Counting the Self-Employed Using Household and Business Sample Data*

Richard J. Boden
Alfred R. Nucci

ABSTRACT. This study compares the number and attributes of self-employed workers using the 1982 and 1987 Characteristics of Business Owners (CBO) data and data from Annual Demographic Files (ADF) of the Current Population Survey (CPS) for the same reference years. Both sources of data have been widely used in empirical studies of entrepreneurship/self-employment. Substantial and inexplicable differences were found in the two data series' estimates of the number of self-employed men and women for both reference years. In terms of individual attributes, the CBO and CPS appear to report reasonably similar profiles of self-employed individuals in terms of marital status and geographic location, and similar systematic gender differences in the industrial distributions of these individuals. However, in terms of other attributes captured by both data series, including age, the two series exhibit notable dissimilarities.

Overview

The rise in the number of self-employed individuals over the past two decades has been accompanied by a number of empirical studies of self-employment. One group of these studies has drawn upon data from the Current Population Survey (CPS). These studies include Moore (1983), who attempts to shed light on employer discrimination by examining racial and gender differences in self-employment earnings, and Evans and Leighton (1989), who assess factors influencing self-employment entry and exit. Other CPS-based inquiries into self-employment issues include Devine (1994), who provides a comprehensive, inter-temporal profile of self-employed women, and Boden (forthcoming), who explores gender differences in the propensity for workers to switch from wage employment to self-employment. Another, parallel body of research has been based upon the Characteristics of Business Owners (CBO) data, including Bates (1989), who examines factors influencing small business viability in central cities, and Bates (1990b), who models the relationship between human and financial capital and the longevity of entrepreneurial endeavors.[1] Carrington and Troske (forthcoming) use the CBO data – supplemented with CPS data – to examine gender segregation of workers in small businesses. Finally, an international comparison (that is based, in part, on U.S. business and household data) of changes in self-employment rates over time can be found in Acs, Audretsch, and Evans (1993).

Both the CPS and the CBO represent important resources for cross-sectional analyses of self-employment phenomena (Aronson, 1991), and the aforementioned studies exploit the respective strengths of these two data sources. But do CBO- and CPS-based studies truly reflect different perspectives of a *common* population of self-employed workers? This paper presents the first comparison of the self-employed individuals represented in the CBO data and data from the Annual Demographic File (ADF) of the CPS. We first compare estimates of the number of self-employed individuals in a presumably comparable

Final version accepted on February 8, 1996

*Richard J. Boden**
Department of Marketing
The University of Toledo
2801 West Bancroft Street
Toledo, Ohio 43606-3390
U.S.A.

and

*Alfred R. Nucci**
Center for Economic Studies
U.S. Bureau of the Census
Washington, D.C. 20233-6300
U.S.A.

Small Business Economics **9**: 427–436, 1997.

sub-population between the two data series for the reference years 1982 and 1987. We then compare the percent distributions of these individuals by age, education, marital status, Census region, industry, and the percent of household or personal income derived from self-employment. All of our comparisons are broken out by gender in deference to Devine (1994), who reports notable gender differences in the attributes of self-employed workers as well as a higher growth rate in the number of self-employed women.

From the outset, it should be noted that there are glaring differences in the number of self-employed workers reported by the CBO and ADF data series. For instance, even after developing comparable sub-populations of self-employed workers from the two data series, the 1987 CBO reports 22 percent more workers than the ADF data series for the same reference year. Moreover, the CBO reports a 1982–1987 growth rate of 30 percent for this major sub-population of self-employed individuals, while the corresponding growth rate reported by the ADF is about 7.5 percent. And these discrepancies are even more profound when the data are broken out by gender. Also, while the two data series report reasonably similar educational and geographic distributions of these individuals, there are marked differences for the other attributes measured by both data series.

Two measures of self-employment

The CBO survey collects demographic and economic information on business owners and firm performance from a sample of approximately 130,000 business owners and their roughly 90,000 businesses, respectively (Bates, 1990a; Boden and Nucci, 1990; Nucci, 1992).[2] What makes the survey a unique resource for the study of self-employment – and entrepreneurship, more generally – is the ability to study owner and firm characteristics, together, and to study minorities and women, demographic groups that are oversampled by the survey.

The Current Population Survey is a monthly survey of approximately 60,000 households and roughly 100,000 household members that serves as the source of household-based labor force statistics, such as the unemployment rate, in addition to period information on fertility and migration (U.S. Bureau of the Census, 1978).[3] Every March, the CPS includes a supplementary set of questions which elicit data on the characteristics of individuals' jobs of longest duration and sources of income in the previous year. The resultant data files – from which our reference samples are drawn – are also known as Annual Demographic Files.[4] For any given reference year, some 8,000 ADF respondents identify self-employment as a source of some or all of their income.

As a primer to the next section of the paper, it is useful to note some of the salient differences between these two sources of data, owing to differences in which the data are collected and reported.

Data collection method

The CBO data are collected from mailed-out survey instruments (which elicit data on age, marital status, and other personal and business attributes) for the respective reference years, as well as from various economic census reports (which yield data on gender, race, industry, receipts, payroll, etc.).[5] ADF data are collected through an interview of a household member who provides data on other household members, as applicable.

Classification as self-employed

In the CBO surveys, individuals are self-employed (business owners) by virtue of having filed a business income tax return. Individuals in the ADF surveys are classified as self-employed from responses to source-of-income and class-of-worker questions.

Treatment of item nonresponse

ADF data are imputed for variables with missing data. No data have been imputed for item nonresponse for either CBO dataset.[6]

Reconciling differences between the two data series

The CBO and ADF data are the respective products of two very different surveys. The CBO

self-employed represent a proper subset of the self-employed individuals covered by the ADF. However, it is possible to select subsamples of data on self-employed individuals from each data source, which should, in principle, correspond to a common population of self-employed workers. This section explains how reference samples of data were drawn from the CBO and ADF, respectively, for the purposes of this study.

Racial coverage of reference samples

The scope of the self-employed individuals in our research sample data is restricted to white, non-Hispanic men and women. This decision was motivated by the sparsity of ADF data on self-employed minorities and the substantially higher item and case nonresponse rates for minorities in the CBO. The exclusion of minorities should also reduce unobserved heterogeneity.

Reconciliation of differences in units of analysis

There is a subtle difference in what each record of these respective data sources represents. Each CBO record corresponds to a particular self-employment activity or "job," while each record of the ADF corresponds to a person. For this paper, a person represented in the ADF was classified as self-employed if he or she had any self-employment earnings during the reference year.[7] Such persons – if they filed a business income tax return – would, in principle, be represented in the CBO data files, subject to other CBO coverage issues, such as those described below.

Insofar as some persons file more than one business income tax return in a given year, such persons might be represented by more than one CBO record for a given reference year, and, in fact, such redundancy does exist. However, the 1982 CBO data reveal this to be a rare phenomenon, with less than 3 percent of CBO records pertaining to multiple returns filed by common owners. Also, it is virtually impossible to quantify the incidence of this phenomenon in the 1987 CBO file. Hence, it was decided that each CBO record would be treated as pertaining to a different person. In short, the unit of analysis in this paper is a self-employed person.

Reconciliation of differences in coverage

The CBO covers "small" (S) corporations, but does not cover regular (C) corporations, while the ADF data presumably cover both regular and "small" corporations. Since the ADF does not distinguish between the two types of corporations, data on small corporations were excluded from the CBO files and data on incorporated self-employed persons were excluded from the ADF data for the purposes of these data analyses. In other words, our reference samples are confined to unincorporated self-employed individuals – i.e., individuals whose businesses are organized as sole proprietorships or partnerships.

Also, a number of two-digit Standard Industrial Classification (SIC) industries are not covered by the CBO.[8] This matter was resolved by simply dropping ADF observations on individuals who were self-employed in any of the industries not covered by the CBO. It should be noted that the only ADF data on industry pertain to the job of longest duration: for the purposes of this study, individuals who were not self-employed in their job of longest duration – but accrued self-employment earnings from some other job or jobs – were assigned to the industry of their (wage-sector) job of longest duration.[9]

After making the foregoing adjustments to the ADF and CBO data, conceptually comparable subsamples of self-employed individuals were obtained for each data source, and for each reference year.[10] Table I summarizes the exclusions described above, and the number of workers represented in our reference samples is summarized in the last three rows of this table.

Comparing population estimates between the CBO and ADF data series

The top portion of Table I shows the population estimates of all self-employed persons as reported by the CBO and ADF data series. There is a dramatic difference in these estimates for 1982, with the ADF showing almost 40 percent more self-employed workers than the CBO. Of course, for reasons previously noted, a comparison of the two data series at this level is not particularly meaningful. Instead, we focus upon the estimates of self-employed workers in our presumably com-

TABLE I
Derivation of ADF and CBO reference samples

	1982		1987	
	ADF	CBO	ADF	CBO
All self-employed				
Men	10,483,093	7,714,695	10,774,593	10,473,646
Women	3,809,503	2,782,606	4,697,068	4,282,994
Total	14,292,596	10,497,301	15,471,661	14,756,640
After dropping incorporated self-employed				
Men	8,223,330	7,184,762	8,370,256	9,198,015
Women	3,365,601	2,615,283	4,048,828	3,799,916
Total	11,588,931	9,800,045	12,419,084	12,997,931
After dropping minorities				
Men	7,555,234	6,651,762	7,499,401	8,394,009
Women	3,090,144	2,405,853	3,615,507	3,425,505
Total	10,645,378	9,057,616	11,114,908	11,819,514
After dropping industries not covered by both data series				
Men	6,105,943	6,637,587	6,274,637	8,350,885
Women	2,848,756	2,401,782	3,352,716	3,414,057
Total	8,954,699	9,039,369	9,627,353	11,764,942

Source: Special tabulations of the 1982 and 1987 Characteristics of Business Owners data and 1983 and 1988 Annual Demographic File data.

parable reference samples as summarized in the last three rows of Table I.

For 1982, the two data series each report about 9 million self-employed workers in our reference samples. When we break these numbers out by gender, though, some dissimilarity emerges, with the ADF reflecting a higher proportion of women (32 percent) compared to the CBO (27 percent). For 1987, the number of self-employed individuals in the CBO reference sample is about 22 percent greater than the corresponding ADF figure. Here, too, a disaggregation by gender reveals disparities that are, in fact, even more pronounced: while the two data series each show about 3.4 million self-employed women, the CBO shows 8.4 million men to the ADF's 6.3 million, a striking difference from any perspective.

Looking at these estimates from a different perspective, the CBO reference sample indicates an overall increase in the number of self-employed workers of 30 percent between 1982 and 1987, with growth rates of approximately 25 and 42 percent for men and women, respectively. By contrast, the ADF reference sample shows an overall growth rate in the number of self-employed workers of 7.5 percent for this same time period, with growth rates of 2.8 and 17.7 percent for men and women, respectively. In short, the CBO reference sample indicates substantially greater growth in the number of self-employed workers.

Comparison of selected attributes of individuals in the CBO and ADF reference samples

The CBO and ADF contain data on six common individual attributes: sex, age, education, marital status, percent of income from self-employment, industry, and region. The manner in which these data are collected and are reported differs between the two data series. For instance, data on the percent of personal (household) income that was attributable to a person's self-employment activity are also explicitly collected by and reported as ordered categorical data for the 1982 (1987) CBO. For the ADF, the fraction of personal (household) income was calculated by dividing personal self-employment earnings by total personal (household) income for the 1982 (1987) reference years. For further details, see Boden and Nucci (1995).

Table II shows the percent distributions of individuals in the ADF and CBO reference samples for these common attributes for 1982 and 1987.

TABLE II
Distribution of unincorporated self-employed workers for selected attributes: white, nonhispanic men and women

	1982		1987	
	ADF	CBO	ADF	CBO
Age				
< 25	9.08	3.59	7.19	3.79
25–34	27.17	21.55	27.00	19.61
35–44	25.33	25.63	26.97	27.83
45–54	17.86	21.94	18.79	21.46
55–64	14.13	18.46	13.95	18.06
65+	6.43	8.83	6.11	9.25
Total	100.00	100.00	100.00	100.00
Marital status				
Never married	13.31	8.12	12.96	9.52
Married	75.14	81.56	73.60	78.90
Divorced/separated	8.99	7.52	10.15	8.91
Widow/widower	2.56	2.80	3.29	2.67
Total	100.00	100.00	100.00	100.00
Education				
Grade school	5.22	6.07	4.01	3.21
Some high school	10.69	8.50	9.15	9.18
High school graduate	35.39	32.57	35.42	30.41
Some college	19.86	20.06	21.72	21.37
College graduate	13.62	14.26	14.59	18.13
Graduate school	15.21	18.55	15.11	17.69
Total	100.00	100.00	100.00	100.00
Percent of income derived from self-employment				
None	14.98	13.38	15.80	13.94
1–9	9.41	18.06	17.23	22.81
10–24	8.33	9.74	15.43	11.93
25–49	8.63	7.26	17.71	9.49
50–74	8.56	7.39	13.69	9.84
75–99	21.16	12.99	12.78	12.16
100	28.94	31.17	7.35	19.84
Total	100.00	100.00	100.00	100.00
Industry division				
Agriculture, forestry, fishing and mining	3.30	3.98	3.79	3.70
Construction	14.76	12.19	14.50	13.53
Manufacturing	8.31	2.35	7.14	3.11
Transportation, communications, public utilities	4.42	3.83	4.18	4.29
Wholesale trade	4.27	1.76	3.74	3.14
Retail trade	20.23	22.07	17.08	16.61
Finance, insurance, real estate	5.40	9.54	6.50	9.90
Services	39.32	44.27	43.08	45.72
Total	100.00	100.00	100.00	100.00
Census region				
Northeast	18.05	19.43	18.14	20.20
North Central	27.33	25.10	26.20	25.97
South	31.28	32.18	31.90	31.79
West	23.34	23.29	23.76	22.04
Total	100.00	100.00	100.00	100.00
Number of observations[a]	6,437	32,642	6,220	19,607

Source: Special tabulations of the 1982 and 1987 Characteristics of Business Owners data and 1983 and 1988 Annual Demographic File data.

[a] Percent distributions are calculated from weighted data.

Tables III and IV show the same percent distributions separately for men and women, respectively.

The degree of dissimilarity between the two data series for these individual attributes can be descriptively summarized with Duncan indices. A Duncan index of dissimilarity between two groups of observations – in terms of percent distribution for a given attribute – is calculated by summing the absolute values of differences in cell percents, and dividing that sum by two (2). Limiting values for this index are zero (0) and one hundred (100), with higher values reflecting greater degrees of dissimilarity.

Table V displays Duncan indices, by gender and reference year, for the six individual attributes captured by both data series. The two series reveal a high degree of similarity in the marital status of women in our reference samples, and a fairly high degree of similarity between the two data series in terms of the educational and geographic distribution of both men and women.

The Duncan indices show considerably more agreement between the CBO and ADF reference samples for the industrial distribution of women than for the industrial distribution of men. However, Tables III and IV imply that the two series do capture systematic gender differences in the industrial distribution of the self-employed workers in our reference samples – viz., the higher representation of men in the construction industry, and the higher representation of women in the retail trade and services industries.

The Duncan indices suggest more dissimilarity between the two data series in terms of age distribution than might be initially expected, since data on age should be less contaminated by response error than data for other attributes. As Tables III and IV show, relative to the ADF reference samples, the age distribution of the individuals in the CBO reference samples for both years are clearly skewed toward older age groups. It is quite possible that CBO respondents provided their contemporaneous age instead of their age as of the reference year (approximately three years prior). Also, very young self-employed workers may be less likely to file business income tax returns and may therefore be under-represented in the CBO sampling frame and sample data. Either or both of these two scenarios may account for some of the systematic difference in the age distribution of the individuals in the two data series.

Finally, the Duncan indices show considerable discordance between the two data series in terms of the percent of personal or household income that individuals purportedly obtained from self-employment during the 1982 or 1987 reference years, respectively. This is not at all surprising given the different manner in which the CBO and ADF surveys collect this information.

Conclusions

Our CBO and ADF reference samples yield weighted estimates of the number of self-employed workers that diverge between 1982 and 1987. While both samples show roughly 9 million self-employed in 1982, the ADF and CBO weighted estimates for 1987 indicate totals of 9.6 million and 11.8 million self-employed individuals, respectively. A break-out of these estimates by gender reveals the source of this divergence. Our CBO reference samples for 1982 and 1987 indicate a growth rate in the number of self-employed men of 26 percent compared to just 3 percent for our ADF reference samples for the same years. The difference in the growth rate in the number of self-employed women in our reference samples is much less pronounced.[11]

Inter-temporal changes in the weighted number of self-employed individuals in the CBO data reflect, in large part, changes in the size of the IRS business income tax universe and the proportion of that universe that is covered by the CBO survey. Changes in weighted estimates of self-employed individuals as reported by the ADF reflect changes in the population of individuals who are candidates for the labor force, as well as changes in the composition – in terms of type of income recipiency – of those individuals as captured by the ADF survey.

In the final analysis, the 1982 and 1987 CBO data series provide more authoritative estimates of the total unincorporated self-employed population for these two years. It follows, then, that the ADF may understate the growth in the number of self-employed individuals, particularly self-employed men. Researchers who use ADF data to study rates of growth in the number of self-employed individuals or self-employment entry rates should be

TABLE III
Distribution of unincorporated self-employed workers for selected attributes: white, nonhispanic men

	1982		1987	
	ADF	CBO	ADF	CBO
Age				
< 25	9.42	3.42	7.21	4.21
25–34	26.18	20.89	26.53	19.26
35–44	24.44	25.07	25.62	27.50
45–54	18.23	22.66	18.90	21.26
55–64	14.91	18.91	14.94	18.37
65 +	6.82	9.05	6.81	9.41
Total	100.00	100.00	100.00	100.00
Marital status				
Never married	15.25	7.83	14.98	9.70
Married	75.24	84.72	73.98	81.77
Divorced/separated	8.41	6.16	9.42	7.23
Widow/widower	1.10	1.30	1.62	1.30
Total	100.00	100.00	100.00	100.00
Education				
Grade school	6.29	6.95	4.6	43.66
Some high school	11.16	8.89	9.32	10.36
High school graduate	33.11	31.21	33.64	29.14
Some college	19.12	18.69	21.07	20.94
College graduate	13.56	14.41	14.91	18.02
Graduate school	16.76	19.85	16.43	17.88
Total	100.00	100.00	100.00	100.00
Percent of income derived from self-employment				
None	13.54	12.13	13.24	13.24
1–9	9.05	16.53	11.62	19.08
10–24	8.11	9.21	12.67	10.17
25–49	8.03	7.00	18.03	9.54
50–74	8.34	7.76	17.20	10.95
75–99	21.49	14.73	17.42	13.81
100	31.44	32.66	9.82	23.21
Total	100.00	100.00	100.00	100.00
Industry division				
Agriculture, forestry, fishing and mining	3.31	4.66	4.60	4.44
Construction	20.84	15.72	21.31	18.24
Manufacturing	9.77	2.51	8.64	3.51
Transportation, communications, public utilities	5.94	4.68	5.56	5.30
Wholesale trade	5.33	2.03	4.92	3.69
Retail trade	15.69	20.60	13.74	14.91
Finance, insurance, real estate	5.06	8.83	6.39	9.21
Services	33.06	40.96	34.83	40.69
Total	100.00	100.00	100.00	100.00
Census region				
Northeast	19.19	19.78	19.22	20.09
North Central	26.72	25.26	24.97	26.15
South	31.27	32.18	32.42	32.10
West	22.82	22.77	23.39	21.65
Total	100.00	100.00	100.00	100.00
Number of observations[a]	4,338	17,881	4,019	11,394

Source: Special tabulations of the 1982 and 1987 Characteristics of Business Owners data and 1983 and 1988 Annual Demographic File data.

[a] Percent distributions are calculated from weighted data.

TABLE IV
Distribution of unincorporated self-employed workers for selected attributes: white, nonhispanic women

	1982		1987	
	ADF	CBO	ADF	CBO
Age				
< 25	8.36	4.08	7.15	2.75
25–34	29.30	23.37	27.88	20.47
35–44	27.25	27.17	29.51	28.65
45–54	17.05	19.95	18.59	21.95
55–64	12.45	17.21	12.09	17.32
65 +	5.59	8.22	4.78	8.86
Total	100.00	100.00	100.00	100.00
Marital status				
Never married	9.13	8.93	9.19	9.09
Married	74.94	72.83	72.89	71.88
Divorced/separated	10.24	11.28	11.51	13.03
Widow/widower	5.70	6.95	6.41	6.00
Total	100.00	100.00	100.00	100.00
Education				
Grade school	2.94	3.61	2.85	2.13
Some high school	9.69	7.42	8.83	6.30
High school graduate	40.28	36.30	38.74	33.52
Some college	21.46	23.84	22.94	22.41
College graduate	13.74	13.85	14.01	18.41
Graduate school	11.88	14.97	12.63	17.23
Total	100.00	100.00	100.00	100.00
Percent of income derived from self-employment				
None	18.06	16.93	20.61	15.67
1–9	10.20	22.39	27.72	31.94
10–24	8.80	11.25	20.61	16.22
25–49	9.90	7.99	17.10	9.34
50–74	9.02	6.36	7.12	7.14
75–99	20.45	8.10	4.10	8.10
100	23.57	26.97	2.74	11.59
Total	100.00	100.00	100.00	100.00
Industry division				
Agriculture, forestry, fishing and mining	1.14	2.10	2.26	1.89
Construction	1.74	2.42	1.75	2.02
Manufacturing	5.16	1.92	4.34	2.12
Transportation, communications, public utilities	1.15	1.46	1.60	1.83
Wholesale trade	2.00	1.02	1.51	1.79
Retail trade	29.98	26.13	23.31	20.76
Finance, insurance, real estate	6.12	11.53	6.69	11.58
Services	52.72	53.42	58.53	58.01
Total	100.00	100.00	100.00	100.00
Census region				
Northeast	15.61	18.47	16.12	20.47
North Central	28.65	24.64	28.49	25.52
South	31.31	32.17	30.94	31.01
West	>24.44	24.72	24.45	23.00
Total	100.00	100.00	100.00	100.00
Number of observations[a]	2,099	14,761	2,201	8,213

Source: Special tabulations of the 1982 and 1987 Characteristics of Business Owners data and 1983 and 1988 Annual Demographic File data.

[a] Percent distributions are calculated from weighted data.

TABLE V
Duncan indices of dissimilarity between the ADF and CBO data series for selected attributes of self-employed workers

	Total		Men		Women	
	1982	1987	1982	1987	1982	1987
Age	11.11	10.79	11.29	10.27	10.29	12.67
Marital status	6.66	5.30	9.68	7.79	2.30	1.52
Education	5.02	6.16	4.60	5.61	6.25	9.00
Percent of income derived from self-employment	12.30	18.06	9.79	20.85	18.05	17.09
Industry	11.62	6.16	16.94	9.85	8.07	5.67
Census region	2.28	2.06	1.81	2.06	4.01	4.42

Source: Special tabulations of the 1982 and 1987 Characteristics of Business Owners Data and 1983 and 1987 Annual Demographic File data.

aware of potential biases resulting from this undercount.

Despite the discrepancies outlined in this paper, both data series have distinguishing strengths and limitations, and neither is likely to supersede the other as a source of data on self-employment. For instance, the ADF contains data on wage workers as well as individuals who are unemployed or not in the labor force (unlike the CBO). This means that the ADF data can be used to assess the determinants of selection into self-employment. Also, the ADF is an annual data series that generally becomes available a little more than a year after each reference year. The CBO data, by contrast, are only available for economic census years – i.e., every five years – and are generally not available to researchers until three years after each reference year. Also, public use files for the ADF are widely available, while there are no public use files for the CBO.

Likewise, the CBO data can accommodate empirical studies for which the ADF data are not well suited. Unlike the ADF, the CBO contains much more information on the characteristics of the businesses owned by self-employed individuals, including receipts, number of paid employees, and start-up capital. Further, the very large number of observations in the CBO – especially on women and minority business owners – means that the CBO can be used to study empirical issues in more depth than can be accommodated by the ADF.

Hopefully, our findings will attract the attention of the government agencies responsible for collecting and disseminating data on the self-employed. These agencies should produce consistent estimates of the number and demographic composition of these individuals.

Notes

* The views expressed in this paper are those of the authors and not necessarily those of the U.S. Bureau of the Census or any other government agency.

[1] Compared to the CPS, fewer studies have been based upon the CBO data because the CBO data are less accessible to researchers. The CBO data are confidential, and no public-use files for these data exist: these data must be accessed by authorized personnel on-site at the Census Bureau in Suitland, Maryland. Public-use CPS data files, on the other hand, are readily available from the Census Bureau and university data consortia.

[2] These are firms registered with the IRS as sole proprietorships, partnerships, or small (sub-chapter S) corporations. Individual owners – for up to ten owners for businesses with more than one owner – of a sampled business are included in the CBO sample. Individuals filing more than one business tax form are counted as separate businesses when their industrial classification is different and are treated as the same businesses when it is the same. Thus, self-employed individuals who are multiple jobholders are included in the CBO estimates.

[3] There are other contemporaneous sources of data on self-employed workers – most notably, the Census Bureau's Survey of Income and Program Participation (SIPP), the National Longitudinal Surveys, and the Panel Survey of Income Dynamics. These panel data series do not contain nearly as many observations, in cross-section, on self-employed workers as the ADF. For an overview of the SIPP literature, see Haber et al. (1987).

[4] The original 1988 ADF was used in this research, as opposed to the 1988 "re-write" file.

[5] Questionnaires are mailed to business owners during the Bureau of the Census quinquennial economic censuses program approximately three years after the reference year (a mail year of 1990 for the reference year 1987, for example).

[6] As with any survey, both the ADF and CBO do not have response data for every survey item. The CBO does not have imputed values for item non-response, although the ADF does. Unfortunately, the ADF files used in this study did not contain flags for imputed variables, other than income-related variables. Such flags became available beginning with the release of the March 1988 CPS "re-write" file. For tabulation purposes, CBO item-nonresponse observations were distributed according to the distribution of observations for which response data are available, thereby insuring consistent totals for the CBO data across tabulations for different response-based variables.

[7] In CPS/ADF parlance, this means that the ADF records used in these data analyses pertain to workers who were: (a) self-employed in their reference-year job of longest duration, and accrued no other reference-year self-employment earnings, (b) self-employed in their reference-year job of longest duration, and accrued self-employment earnings from some other reference-year job or jobs, or (c) were not self-employed in their reference-year job of longest duration, but accrued self-employment earnings from some other reference-year job or jobs.

[8] These industries are: agricultural production (SICs 01 and 02), railroad transportation (SIC 40), U.S. Postal Service (SIC 43), membership organizations (SIC 86), private household services (SIC 88), public administration (SICs 91 through 97). Also, the CBO files do not cover the trusts (SIC 673) and miscellaneous investing (SIC 679) industries. However, since the ADF data cannot expressly identify these three-digit industries, observations pertaining to SICs 62 and 67 were excluded from both the CBO and ADF files for these data analyses.

[9] Approximately 20 percent of the ADF observations in this ADF reference sample pertain to individuals who were not self-employed in their jobs of longest duration in the respective reference years. Also, a very small number of ADF records on individuals who were not in the labor force or otherwise not in the survey universe in the reference year were excluded from these data analyses.

[10] Unfortunately, there are some differences between the coverage of the ADF and CBO files that could not be reconciled. First, the 1987 CBO data do not cover owners of businesses with annual receipts of less than $500, while the 1982 data do. To attain comparability between the 1982 and 1987 CBO files, 1982 CBO data on owners of businesses with annual receipts of less than $500 were excluded from these data analyses. The ADF contains no data on annual receipts, so self-employed individuals whose businesses generated less than $500 in annual receipts could not be excluded from these data analyses. In principle, then, the ADF data cover more people than the CBO data used in these data analyses. Second, the 1982 CBO data only cover businesses with fewer than 10 common owners, while the 1987 CBO data has no such limitation. However, the impact of this latter asymmetry in coverage upon these data analyses is somewhat moderated by our exclusion of data on incorporated self-employed workers/business owners.

[11] It is worth noting that the (unweighted) number of observations on self-employed women in the ADF increase by about 10 percent between the 1982 and 1987 reference years, while the number of observations on self-employed men actually declined by about 7 percent. This is obviously a major reason for the gender difference in the growth rates of the weighted number of self-employed individuals reported by the ADF.

References

Acs, Zoltan, David Audretsch and David Evans, 1994, 'The Determinants of Variations in Self-Employment Rates Across Countries and Over Time', Discussion Paper No. 92-3, Wissenschaftszentrum Berlin für Sozialforschung.

Aronson, Robert L., 1991, *Self-Employment, A Labor Market Perspective*, Ithaca, New York: IRS Press.

Bates, Timothy, 1989, 'Small Business Viability in the Urban Ghetto Milieu', *Journal of Regional Science* **29**, 625–643.

Bates, Timothy, 1990a, 'The Characteristics of Business Owners Database', *Journal of Human Resources* **23**, 752–756.

Bates, Timothy, 1990b, 'Entrepreneurial Human Capital Inputs and Small Business Longevity', *Review of Economics and Statistics* **72**, 551–559.

Boden, Richard, (forthcoming), 'Gender and Self-Employment Selection: An Empirical Assessment', *Journal of Socio-Economics*.

Boden, Richard and Alfred Nucci, 1990, 'The 1982 Characteristics of Business Owners Dataset', *Small Business Economics*.

Boden, Richard and Alfred Nucci, 1995, 'Counting the Self-Employed from Two Perspectives: Household vs. Business Sample Data', U.S. Bureau of the Census, Center for Economic Studies, Discussion Paper No. 95-11, Washington, D.C.

Borjas, George and Stephen Bronars, 1988, 'Consumer Discrimination and Self-Employment', National Bureau of Economic Research Working Paper No. 2627.

Carrington, William and Kenneth Troske, forthcoming, 'Gender Segregation in Small Firms', *Journal of Human Resources*.

Devine, Theresa, 1994, 'Characteristics of self-employed women in the United States', *Monthly Labor Review* **117**, 20–34.

Evans, David and Linda Leighton, 1989, 'Some Empirical Aspects of Entrepreneurship', *American Economic Review* **79**, 519–535.

Haber, Sheldon, Enrique Lamas and Jules Lichtenstein, 1987, 'On Their Own: the Self-Employed and Others in Private Business', *Monthly Labor Review* **110**, 17–23.

Holmes, Thomas and James Schmitz, 1992, 'On the Turnover of Business Firms and Business Managers', U.S. Bureau of the Census, Center for Economic Studies, Discussion Paper No. 92-6, Washington, D.C.

Moore, Robert, 1983, 'Employer Discrimination: Evidence from Self-Employed Workers', *Review of Economics and Statistics* **65**, 496–501.

Nucci, Alfred, 1992, 'The Characteristics of Business Owners Database', U.S. Bureau of the Census, Center for Economic Studies, Discussion Paper No. 92-7, Washington, D.C.

U.S. Bureau of the Census, 1978, 'The Current Population Survey: Design and Methodology', Technical Paper No. 40, Washington, D.C.: Government Printing Office.