Comparing Survey Data and Tax Data: Differences in Reporting across Businesses

Tami Gurley-Calvez
Assistant Professor
Department of Health Policy and Management
University of Kansas Medical Center
Mail Stop 3044, 3901 Rainbow Blvd.
Kansas City, KS 66204
Phone: (913) 588-0869
Fax: (913) 588-8236
(tgurley-calvez@kumc.edu)

Donald Bruce
Center for Business and Economic Research
and Department of Economics
Haslam College of Business
The University of Tennessee
916 Volunteer Blvd, 722 Stokely Management Center
Knoxville, TN 37996
Phone: (865) 974-6088
(dbruce@utk.edu)

E.J. Reedy
Kauffman Foundation
Director in Research and Policy
4801 Rockhill Road
Kansas City, MO 64110
ereedy@kauffman.org

Josh Russell
Kauffman Foundation
Senior Research Assistant
4801 Rockhill Road
Kansas City, MO 64110
ereedy@kauffman.org

April 2016

Comparing Survey Data and Taxpayer Data: Differences in Reporting across Businesses

Abstract

Survey data and tax data are two common sources of information for research on firm size and growth. Despite several recent studies of self-employment income reported at the household level, little is known about how information reported in tax data is related to survey data across all firm types. This analysis makes use of a matched administrative and survey data file that includes sole proprietors, LLCs, partnerships, S Corporations, and C Corporations to assess differences in reported business receipts, expenses, and profits across data sources. Our findings suggest that firm owners report higher receipts in survey data as well as higher expenses. The difference in expense reporting is large enough that reported profits are higher in the tax data. These differences vary by owner gender and age and the employer status of the firm.

JEL: H2, L2, D22, C8

Keywords: tax data, survey data, entrepreneurship, business receipts, profit, small business

Introduction

The literature on small business and entrepreneurship entry, survival, and growth is interesting to a broad audience of tax administrators, researchers, advocates and policy makers. However, little attention has been focused on the possibility that some of the disparate answers in the literature are driven by differences in data sources. Survey data and tax return data are two common sources of information for entrepreneurial research, but it is unclear whether empirical results based on survey data are useful for tax administration purposes. Business owners might respond to surveys with optimistic estimates of their business performance but take action to limit their reporting requirements to avoid taxation leading to conservative income reports in tax data. Biases in either direction in terms of receipts and expense reporting would limit the usefulness of survey-based elasticity estimates for predicting tax revenue response to policy changes.

Thus, an open question in the literature is whether tax data and survey data include similar information and produce equivalent research results. Evidence from the literature suggests underreporting of income on tax returns and in self-reported survey information (Black et al., 2012; Hurst, Li, and Pugsley, 2014; Alm and Erard, 2015). Estimates from tax data are based on audits of self-employed households (Black et al., 2012) and survey estimates are based on comparisons of income and consumption patterns between the self-employed and wage and salary workers (Hurst, Li, and Pugsley, 2014). We use a matched dataset to directly compare information across data sources. Survey data from the Kauffman Firm Survey (KFS), an eight-year panel of new firms beginning in 2004, are matched to tax data from Schedule C sole proprietors (Form 1040), partnerships (Form 1065), Subchapter S-Corporations (Form 1120S) and C-Corporations (Form 1120). We assess differences in reported business receipts, expenses, profits, and employer status by organizational form and employer status, factors available in tax data, as well as owner and firm characteristics available in the KFS. A strength of the matched dataset is that it covers firms regardless of organizational form (e.g. sole proprietorship, LLC, partnership) and provides evidence beyond the self-employment measure used in much of the previous literature.

Our results suggest that survey data and tax return data contain different information. We find a nuanced relationship between tax data and survey information, which provides insights into the usefulness of reported survey data for informing tax administration. In the matched data file, reported receipts were higher in the survey data and reported expenses were higher in the survey data to such a large degree that despite lower reported receipts, profits were higher in the tax data. These results might indicate that respondents report expenses not allowable for tax purposes in survey responses.

Background

Since Evans and Leighton (1989) and Hamilton (2000) raised the topic of entrepreneurial returns in the economics literature, much work has been done to understand how financial returns to entrepreneurship relate to those of wage and salaried workers. This literature has largely relied upon survey-based data such as the Panel Study on Income Dynamics or the National Longitudinal Survey of Youth and the results are puzzling in that self-employment persists even in light of lower earnings compared to the wage and salary sector (Hamilton, 2000; Åstebro and Chen, 2014). Researchers have posited that some of the puzzle can be explained by nonpecuniary benefits and differences in preferences, risk attitudes and beliefs (e.g. Camerer and Lovallo, 1999; Moskowitz and Vissing-Jørgensen, 2002; Puri and Robinson, 2013).

Another possible explanation is that self-employment income is underreported in survey data and a number of recent studies have assessed the degree of underreporting in surveys. Using consumption data for wage and salary workers, Hurst, Li, and Pugsley (2014) estimate that the self-employed underreport their income by about 25 percent or about half of the estimated underreporting (56 percent) found in audits of tax return data (Black et al, 2012). Sarada (2015) finds that self-employed workers consume and save more than their wage and salary counterparts despite reporting lower income, suggesting that there might be measurement error in reported earnings. Additionally, recent evidence suggests that business survival estimated from KFS data yields different results than estimates from matched administrative tax data (Gurley-Calvez et al., 2015).

We build on these efforts by directly comparing survey and tax data for a matched panel of firms. The KFS is a longitudinal survey of nearly 5,000 new firms that began operations in 2004 and were followed over an eight year period. KFS survey data are matched to tax data allowing for comparisons of reported receipts, expenses, profits, and employer status by firm and year. We

assess whether reporting gaps vary by owner characteristics and organizational form in the spirit of Levine and Rubinstein (2013) and a growing literature that indicates differences in firm behavior by owner demographics (Fairlie and Meyer, 1996; Fairlie and Robb, 2009; Gurley-Calvez, Biehl, and Harper, 2009; Hundley, 2001; Lombard, 2001).

Data

A direct firm-level comparison of reported information in KFS and tax data is possible with the matched KFS-tax dataset. A strength of the KFS-tax matched data file is the inclusion of businesses regardless of their legal (e.g. sole proprietorship, Subchapter S-Corporation, partnership) or tax reporting (e.g. Schedule C, 1120 Subchapter S-Corporation form, 1065 Partnership return) status. Most economics research on business formation and outcomes has been based on the self-employment identified through survey responses (Bates, 1990; Bruce, 2002; Headd, 2003; Bates, 2005; Bruce and Mohsin, 2006; Fairlie and Robb, 2007; Fairlie and Robb, 2009; Robb and Watson, 2012) or filing a Schedule C (Holtz-Eakin, Joulfaian, and Rosen, 1994; Carroll et al., 2001; Gurley-Calvez and Bruce, 2008). In the matched data file about one third of new businesses report being sole proprietorships or file a Schedule C through the individual tax system.

Kauffman Firm Survey

The KFS is a recent longitudinal survey of nearly 5,000 new firms that began operations in 2004.[1] The initial survey was administered in 2005 and 2006, and seven follow-up surveys were administered to collect annual data through2011. Response rates exceeded 80 percent at each of the follow-ups and concerted efforts were made to determine whether non-response was a result of business closure or for some other reason. As a result, the KFS provides researchers with a unique opportunity to study a panel of new businesses from start-up to sustainability or closure.

[1] http://www.kauffman.org/what-we-do/research/kauffman-firm-survey-series

Tax Data

Tax return data are pulled from the Business Returns Transaction File (BRTF) and the Individual Returns Transaction File (IRTF) for tax years 2004-2013. These files contain the population of business and individual filings. One potential concern for linking the KFS with tax data is the possibility that firms might change legal status and file different tax forms. These changes might be particularly difficult to track in the tax data if filers change tax identification numbers (TINs). For this project, firms were linked across datasets using name and address for tax years 2004 through 2008 to capture the business regardless of TIN. Firm data are matched by firm name for forms 1065, 1120, and 1120S and owner name for form 1040, Schedule C and address for each year. Table 1 includes the results from iterative data searches for KFS firms in the tax data. Almost 90 percent of matches to the BRTF (1065, 1120, 1120S) are exact matches or matches with similar addresses (e.g. street names spelled differently). About 80 percent of matches to the IRTF (1040, Schedule C) are exact matches or similar addresses. A higher percent of IRTF matches (17 percent versus 7 percent) have non-matching addresses, but were matched on owner name. This is perhaps not surprising as the tax form is filed under the individual's address, which is not necessarily the same as the business address. Note that the original matching resulted in more matches than the number of firms in the KFS (4,928). This is because BRTF matches indicate actual firm-level matches whereas IRTF matches indicate that a tax return was located for the business owner, but further validation is needed to ascertain whether the return includes a Schedule C matching the KFS business profile.[2]

Table 1: Matching Results from Multiple Passes through Individual and Business Tax Files

	IRTF		BRTF	
Match Notes	Number	Percent	Number	Percent
Exact Match	1,836	63	1,409	62

[2] For each firm-year combination without a BRTF match, any IRTF matches are assessed for presence of a Schedule C. Filers with multiple Schedule C's are reviewed by year to locate the KFS firm.

Addresses Similar	444	15	604	27
Address Does Not Match	512	17	165	7
Names Similar	99	3	83	4
Other	44	1	17	1
Total	2,935	100	2,278	100

Matched File

Final matching statistics by KFS reported legal structure and tax form are presented in Table 2. Overall, 80 percent of firms are matched to tax files. The matched data file includes 3,940 firms and 22,444 firm-year observations. Match rates are highest for Subchapter S-corporations (88 percent) and lowest for sole proprietorships (71 percent). For our analysis, we limit our sample to the firms that matched a Schedule C, 1065, 1120, or 1120S in at least one year of tax return data.

Table 2: KFS-Tax Return Match Rates by Legal Status and Form[3]

KFS Legal	Match Rate	Sched. C	F1065	F1120	1120S	Other	N
1 - Sole Proprietorship	0.71	0.607	0.013	0.017	0.037	0.036	1,635
2 – LLC	0.842	0.331	0.306	0.028	0.071	0.106	1,557
3 - Subchapter S	0.88	0.122	0.014	0.117	0.598	0.028	1,040
4 - C-Corporation	0.816	0.129	0.018	0.465	0.125	0.079	441
5 - Partnership and Other	0.757	0.311	0.487	0.036	0.083	0.083	255
Overall	0.800	0.355	0.125	0.082	0.175	0.062	4,928

[3] These statistics include about 240 cases from the BRTF where we have located a match in the entity-level tax information file but not to one of the three main forms. We conducted more data searches for these filers and determined that most were in the tax information file because they submitted payroll tax withholding reports. However, these filers did not file income tax returns. In some cases, it does not appear that an income tax return was required.

Reporting Differences

Comparisons across the datasets indicate a substantial amount of variance in the reporting of receipts, expenses, and profits. Reported receipts are higher in the survey data than in the tax data (Table 3). The difference in reported receipts (tax receipts – KFS receipts) indicates that the median firm reported $800 more in the KFS survey data than the tax data. The mean is substantially higher at just under $500,000. There are several reasons why these numbers might differ. Owners might be more likely to overestimate receipts when responding to the survey than when documenting receipts for tax purposes. It is also possible that KFS respondents report aggregate receipts for more than one closely-related business, but that these earnings are reported on separate tax forms. Whatever the reason behind these findings, the differences are substantial and suggest that research from the separate data sources might not yield qualitatively similar conclusions.

Reported expenses are also higher in the KFS data. The median difference in expenses is -$3,300 indicating that the median firm reported about $3,000 more in expenses in the KFS. Again, the mean difference in expenses was higher. This result is perhaps less intuitive as firms would have an incentive to report all eligible expenses to reduce tax payments. However, it seems plausible that owners might report the full amount of expenses in the KFS, but major expenses would be depreciated over several years in the tax data. Another reason for the observed differences could be that different individuals are responsible for tax and survey reporting and different methods might be used to calculate each. Additionally, although the survey was carefully designed to capture randomly selected new firms, respondents might have included reported survey information on closely related businesses or might have reported survey information at a different level than the tax data (e.g. parent company or subsidiary).

Interestingly, the expense reporting in the KFS was high enough relative to tax data that although receipts were lower in the tax data, overall profits were higher. The median difference in

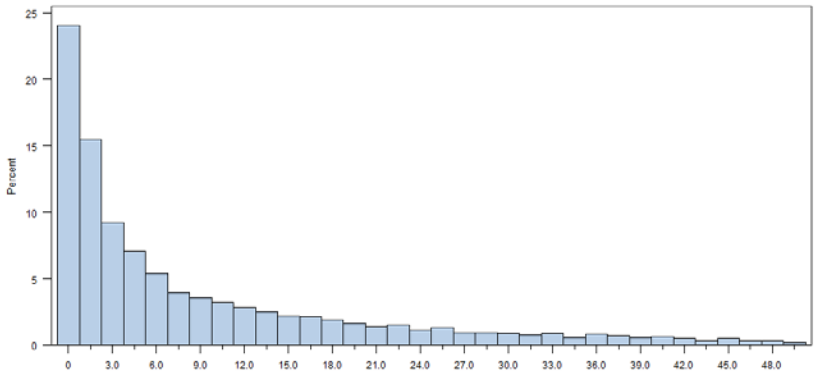
profits was $12,200 and the mean difference was just over $200,000. These results are consistent with a large literature indicating underreported income in tax data (e.g. Andreoni et al., 2008; Slemrod, 2007; Feldman and Slemrod, 2007) and recent evidence suggesting that underreporting is greater for households with larger shares of business income (Kukk and Staehr, 2015). However, the evidence suggests that expenses are systematically over-reported in the survey data and net profit is lower in the survey data leading to opposite conclusions about the relationship between tax data and measures of business profit.

Distributions of business receipts, expenses, and profits are explored in more detail using histograms (Figure 1 – Figure 3). Receipts reported in Figure 1 indicate that zero KFS reported values are less common (about 16 percent versus 24 percent in the tax data) and the KFS distribution has more bulk in the right tail. As indicated by the summary statistics presented in Table 3, differences in receipts are centered near zero, but there are higher frequencies in the left tail, indicating lower reported tax receipts. Expenses follow a similar pattern (Figure 2). KFS firms were more likely to have zero or negative net profit (Figure 3) .

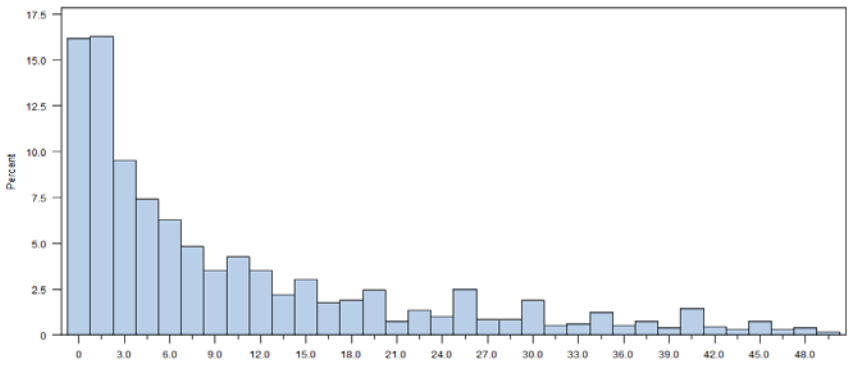
Table 3: Receipts, Expenses and Profits across Survey and Tax Datasets

	Receipts ($10,000)			Expenses ($10,000)			Profit ($10,000)		
	Tax Data	KFS Data	Difference	Tax Data	KFS Data	Difference	Tax Data	KFS Data	Difference
Mean	41.74	55.22	-47.83	18.84	36.90	-37.25	16.96	2.97	20.71
Median	6.63	9.20	-0.08	3.62	5.70	-0.33	2.41	0.50	1.22
99th Percentile	743.42	1150.00	84.70	268.00	745.00	45.23	247.80	81.00	251.11
75th Percentile	28.10	35.00	0.23	15.20	23.60	0.45	14.21	3.10	12.26
25th Percentile	1.18	2.05	-2.85	0.08	1.20	-5.69	0.05	-0.30	-0.11
1st Percentile	0.00	0.03	-380.35	0.00	0.00	-549.14	-4.20	-44.00	-22.55
N	11328	11328	11328	11293	11328	11293	11328	11034	11034

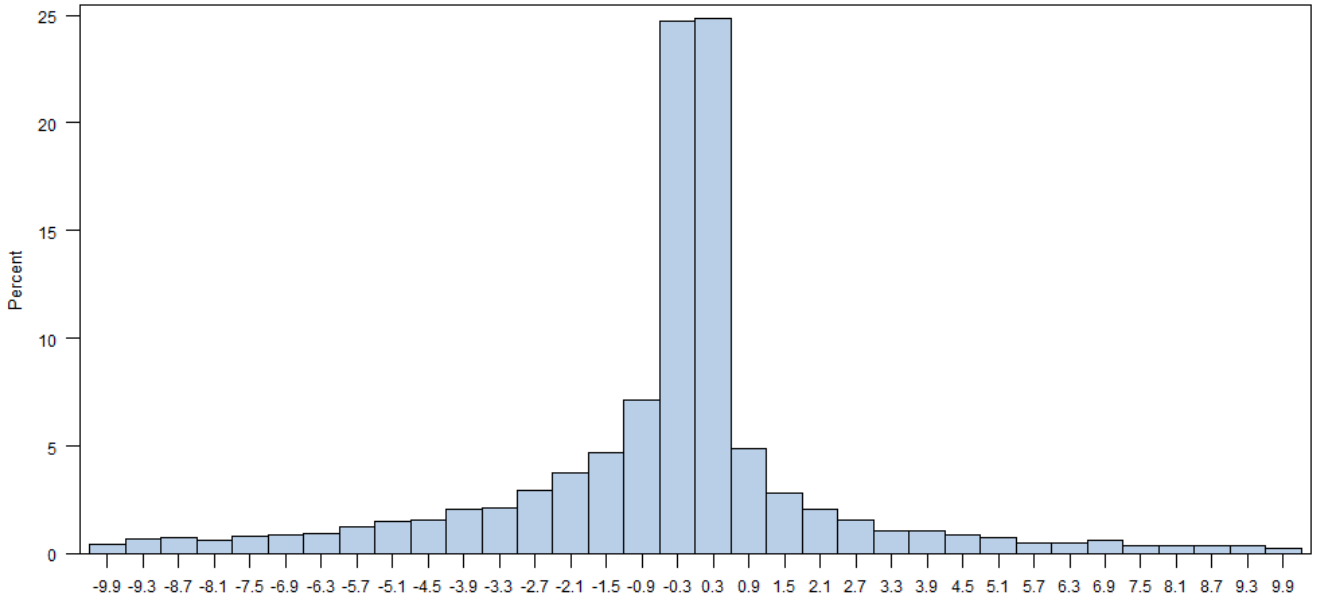
Figure 1: Distribution of Business Receipts by Source[4]



Tax Business Receipts ($10,000)



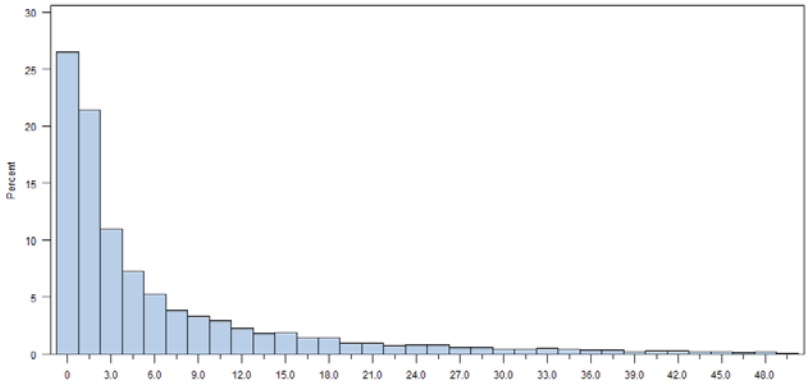
KFS Receipts ($10,000)



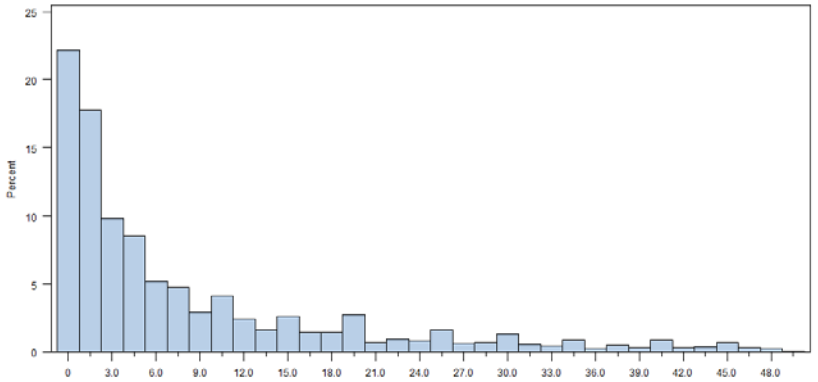
Difference in Receipts: Tax - KFS ($10,000)

[4] For ease of presentation, the distribution is presented for up to 500,000 in receipts and differences of up to 100,000.

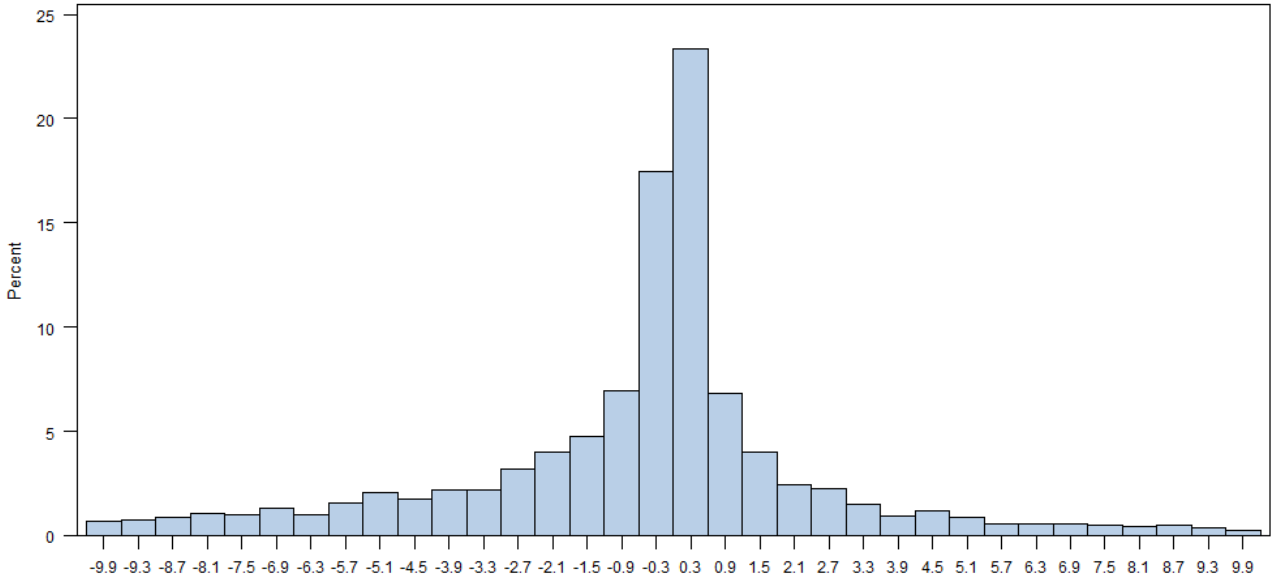
Figure 2: Distribution of Business Expenses by Source[5]



Tax Business Expenses ($10,000)



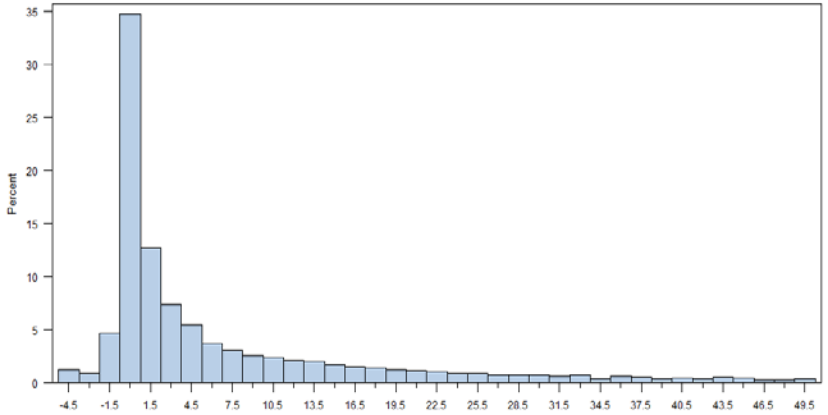
KFS Expenses ($10,000)



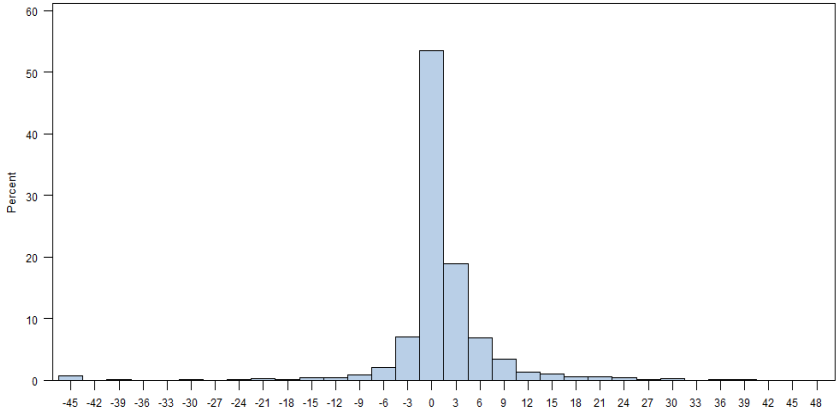
Difference in Expenses: Tax - KFS ($10,000)

[5] For ease of presentation, the distribution is presented for up to 500,000 in expenses and differences of up to 100,000.

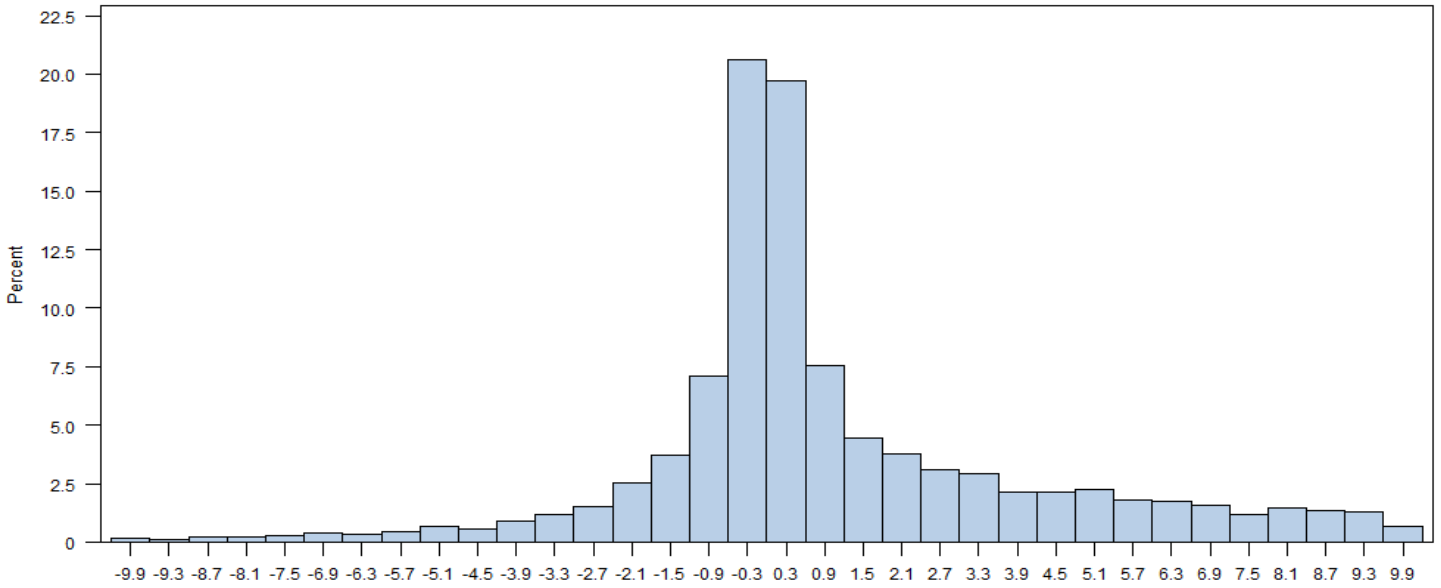
Figure 3: Distribution of Business Profits by Source[6]



Tax Business Profits ($10,000)



KFS Profits ($10,000)



Difference in Profits: Tax - KFS ($10,000)

[6] For ease of presentation, the distribution is presented for up to 500,000 in profits and differences of up to 100,000.

We use the rich owner and firm characteristic data available in the KFS as well as KFS and tax information on firm legal structure to assess which factors are correlated with reporting differences in a multivariate regression framework. We estimate linear regression models with standard errors clustered at the firm level, year fixed effects, characteristics for the primary owner (gender, age, education, race, ethnicity, citizenship status), and firm characteristics (presence of a patent or trademark, percent of sales made directly to individuals, number of owners). Table 4 includes select regression results for reporting differences in receipts, expenses and profits.

Owner gender and employer status of the firm (firm reports salary payments>0 in the tax data[7]) are generally related to reporting differences. Firms with female owners have smaller differences in receipts (Figure 1), expenses (Figure 2), and profits (Figure 3) while the reverse is true for expenses and profits among firms that report salary expenses. Relative to the youngest owners (under 30), firms with owners aged 55-65 have larger differences in reported receipts and expenses. Firms with primary owners identifying as Hispanic have smaller differences in receipts and expenses, while firms with owners who report being US citizens report larger expenses in the KFS and higher profits in the tax data. Firms filing forms 1120 or 1120S have higher reported profits in the tax data relative to those that file a Schedule C. Relative to firms with the highest credit scores, firms with lower scores were more likely to report higher expenses and lower profits in the tax data. The reporting differences were largest for between firms with the highest and lowest credit scores.

[7] These payments do not include sole-proprietorship earnings accruing to the owner.

Table 4: Regression Results for Reporting Differences

	(1) Difference in Receipts	(2) Difference in Expenses	(3) Difference in Profits
Any Salary	12.1045	-25.3340***	21.6857**
	(31.6062)	(9.1456)	(9.7684)
Female	48.8091***	20.0388***	-7.1319*
	(16.7959)	(5.6309)	(3.9652)
Owner Age 30-40	1.6957	1.8915	3.4980
	(12.7469)	(8.8805)	(3.6556)
Owner Age 40-55	-34.8366	-7.2465	4.8349
	(23.0257)	(7.1685)	(4.5516)
Owner Age 55-65	-60.1075*	-22.8482**	-2.4248
	(31.8212)	(9.4261)	(3.4943)
Owner Age 65+	-29.2681	11.1600	-4.3709
	(34.4123)	(7.4789)	(5.3019)
Hispanic	32.5625**	19.3731**	-2.4281
	(15.0250)	(7.6169)	(4.5622)
Citizen	34.6194	-18.9767**	11.2401**
	(67.7842)	(7.780)	(4.5381)
Patent or Trademark	-27.4213	-2.8604	3.1038
	(41.4210)	(10.8381)	(5.2506)
High Credit Score	182.5091	96.6061*	-8.3393*
	(140.8407)	(56.9169)	(5.0240)
Mid Credit Score	194.0293	101.3100*	-5.6755
	(140.1289)	(55.8513)	(8.6974)
Low Credit core	212.0981	99.9981*	-6.4716
	(136.0089)	(54.4311)	(5.9436)
Lowest Credit Score	199.6804	104.0392*	-10.4297*
	(154.2711)	(56.7440)	(6.1301)
Form 1065	23.9158	7.6525	2.6036
	(25.2874)	(13.6393)	(5.9727)
Form 1120S	-145.8538	-40.7154	18.6184***
	(94.8149)	(36.3187)	(5.4582)
Form 1120	-16.8684	6.9062	19.4910*
	(26.8614)	(9.0093)	(9.6549)
N	10,255	11,808	11,844
Mean Difference	-47.6659	-27.3584	18.2362

Standard errors clustered at the firm level. All models include controls for any salary paid (tax data), owner education, owner race, tax form filed, number of owners, presence of a patent or trademark, percent of sales made direct to consumers, tax year, and a constant. Differences are calculated as the reported tax amount minus KFS reported amount. *** p-value significant at 10 percent level, ** p-value significant at 5 percent level, * p-value significant at 1 percent level.

Discussion

A direct comparison of tax data and KFS survey data indicates that firms do not report equivalent data to each source. On average, firms reported higher receipts and expenses in the KFS data and higher profits in the tax data. In terms of applying research results for tax administration and revenue forecasting purposes, estimates based on business receipts from survey data are likely to be overstated as suggested in the previous literature, but there are also large discrepancies in expenses reporting. Expense differences could be due to a number of reasons including reporting full expenses not the tax allowable depreciation amount, using different estimation methods or reporting numbers for different business level to each source. Overall, we find higher reported profits in the tax data suggesting that using reported profits from survey sources would understate tax revenue predictions. The results suggest a nuanced relationship between tax and survey information for the purposes of informing tax administration and suggest caution in using owner-reported survey information to anticipate tax reporting or revenue. Further research is needed to uncover whether these differences in reporting lead to different conclusions about business size and growth.

References

Alm, James and Brian Erard. "Using Public Information to Estimate Self-Employment Earnings of Informal Suppliers" Tulane Economics Working Paper Series.

Andreoni, James, Brian Erard, and Jonathan Feinstein (1998). "Tax Compliance," *Journal of Economic Literature*, 36(2), pp. 818-60.

Åstebro, Thomas and Jing Chen, "The entrepreneurial earnings puzzle: Mismeasurement or real?," Journal of Business Venturing, 2014, 29 (1), 88–105.

Black, Theodore, Kim Bloomquist, Edward Emblom, Andrew Johns, Alan Plumley, and Esmeralda Stuk. 2012. Federal Tax Compliance Research: Tax Year 2006 Tax Gap Estimation. IRS Research, Analysis & Statistics Working Paper. Internal Revenue Service, Washington, DC.

Bruce, Donald. 2002. "Taxes and Entrepreneurial Endurance: Evidence from the Self-Employed." *National Tax Journal* 55(1): 5-24.

Bruce, Donald and Mohammed Mohsin. 2006. "Tax Policy and Entrepreneurship: New Time Series Evidence. *Small Business Economics* 26(5): 409-425.

Camerer, Colin and Dan Lovallo, "Overconfidence and Excess Entry: An Experimental Approach," The American Economic Review, 1999, 89 (1), pp. 306–318.

Carroll, Robert, Douglas Holtz-Eakin, Mark Rider, and Harvey S. Rosen. 2001. Personal Income taxes and the Growth of Small Firms. In *Tax Policy and the Economy, Vol 15*. Ed. J. Poterba.

Evans, David S. and Linda S. Leighton, "Some Empirical Aspects of Entrepreneurship," The American Economic Review, 1989, 79 (3), pp. 519–535.

Fairlie, Robert W. and Bruce D. Meyer. 1996. "Ethnic and Racial Self-Employment Differences and Possible Explanations." *Journal of Human Resources* 31 (4): 757-93.

Fairlie, Robert and Alicia Robb. 2009. "Why do Female-Owned Businesses have Lower Survival Rates, Profits, Employment, and Sales, than Male-Owned Businesses?," *Small Business Economics*, 33(4).

Feldman, Naomi and Joel Slemrod (2007). "Estimating Tax Noncompliance with Evidence from Unaudited Tax Returns", *Economic Journal*, vol. 117 (518), pp. 327-52.

Gurley-Calvez, Tami, Amelia Biehl, and Katherine Harper. 2009. "Time-Use Patterns and Women Entrepreneurs." *American Economic Review Papers and Proceedings*, 99(2): 139–44.

Gurley-Calvez, Tami and Donald Bruce. "Do Tax Cuts Promote Entrepreneurial Longevity?" *National Tax Journal* 61(2): 225-250

Gurley-Calvez, Tami, Donald Bruce, John Deskins, and Brian Hill. 2015. "A Tale of Two Datasets: Business Survival in Administrative Versus Survey Data." Statistics of Income Working Paper.

Hamilton, Barton H., "Does Entrepreneurship Pay? An Empirical Analysis of the Returns of Self-Employment," The Journal of Political Economy, 2000, 108 (3), pp. 604–631.

Holtz-Eakin, Douglas, David Joulfaian, and Harvey S. Rosen. 1994. "Sticking It Out: Entrepreneurial Survival and Liquidity Constraints." *Journal of Political Economy* 102(1): 53-75.

Hundley, Greg. 2001. "Why Women Earn Less Than Men in Self-Employment." *Journal of Labor Research* 22 (4): 817-29.

Hurst, Erik, Geng Li, and Ben Pugsley. "Are Household Surveys Like Tax Forms: Evidence from Income Underreporting of the Self Employed." *Review of Economics and Statistics* 96(1): 19-33.

Kukk, Merike and Karsten Staehr. 2015. "Identification of Households Prone to Income Underreporting: Employment Status or Reported Business Income?". *Public Finance Review,* published on-line December 1, 2015:

http://pfr.sagepub.com/content/early/2015/11/30/1091142115616182.abstract. Accessed December 21, 2015.
Lombard, Karen V. 2001. "Female Self-Employment and Demand for Flexible, Nonstandard Work Schedules." *Economic Inquiry* 29 (2): 214-17.
Moskowitz, Tobias J. and Annette Vissing-Jørgensen, "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?," The American Economic Review, 2002, 92 (4), pp. 745–778.
Puri, Manju and David T Robinson, "The Economic Psychology of Entrepreneurship and Family Business," Journal of Entrepreneurship and Management, 2013.
Robb, Alicia M. and John Watson. 2012. "Gender Difference in Firm Performance: Evidence from New Ventures in the United States." *Journal of Business Venturing* 27(5): 544-558.
Ross Levine & Yona Rubinstein, 2013. "Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?," NBER Working Papers 19276, National Bureau of Economic Research, Inc.
Sarada, The Unobserved Returns from Entrepreneurship (June 12, 2015). Available at SSRN: http://ssrn.com/abstract=2650636 or http://dx.doi.org/10.2139/ssrn.2650636
Slemrod, Joel (2007). "Cheating Ourselves: The Economics of Tax Evasion," *Journal of Economic Perspectives,* 21(1), pp. 25-48.