxies (estimates) for these estimands are described in the “Descriptive Statistics” section below.

Definitions:

- Let N be the set of 2-digit NAICS codes¹ used by the federal government to track different sectors (i.e., broad categories of industries) in which businesses work. See the [Census Bureau's web page](#) for more context on what these sectors represent.
- Let B denote a particular business in the United States.
- Let R_B denote a binary variable where a value of 1 indicates that the business is registered to do business with the federal government in [SAM.gov](#).²
- Let S_B denote a binary variable where a value of 1 indicates that the business is a small business.³
- Let U_B denote a binary variable where a value of 1 indicates that the business is a rural business.
- Let SU_B denote a binary variable where a value of 1 indicates that the business is a small rural business.
- Let N_B be the 2-digit NAICS code of the business.
- Let $SBAI_B$ denote a series of binary variables where a value of 1 indicates that the business has participated in SBA contracting assistance program I , where $I \in I$ and I is the set of all SBA contracting assistance programs.
- Let SBA_B denote a binary variable where a value of 1 indicates that the business has participated in at least one SBA contracting assistance program.
- Let EI_B denote a series of binary variables where a value of 1 indicates that the business is likely eligible for SBA contracting assistance program I , where $I \in I$. (See the table under “Statistics, Tables, and Graphs” for details on how likely eligibility will be assessed for each SBA program.)
- Let E_B denote a binary variable where a value of 1 indicates that the business is likely eligible for one or more SBA contracting assistance programs.

¹ The NAICS code is a 6-digit code indicating the industry of a business. The first two digits represent individual “sectors.”

² As some states require [SAM.gov](#) registration in order to obtain state contracts, it is possible that not all businesses from these states registered in [SAM.gov](#) intend to do business with the federal government. The number of businesses may be overestimated for these states. We will note this possibility when we present our findings.

³ The definition of a “small” business per the SBA varies depending on the 6 digit NAICS code (see <https://www.sba.gov/document/support-table-size-standards>).

- Assume we observe the population of federal contracting awards. In this data, the unit of observation is the award, denoted by A .
- Let N_A be the 2-digit NAICS code associated with the award.
- Let P_A denote a binary variable where a value of one indicates that the award is a primary award.
- Let $C_A = 1 - P_A$ denote a binary variable where a value of one indicates that the award is a subcontract award.
- Let B_A be the business awarded the opportunity. With some abuse of notation, let $B_A = S$ denote a binary variable indicating that B_A is a small rural firm (i.e., the firm has $S_B = 1$ and $U_B = 1$, or equivalently $SU_B = 1$), as defined above.

Research question	Target estimand(s)
1. How many rural small businesses (i.e., small businesses with physical headquarters in rural counties) are currently registered to do business with the federal government? What proportion of registered businesses are rural small businesses?	Count of all $[SU_B R_B = 1]$ $E[SU_B R_B = 1]$
2. What sectors do these rural small businesses work in?	Count of all $[SU_B R_B = 1, N_B = n]$ for each n in N $E[SU_B R_B = 1, N_B = n]$ for each n in N
3. What are the rates of prime versus subcontract awards to rural small businesses? Relative to all prime awards in general and to prime awards for rural businesses?	$E[B_A = S P_A = 1]$ $E[B_A = S P_A = 1, U_B = 1]$ $E[B_A = S P_A = 1, N_A = n]$ for each n in N $E[B_A = S C_A = 1]$ $E[B_A = S C_A = 1, U_B = 1]$ $E[B_A = S C_A = 1, N_A = n]$ for each n in N
4. What proportion of rural small businesses registered to do business with the US government participate in SBA's contracting assistance programs (e.g., HUBZone)? Compared to potential eligibility, which programs are underutilized?	$E[SBA_B S_B = 1, R_B = 1, U_B = 1]$ $E[SBA_I S_B = 1, R_B = 1, U_B = 1]$ for each I in I $E[E_B S_B = 1, R_B = 1, U_B = 1]$ $E[E_I S_B = 1, R_B = 1, U_B = 1]$ for each I in I

Data and data structure

This section describes the variables that will be analyzed, as well as the changes made to the raw data regarding data structure and variables.

Population and expected number of observations:

There are two relevant populations for this study:

- **Population 1, R_B :** Businesses registered to compete for federal government contracts
- **Population 2, A :** Federal contract opportunities awarded to firms

Though we are not intentionally sampling from these populations, data limitations mean that we cannot necessarily observe any of these entire populations in a given year (see “Data Sources” below). However, we are drawing on the best data available to federal employees as well as, in many cases, to the broader public.

Data source(s):

Population 1, R_B :

- Firms register to do business with the federal government in [SAM.gov](https://sam.gov), which yields a Unique Entity Identifier (UEI) that identifies them across the federal procurement sphere. [SAM.gov](https://sam.gov) provides access to a variety of information about firms registered to do business with the government, including their state and address of incorporation, other potential addresses such as office address, self-attestations regarding whether they fall into categories such as “veteran-owned,” the industries (NAICS) they report working in, and their number of employees.
 - When firms register in SAM, they report the reason for doing so, which is tracked in a “purpose of registration” field within the SAM entity data. Firms interested in pursuing federal contracts are required to provide information on the NAICS codes in which they operate (they also report a “primary” NAICS code out of those listed).
 - We plan to download a SAM entity data extract for FY 2024 to address the other Research Questions.
- The Small Business Administration's (SBA) [Dynamic Small Business Search](https://dynamic.sba.gov) (DSBS) tool allows users to look up the UEIs of firms actively participating in government certification programs (e.g., certifying that a firm is a service-disabled veteran-owned business).

Population 2, A :

- We will consider two sources of data on federal contract awards (FY 2024):
 - Prime contract awards, P_A , are submitted to the *Federal Procurement Data System* (FPDS). FPDS has an [online portal](https://fpds.gov) you can use to search, and FPDS data can be downloaded from either a [SAM.gov](https://sam.gov) report builder (by federal employees), or in bulk from [USA Spending](https://usa.spending.gov) (by members of the public; select fields only). There are a [variety of conditions](#) determining which awards need to be reported to FPDS, including the dollar amount: awards of \$10,000 or more must generally be reported, all else equal. Note that some agencies, like DOD, report awards on a several-month delay for security reasons.

- Subcontract awards, C_A , were historically submitted to the *Federal Subaward Reporting System* (FSRS), but as of 5/8/2025 [this is being moved to SAM.gov](#). These data can also be downloaded from USA Spending. Again, there are a variety of conditions influencing whether a subcontract award must be reported. Among them is the dollar amount: only subcontracts of \$30,000 or more are required to be reported. This means that there will be prime awards in FPDS with subcontracts that are simply too small in terms of dollar value to appear in our data.
- Both prime contract award data and subcontract award data sources provide information on the recipient of the prime/subcontract award and the corresponding industry (NAICS code). These sources are not an exhaustive list of all contracts awarded across the federal government but represent the vast majority of prime contracts awarded across executive branch agencies and are the best available data on subcontract awards. In our analysis, we will describe the limitations of the datasets described above and the implications of those limitations.

Other data inclusion and measurement notes:

The current [SAM.gov](#) data does not explicitly identify whether a business is based or located in a rural region. Rural county designations will be mapped using Rural-Urban Commuting Area (“RUCA”)⁴ data mapped to the zip code level. To identify urban versus rural areas, we will use the Primary RUCA Code 2010⁵ variable in the [Rural-Urban Commuting Area](#) (RUCA) dataset provided by the U.S. Department of Agriculture. These data are provided at the Census tract level. RUCA code values of 1-3 will be coded as RURAL=0 and RUCA code values of 4-10 will be coded as RURAL=1. This indicates that metropolitan areas are classified as non-rural and micropolitan, small town, and rural areas as rural. A binary classification of zip-codes as metropolitan (urban) or not (rural) is coarse, but facilitates an initial investigation into which firms are participating in the federal marketplace that is easy to interpret, while also setting the stage for a more granular analysis of firms in that “rural” category in the future.

Datasets and variables:

This project will combine the data sources described above to generate two datasets. One will be drawn from [SAM.gov](#), DSBS, and the [Rural-Urban Commuting Area](#) (RUCA) dataset, at the level of the business (BUSINESS), and will be used to answer research questions 1, 2, and 4. The second will be drawn from FPDS/FSRS data in USA Spending at the level of the award (AWARDS), and, once merged on `entity` with the BUSINESS dataset, will be used to answer research question 3.

⁴ See [this resource](#) for more on characterizing geographies as urban/rural.

⁵ The 2020 updated RUCA data are not yet available. While there will be measurement error in using the 2010 data, we have no reason to believe this error would systematically bias our findings.

The variables in these datasets will be as follows:

Dataset	Variable name	Variable description	Original source
BUSINESS	entity	Unique entity ID	SAM.gov entity data
BUSINESS	n_employees	Number of employees	SAM.gov entity data
BUSINESS	primary_naics	6-digit NAICS code that the firm lists as its “primary” NAICS code	SAM.gov entity data
BUSINESS	all_naics	All NAICS codes a firm lists in SAM, concatenated by “~”. Including a flag for whether the business is small in each NAICS, as defined below.	SAM.gov entity data
BUSINESS	is_rural	Whether a small business is physically located in a rural area in the SAM set, per the classification rule discussed above	Rural-Urban Commuting Area (RUCA) dataset
BUSINESS	primary_naics2	First two digits of the firm’s primary NAICS code	SAM.gov entity data
BUSINESS	small_criteria	Maximum number of employees for small business designation in a firm’s primary NAICS	“Table of Size Standards” in Footnote 1
BUSINESS	is_sb_emp_def	Binary indicator for whether $employees < small_criteria$ in their primary NAICS	SAM.gov entity data
BUSINESS	is_small_sam	Binary indicator for whether a business is considered small based on their primary NAICS code (Y/N/Exception -> Y indicates small, E indicates in some spaces)	SAM.gov entity data
BUSINESS	physical_zip	Zip code of firm’s physical address	SAM.gov entity data

BUSINESS	cert_8a	Binary indicator for 8(a) certification	DSBS certification data
BUSINESS	cert_HUB	Binary indicator for HUBZone certification	DSBS certification data
BUSINESS	cert_vet	Binary indicator for VOSB or SDVOSB certification	DSBS certification data
BUSINESS	cert_wosb	Binary indicator for WOSB or EDWOSB certification	DSBS certification data
BUSINESS	self_SDB	Binary indicator for Small Disadvantaged Business (SDB) self-certification	SAM.gov entity data
BUSINESS	self_woman	Binary indicator for women-owned self-certification	SAM.gov entity data
BUSINESS	self_vet	Binary indicator for veteran-owned self-certification	SAM.gov entity data
AWARDS	awardID	AwardID	FPDS/FSRS (via USA Spending).
AWARDS	award_naics	6-digit NAICS code	FPDS/FSRS (via USA Spending).
AWARDS	award_naics2	2-digit NAICS code	FPDS/FSRS (via USA Spending).
AWARDS	award_amount	Amount of the award	FPDS/FSRS (via USA Spending).
AWARDS	award_set-aside	Small business set-aside utilization	FPDS/FSRS (via USA Spending).
AWARDS	prime	Binary indicator for prime award	FPDS/FSRS (via USA Spending).
AWARDS	entity	Unique entity ID of the awarded business	FPDS/FSRS (via USA Spending).

Data exclusion:

We will not exclude any data from our analysis.

Treatment of missing data:

We do not anticipate significant missingness on the key variables for this study: (1) those needed to classify firms as small businesses in different industries or to identify them as rural, (2) those needed to count awards and opportunities in different industries and determine which firms are receiving awards, or (3) the dollar value of awards. We provide more context below.

In the minority of cases where observations could be missing data for key variables, or when any other variables are missing, we will determine whether missingness is plausibly “missing-at-random” (quasi-random, at least after conditioning on covariates). If missingness does appear to be quasi-random, we will use listwise deletion for analyses relying on these variables, with appropriate caveats to our interpretation where necessary. If we encounter a substantial missingness pattern that is not plausibly missing-at-random, we will adhere to best-practice methodological guidance to address the issue and justify our solution in the Record of Analysis.

For instance, in the BUSINESS data, we need to measure whether a firm is a small business in different industries (NAICS and employees). All entities indicating interest in contract awards alone when registering in SAM are required to fill out the fields we use to classify them as a small business. There are a small number of firms indicating registration for multiple purposes that do not always provide this information, but in 2024, for instance they provide complete information in ~97% of cases. The only entities we cannot classify as small businesses are those registering only to apply for federal grant assistance, which we exclude from our analyses.⁶ Similarly, the most important fields we need from the AWARDS data, entity, and award_naics, are either required or automatically propagated for different kinds of award records. In FY 2022, for instance, only .0023% of awardIDs do not have a NAICS code, and all have both entity and award_amount.

Statistics, tables, and graphs

Anticipated descriptive statistics, tables, and graphs:

As mentioned above in the section on Target Estimands, we do not have access to data that would enable us to calculate each estimand perfectly. This section thus describes the proxies we will use as estimators for each estimand, given data limitations. See above more context on details like the years of data we will use and the exact data sources we will download.

⁶ The [FPDS data dictionary](#), page 115, indicates that an award entry cannot be registered in FPDS if the recipient firm's UEI has an incomplete SAM registration, including fields in the “Assertions” section we rely on for small business measurement. We take this as evidence that firms cannot be awarded contracts if they are in the “grants only” category in SAM entity data that does not provide fields we need.

We create the following estimates for our estimands of interest:

RQ	Estimand	Estimator	High level goal
1	Count of all [SU _B R _B =1]	Number of small rural firms registered in SAM	How many small rural firms are registered to do business with the government?
1	Count of all [U _B R _B =1]	Number of rural firms registered in SAM	How many rural firms are registered to do business with the government?
1	E[SU _B R _B =1]	Proportion of businesses registered in SAM that are rural and small	What fraction of those registered firms are small/rural?
1	E[U _B R _B =1]	Proportion of businesses registered in SAM that are rural	What fraction of those registered firms are any rural business?
2	Count of all [SU _B R _B =1] for each n in N	Number of small rural firms with NAICS sector n registered in SAM - for each n in N	How many small rural firms in SAM.gov work in each sector?
2	E[SU _B R _B =1] for each n in N	Number of small rural firms with NAICS sector n registered in SAM / Number of firms with NAICS sector n - for each n in N	What fraction of registered firms in each NAICS sector are small/rural?
3	E[B _A =S P _A =1]	Number of prime awards made to small rural firms / Total number of prime awards	What fraction of prime awards go to small rural businesses?
3	E[B _A =S P _A =1, U _B =1]	Number of prime awards made to small rural firms / Total number of prime awards to rural firms	What fraction of prime awards to rural businesses go to <i>small</i> rural businesses specifically?
3	E[B _A =S P _A =1, N _A =n] for each n in N	Number of prime awards with NAICS sector n made to small rural firms / Total number of prime awards with NAICS sector n - for each n in N	What fraction of prime awards in each sector go to small rural businesses specifically?
3	E[B _A =S C _A =1]	Number of subcontract awards made to small rural firms / Total number of subcontract awards	What fraction of subcontract awards go to small rural businesses?
3	E[B _A =S C _A =1, U _B =1]	Number of subcontract awards made to small rural firms / Total number of subcontract awards to rural firms	What fraction of subcontract awards to rural businesses go to <i>small</i> rural businesses specifically?
3	E[B _A =S C _A =1, N _A =n]	Number of subcontract awards with NAICS sector n made to small rural firms	What fraction of subcontract awards in each sector go to small rural

	for each n in N	/ Total number of subcontract awards with NAICS sector n - for each n in N	businesses specifically?
4	$E[SBA_B R_B=1, SU_B=1]$	Number of small rural firms in SBA data that have participated in at least one SBA contracting assistance program / Number of small rural firms registered in SAM	What fraction of small rural firms in SAM are certification program participants?
4	$E[SBA_I R_B=1, SU_B=1]$ for each I in I	Number of small rural firms in SBA data that have participated SBA contracting assistance program I / Number of small rural firms registered in SAM for each I in I	What fraction of small rural firms in SAM are certification program participants, by program?
4	$E[E_B R_B=1, SU_B=1]$	Number of small rural firms registered in SAM and categorized as likely eligible for at least one SBA contracting assistance program / Number of small rural firms registered in SAM	What fraction of small rural firms in SAM are eligible to be certification program participants?
4	$E[E_I R_B=1, SU_B=1]$ for each I in I	Number of small rural firms registered in SAM and categorized as likely eligible for SBA contracting assistance program I / Number of small rural firms registered in SAM for each I in I	What fraction of small rural firms in SAM are eligible to be certification program participants, by program?

Statistical inference: Our priority in this study is providing a descriptive exploration of the samples of businesses and contract awards outlined above, with the understanding that they are not perfect representations of our underlying populations of interest. That said, in any situation where it is meaningful to do so (i.e., where it is necessary for drawing appropriately careful inferences outside of our study sample), we will report 95% confidence intervals to represent our uncertainty about the “true” values of quantities above. We will document the reasoning for our decisions either way in the Record of Analysis.

Notes on definitions:

Small business: The definition of “small” varies by sub industry at the 6-digit NAICS code level. Each 6-digit NAICS code has an associated cutoff ranging from 500 to 1,500,

indicating the maximum number of employees a business can have to qualify as small.⁷ To calculate the number of small businesses registered in [SAM.gov](https://sam.gov), we will use information on a firm's number of employees available in the FOUO ("For official use only") entity data.

Eligibility for and participation in SBA Contracting Assistance Programs: There are [eight SBA contracting assistance programs](#). However, we do not include the SBA Mentor-Protege program, the Joint Ventures program, or the Natural Resources program, due to the specific natures of the programs (i.e., the difficulty of accurately incorporating them into this high-level analysis). Additionally, because there is no SBA certification for Small Disadvantaged Businesses, we also exclude that program. The table below lists each program and its eligibility criteria. It also defines how we will categorize the businesses in our data as "likely eligible" for each program. Since we do not directly observe true eligibility, we use these as proxies for it. Lastly, it defines program participation.

Program	Eligibility ⁸	"Likely eligible" definition	"Participated" definition
8(a) Business Development Program	Small business by SBA standards At least 51% owned and controlled by one or more persons who are socially and economically disadvantaged (with criteria on net worth, AGI, and assets) Demonstrates good character and potential for success	Business is small and self-certifies as a Small Disadvantaged Business (SDB)	Business has 8(a) certification from SBA
Women-Owned Small Business (WOSB) and Economically Disadvantaged Women-Owned Small Business (EDWOSB)	Small business by SBA standards At least 51% owned and controlled by women who are US citizens Have women manage day-to-day operations who also make long-term decisions	Business is small and self-certifies as a WOSB or EDWOSB	Business is certified by SBA as a WOSB or EDWOSB
Veteran-Owned Small Business (VOSB) and Service- Disabled Veteran-Owned Small Business	Small business by SBA standards At least 51% owned and controlled by one or more individuals identified by the VA as a veteran or	Business is small and self-certifies as a VOSB or SDVOSB	Business is certified by SBA as a VOSB or SDVOSB

⁷ Within some NAICS codes, businesses can obtain an exception to be categorized as small, even if their number of employees is greater than the cutoff. We will conduct our analysis assuming there are no exceptions, which could underestimate the number of small businesses in the affected NAICS codes. We will note the affected NAICS codes and discuss how this underestimation could affect our findings.

⁸ Source: <https://www.sba.gov/federal-contracting/contracting-assistance-programs> as of May 13, 2025.

(SDVOSB)	service-disabled veteran		
HUBZone Program	<p>Small business by SBA standards</p> <p>Be at least 51% owned and controlled by US citizens, or be a CDC, agricultural cooperative, ANC, NHO, or Indian tribe</p> <p>Have its principal office located in a HUB Zone</p> <p>Have at least 35% of employees living in a HUB Zone</p>	Business is small and main location is in a HUBZone, based on ZIP code. To identify main location, we will use physical address. Where physical address is not provided, we will use the mailing address.	Business has HUBZone certification from SBA

Limitations:

There are two primary limitations to the analysis. First, in answering Research Question 3, we note that there is significant under-reporting of subcontracts in FPDS, due, for instance, to the lack of a requirement to report subcontracts under certain value thresholds. Since we have reason to believe that subcontracts to small businesses may be particularly vulnerable to under-reporting, this omission will likely result in the underestimation of small businesses receiving subcontracts. We plan to note this in our findings.

Second, this study is descriptive, and we will be unable to draw conclusions about the causal relationships underlying our findings. We may find that there are ample federal awards in a particular NAICS code, but there are not many small rural entities within this NAICS code that are registered in [SAM.gov](https://sam.gov). This could be because there is a lack of awareness of the potential for federal contracts, suggesting that a future intervention to raise awareness could be effective. However, it could also be because particular NAICS codes are not appropriate for a small business in a rural area. It will be critically important to precisely discuss and present our findings to guard against assuming causal relationships that may not exist.

Preregistration details

This Analysis Plan will be posted on the OES website at oes.gsa.gov before analyzing outcome data.