Employment, Earnings, and Deferred Compensation During the Great Recession*

Jeremy Skog
U.S. Census Bureau

December 31, 2011

Abstract

This paper investigates how well the labor force behavior in the GSF sample in the Survey of Income and Program Participation (SIPP) Gold Standard File (GSF) matches that of universe of W-2 recipients in the United States. For both groups, I use tax records from 2005 to 2009 to examine patterns in employment, earnings and deferred compensation. I examine how these patterns vary across the earnings distribution. I also link these administrative data to demographic data from the GSF, to examine how changes in employment and presence of deferred compensation were distributed across demographic groups during the Great Recession.

1 Introduction

This paper investigates how well the labor force behavior of the population in the Survey of Income and Program Participation (SIPP) Gold Standard File (GSF) matches that of the universe of W-2 recipients in the United States. This project is undertaken as part of ongoing research intended to identify potential opportunities for improvements to the Census Bureau's data products, in this case the public-use SIPP Synthetic Beta (SSB) which contains administrative data linked to survey data. The SSB data is generated from the restricted-access GSF and uses imputation and data synthesis methods with the goal of allowing unbiased estimates compared to those on the GSF while preserving the confidentiality of the survey respondents. I examine whether the underlying GSF sample behaves comparably to the overall population of W-2 recipients in a variety of workforce variables even in years when the sample was not surveyed. I find that the GSF sample behaves comparably to the universe of W-2 recipients in trends in earnings and

*I would like to thank Graton Gathright, Martha Stinson, Judith Eargle, Shawn Klimek, Sarah Tekansik, David Hedengren, and Gary Benedetto for helpful comments. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau. All of the data used in this paper are confidential data. The U.S. Census Bureau supports external researchers' use of some of these data items through the Research Data Center network (www.census.gov/ces).

deferred compensation.[1] The behavior in the GSF sample differs in having a lower average yearly number of W-2s reported per person than the overall W-2 population, and a lower proportion of people who exit or enter employment.

The SIPP GSF is a restricted-access data set consisting of person-level linked survey and administrative data which is then used to create the SSB. The SSB is a data product designed to make the benefits of using linked survey and administrative data available to researchers outside of the Census Bureau. The GSF is constructed from data from the SIPP, W-2 earnings, and OASDI and SSI benefit information and its confidential nature means that it can only be used within the Census Bureau secure network and Research Data Centers by Census staff and Special Sworn Status researchers on approved projects. A set of variables from the SIPP panels are standardized to create the GSF and missing values are then multiply imputed. The multiple imputations are then synthesized several times to create the 16 SSB implicates. This process is designed to preserve the confidentiality of the SIPP respondents while providing a potentially analytically valid set of data containing survey responses and longitudinal administrative data for public use on any topics for which the data is appropriate. The Census Bureau is actively engaged in evaluating the validity of the synthetic data. [2]

The GSF sample contains respondents to SIPP panels from 1990 to 2004. Administrative records for survey respondents in the GSF sample are then linked to the survey data, including administrative data for time periods before and after the reference period of a particular SIPP panel. The survey design cannot guarantee that a particular SIPP sample is nationally representative in years outside the survey reference period. This study will inform our goal to make the GSF sample comparable to the W-2 population by comparing the unweighted estimates of the behavior of the GSF sample with the overall W-2 population on a variety of workforce variables including during years when the constituent SIPP panels were not surveyed.

A weighting scheme could potentially eliminate any observed differences in behavior between the GSF sample and the W-2 population so that research could be conducted using data for any given year and the results would be representative of the overall W-2 population. Although development of a weighting scheme to make the W-2 and GSF data comparable is in progress, no such weighting scheme is currently available. This paper advances that goal by highlighting differences in behavior between the populations which a weighting scheme would need to address. The tests performed in this study could be repeated once a potential weighting scheme has been applied to the sample to determine if estimates of the behavior of the weighted sample differ significantly from that of the W-2 population.

The Census Bureau's "Testing Household Survey and Business Register Enhancements via Longitudinal Integration" (BRELI) project explores ways to improve and expand the Census Bureau's household surveys and Business Register data products by using administrative data from the Internal Revenue Service's (IRS) Form W-2. One of the project's goals is to research improvements to the Census Bureau's data products and household surveys, such as the SIPP, by using W-2 data on employment, wages, and contributions to tax-deferred retirement saving plans. To further this goal, I identify patterns in these vari-

[1] All comparative statements in this report have undergone statistical testing, and, unless otherwise noted, all comparisons are statistically significant at the 5 percent significance level.

[2] The Census Bureau has procedures allowing the validation of results from the SSB by replicating them on the confidential GSF. Further information on the SSB and its construction is available at www.census.gov/sipp/synth_data.html

ables over time for the entire W-2 population and examine whether the survey sample behaves similarly to the W-2 population. This is a test of whether the GSF sample is representative of the overall population of W-2 recipients and this test can identify potential differenes in the behavior of the GSF sample which would need correction through a weighting scheme.

I use administrative data on earnings and deferred compensation for both the overall W-2 population and the GSF sample and focus on changes in earnings, changes in employment status, the probability of having deferred compensation, and the levels of deferred compensation. A GSF sample which is representative of W-2 recipients would not be expected to have statistically significant differences in those variables from the overall W-2 population.

Studies which have used the GSF include Benedetto, Gathright, and Stinson (2010) who investigate the effects of graduating college in a recession; Abowd, Benedetto and Stinson (2007) who investigate the covariance of earnings and hours worked; Matthew Rutledge (2011) who examines the relationship between lifetime earnings volatility and health insurance coverage; Shaefer (2011) who examine earnings drops by educational group, and a variety of other research projects funded by the Census Bureau and National Science Foundation. Researchers face the dilemma of not knowing how representative their results are of the overall population as they have been unable to repeat their tests for the overall population. This issue is of particular concern when using administrative data for time periods before and after the survey reference periods as the representativeness of a particular SIPP sample for these time periods cannot be guaranteed by the survey design.

Several recent studies and reports have examined changes in the labor market during the recent recession. Among researchers for this subject are: Sahin, Song, and Hobijn (2009) who document an increase in the inflow rate to unemployment from employment and a disproportionate increase in inflow rates for men; Hurd and Rohwedder (2010) who document increases in households experiencing unemployment and increased expectations of unemployment; and Elsby, Hobijn, and Sahin (2010) who document increases in unemployment, and declines in employment and labor force participation as well as disparate effects across population subgroups. Autor (2010) has noted an increase in labor market polarization in recent decades. A report by Munnell, Muldoon, and Sass (2009) shows variations in the recession's effect on employment and labor force participation across age groups for men. This paper also examines changes in employment and earnings, while also examining changes in deferred compensation.

The labor market changes during great recession that are observed for the W-2 population in this paper include a reduction in earnings, average number of jobs, deferred compensation rates, and deferred compensation amounts. The GSF sample also shows a reduction in the number of jobs, saving rates, and savings amounts, but there is no statistically significant reduction in earnings for the GSF sample. Compared to the W-2 population the GSF sample has higher earnings and is more likely to save, but conditional on saving, saves lower amounts. The GSF sample works fewer jobs on average but they are less likely to lose them and a lower proportion are new entrants to the W-2 population.

The W-2 population experiences a decrease in earnings in 2008 and 2009 while the GSF sample shows no significant change in earnings. Both the GSF sample and universe of W-2 recipients experience decreases in the rate of deferred compensation in 2009 and a decline in average amounts of deferred compensation in 2009. Savers in both the GSF and the W-2 universe exhibit declining deferred compensation contributions

in 2008, and 2009. The proportion exiting employment increases each year after 2005 for the W-2 population and the GSF sample, although it increases more quickly for the GSF sample than for the W-2 population in 2006 and 2007 and more slowly for the GSF sample than for the W-2 population in 2008. Likewise, the chance of entering W-2 employment drops in 2006, 2007, 2008, and 2009 for the overall population and for 2008 and 2009 for the GSF sample. It drops more slowly for the GSF sample than for the W-2 population in 2008 and 2009.

As part of the project to research potential ways of enhancing the GSF, I also present initial results examining the importance and feasibility of adding a person's earnings percentile as a variable on the GSF. Including the earning percentile could eventually allow the Census Bureau to test how much additional information is added by using administrative records for the universe of W-2s versus just using administrative records for the GSF sample. By recalculating a person's location within the earnings distribution using only administrative data for the GSF sample and repeating the analysis using this measure of rank I will eventually be able to test whether estimates on workforce variables, such as the likelihood of saving, vary across an earnings distribution and whether estimates are significantly affected by placing a person within the a distribution calculated from the entire universe of W-2 recipients versus calculated just from data on members of the GSF sample.

To show how including such a potential variable could enhance the GSF or SSB, I demonstrate the usefulness of using a worker's position in the earnings distribution to describe how important labor market patterns differ across the distribution. I show that workforce variables such as the likelihood of saving and the likelihood of exiting employment vary significantly across the earnings distribution.

Using demographic information from survey data I show that subpopulations in the GSF vary in their likelihood of saving and likelihood of exiting employment.

The paper proceeds as follows: First I discuss the data sources and compare the summary statistics on earnings and saving over time. I then examine changes in earnings, deferred compensation, entering, and exiting employment and compare the GSF sample to the overall W-2 population. I then examine how the likelihood of exiting employment and having deferred compensation varies across the earnings distribution and across several demographic groups. Finally, I offer conclusions and suggestions for future research.

2 Data

This study uses administrative data on earnings and deferred compensation for the universe of W-2 recipients as well as the GSF sample. The use of administrative data enables me to analyze and compare the employment behavior of the GSF sample to that of the entire population of W-2 recipients. Demographic information on education, age, race, and sex from the SIPP is used to analyze subpopulations of the GSF sample.

2.1 SIPP GSF Data

The GSF sample data consists of information on 276,458 people from the SIPP Gold Standard file, a confidential Census dataset which contains linked administrative and survey data and is the basis for the SIPP Synthetic Beta data product. The GSF is created as part of a joint project with the IRS and the SSA. The goal of the project is to make some of the benefits of using linked survey and administrative data available to researchers outside of restricted-access Census Bureau facilities while protencting the confidentiality of the underlying data.

The GSF contains a set of variables standardized across the 1990, 1991, 1992, 1993, 1996, 2001, and 2004 SIPP Panels. The GSF sample consists of survey respondents to any of these SIPP panels. The SIPP is a continuous series of national panels with content built around a core of labor force, program participation, and income questions. A multistage–stratified sample of the U.S. civilian noninstitutionalized population is created for each panel and then the sample is followed for 3–4 years and interviewed periodically, although the length of time the sample is followed varies by panel.

Variables on the GSF obtained from SIPP survey data cover topics such as education, pension plan participation, occupation, fertility, and marital history. Data from each of the constituent SIPP panels is standardized to create each of the survey-based variables on the GSF and produce consistent measures across panels. The survey data is then matched to longitudinal administrative data on earnings, deferred compensation and select government benefits extracted from several files provided by the Social Security Administration.

The files which are the sources of the linked administrative data are the Master Earnings File which provides the Detailed Earnings Records (1978 – 2009) and the FICA-maximum-capped Summary Earnings Records (1951 – 2006), the Master Beneficiary Records (MBR), the Supplemental Security Records (SSR), the 831 Disability File (F831), and the Payment History Update System (PHUS). Earnings data is available for the years indicated. The availability of benefit data varies and exact dates for each variable can be found in the SSB technical documentation[3].

After the survey and administrative data is linked, the GSF data set is completed by imputing missing data multiple times. The result is four data sets with independent imputations known as the completed files. Both the completed files with imputed data and the underlying GSF are confidential and may only be accessed on the Census Bureau's secure network by Census employees or Special Sworn Status researchers on approved projects.

The data on each of the completed files are then synthesized conditional on their values in the completed files. Only sex, spouse link, and OASDI benefit type are not synthesized. After the data is synthesized, the SSB sample is limited to those persons whose synthesized age is greater than or equal to 15 on January first of the first year of his SIPP panel. This censoring results in slightly different numbers of people in each implicate. The result of this data creation process is the sixteen implicates which comprise the SSB data product.

The creation of the SSB aims to create analytically valid data which supplies point estimates which

[3]Available at http://www.census.gov/sipp/synth_data.html

are unbiased and estimated variances which support similar inferences as would be found by using the underlying imputed data while preserving the confidentiality of the SIPP respondents. The SSB is publicly available for use and may be used for any research for which the data is appropriate. Researchers who wish to use the data product can complete an application available from the SSB website to apply for a free account on the Synthetic Data Server housed at the virtual RDC at Cornell University. This account allows the researcher to accesss the SSB and use SAS and STATA software.

Using the GSF or SSB allows researchers to include more observations by pooling the SIPP panels. They can also take advantage of the properties of administrative data which exists for longer time periods than the years when a particular SIPP panel was surveyed, but the cost of using data for this longer time period is that they also lose that guarantee that their data set is representative of the overall population in all years.

Each panel of the SIPP is designed to be nationally representative for the time period the survey is conducted. The Census bureau creates sampling weights for each SIPP panel and performing unweighted analysis may produce biased results. While weights are constructed for each SIPP panel, there are currently no weights for the GSF or SSB to account for the inclusion of administrative data for periods outside the reference period for a particular SIPP sample, which introduces a potential source of bias. Investigating differences in the GSF estimates and the W-2 universe estimates as a step toward producing such weights for the GSF and SSB motivates this research.

The number and distribution of persons included in this study by SIPP panel is shown in Table 1. Inclusion in the GSF sample is dependent on being matched to administrative data on the W-2 population and therefore on having W-2 earnings at some time between 2005 and 2009. While each panel was representative of households when it was designed, survey respondents may have left or entered the universe of W-2 recipients before and after that time period. The survey design does not guarantee that the GSF sample reflects the current wage-earning population. The number of GSF persons with earnings in the administrative records varied by year from a high of 232,314 in 2005 to a low of 215,903 in 2009.

Comparisons with the administrative data confirm that the GSF survey sample is distributed fairly uniformly across the earnings distribution. I examined what proportion of the GSF sample was in each earnings percentile and decile in each year. Results for earnings deciles are presented in Table 2. In 2005 each earnings decile contained between 9.39% and 10.55% of the GSF survey population. In 2009 the deciles contained between 9.38% and 10.72% of the GSF survey population. The W-2 population is distributed evenly across the earnings distribution.

The distribution, though, is slightly skewed towards higher earnings with the average decile below the median earnings level having a size of 9.64% of the GSF sample in 2005 and the average percentile at or above the median earnings level having 10.36% of the GSF sample. This could be because the GSF includes older SIPP panels. A SIPP panel can be weighted to be representative when it is conducted, but over time the members of the panel age and are eventually the average panel member is older than the average member of the overall population.

Table 1: Panel Person-Observations from SIPP Gold-Standard File)

Panel	N	Percent
1990	37,783	16.76%
1991	23,594	8.53%
1992	33,575	12.14%
1993	33,029	11.95%
1996	58,261	21.07%
2001	30,763	11.12%
2004	59.480	11.12%
Total	276,458	100%

Table 2: Percent of GSF Sample in Each Earnings Decile by Year

Earnings Decile	2005	2006	2007	2008	2009
0	9.39	9.4	9.45	9.38	9.23
1	9.55	9.53	9.49	9.52	9.6
2	9.61	9.65	9.59	9.52	9.6
3	9.8	9.62	9.56	9.57	9.61
4	9.85	9.93	9.89	9.78	9.78
5	10.15	10.17	10.11	10.23	10.13
6	10.31	10.34	10.29	10.31	10.36
7	10.52	10.51	10.6	10.61	10.61
8	10.55	10.57	10.6	10.72	10.72
9	10.29	10.32	10.39	10.34	10.34

2.2 Administrative Data

This study also uses administrative data from the universe of W-2 tax filings between 2005 and 2009. These data are obtained from the IRS by the Census Bureau as part of the BRELI project which uses administrative data to explore ways to improve and enhance the Bureau's data products.

These administrative data are used to compare estimates for the GSF sample with the population of W-2 recipients.

The administrative data contains data for each Employer Identification Number (EIN) for each person for a whom that EIN filed a W-2. The data contains information on wages paid and deferred compensation. Employers file W-2 forms annually with the IRS. Each W-2 record should in theory represent a single job held by the respondent at some point during the year. Each person-record in the GSF and administrative data is identified by a Protected Identification Key (PIK) and a Employer Identification Number. The PIK is a confidentiality-protected version of the Social Security Number used by the Census Bureau to prevent identification of individuals in the data. An EIN can identify either a business enterprise or a unit within an enterprise. These identifiers allow data to be summarized at either the person level across multiple employers, or at the employer level across multiple employees.

In addition to the identifiers, W-2 information is provided on reported wages and tips and the amount of deferred compensation. Earnings are calculated as the sum of the "Wages, Salaries, and Tips" field and the "Deferred Compensation" field. Deferred compensation is tax-advantaged saving set aside in a formal arrangement with the employer, such as a 401(k) or 403(b) plan, and is reported in the "Deferred Compensation" field. These values are then summed for each person across all EINs to calculate total earnings and deferred compensation for each person in each year. Individuals are recorded as having no employment in the W-2s population in a year if they have no reported earnings for that year. All total values are converted to 2009 constant dollars using the Consumer Price Index for urban wage-earners.

Deferred compensation receives tax advantages, but there is a tradeoff in reduced liquidity. The funds allocated to deferred compensation accounts must be spent in specific ways if they are accessed before age 59 1/2 or the employee must pay a penalty. When the worker makes the decision to defer some of their compensation they are making a tradeoff between tax advantages and liquidity. Deferred compensation is more valuable to the employee as liquidity constraints decrease and marginal tax rates increase. I posit that both of these conditions are more likely as the level of earnings rises.

I also calculate entry into and exit from the W-2 population. Because data is observed on a yearly basis, I cannot analyze intra-year changes. I record a person as having exited the W-2 population if he had earnings in one year and then not in the next year. Because this requires two years of observation I can only calculate exiting W-2 employment for 2005, 2006, 2007, and 2008. The estimated values are the proportion of the W-2 population which has earnings in the base year, but no earnings in the next year. The reasons why a person might exit the W-2 workforce include exiting the labor force, entering long-term unemployment, or shifting to employment which is not reported via a W-2 form. I also calculate whether a person entered W-2 employment. This is calculated as people with no earnings in the base year, but with earnings in the next year. This is calculated for years 2006, 2007, 2008, and 2009. The estimated values give the proportion of the labor force in each year which had no W-2 earnings in the prior year. A person

could enter the W-2 population if they begin working in a W-2 reporting job, or if their employer shifts to reporting the person's earnings via a W-2 form. A person must have had W-2 earnings in at least one year to be in the W-2 population.

Because the administrative data is provided by the employer, common household survey errors such as non-response or recall error should be minimal as these reporters are subject to legal sanctions for false statements. There may be measurement error because of reporting issues, data entry errors, or other sources of error associated with administrative data. Checks for unusually large values revealed that the administrative data contained 212 observations with reported yearly earnings greater than $100,000,000; 3,399 observations with reported deferred compensation greater than $100,000, and 345 observations with more than 200 W-2s. These observations were dropped from this analysis. The GSF had one observation which reported more than $100,000 in deferred compensation and one observation which reported more than 200 W-2s.

Conversely, I do not drop people from the lower end of the distribution and therefore the earnings percentiles which I calculate may be lower than those shown in other studies which do drop people who have low earnings. A commonly-used methodology is to drop anyone earning less than one quarter of the full-time minimum wage or about $2,600. However, I analyze all workers with reported earnings in a given year which do not exceed the maximum values described above.

The resulting administrative data set consists of 194,467,544 individuals who were employed at least once in the years 2005 through 2009.

2.3 Labor Market Characteristics and Yearly Comparisons

I calculate various summary statistics for the population of W-2 recipients and the GSF including the total number of people receiving W-2s, the number of W-2s observed in a given year, the number of EINs issuing W-2s per year, the median earnings percentile of the sample, the mean and median earnings, the percent of the W-2 population with deferred compensation and the average dollars of deferred compensation per person. I examine whether the values presented are different from those in the prior year by performing a Wald Test on the estimated coefficients from the dummy variable regressions presented in the results section. This tests whether the the year-on-year values are statistically significantly different from each other and therefore represent a change in the values over time. These statistics are presented in Table 3.

The number of people receiving W-2s in a given year peaked in 2007 with 160,435,190 people who had positive earnings. In 2009 the number of people receiving W-2s was 155,091,971. The overall number of W-2s dropped from a peak of 236,155,334 in 2007 to 208,917,311 in 2009. There are a number of a reasons a person might have more than on W-2 form. It could represent a second or third concurrent job, multiple jobs within a year either due to job transitions or multiple employment spells or it could be due to an administrative realignment or corporate merger. The GSF sample received a peak number of 347,202 W-2s in 2006. This number declined to 291,773 in 2009, a statistically significant reduction.

I present the counts of the numbers of EINs filing W-2 forms for the overall population and for those who were in the GSF, but do not test them for significance as the GSF is not intended to be an employer-level survey. The counted number of EINs for the W-2 population was 6,422,415 in 2007 and the counted

value in 2009 was 6,232,548, although the values were not tested to determine if the difference is statistically significant. If the change is significant it could be due to either fewer companies in operation or to fewer administrative changes occurring requiring an intra-year change in EIN.

The GSF sample earns more than the population as a whole on average, although inter-year changes are not significant for the GSF sample while there are significant changes in average earnings occurring among all W-2 recipients. The higher earnings for the GSF population show up in a higher median earnings percentile than that for the population as a whole. Average earnings for the W-2 population in 2009 were below their 2007 value and not significantly different from the value in 2006.

For both the W-2 population and the GSF sample, the number of people with deferred compensation peaked in 2008 and dropped in 2009, although it was still above levels seen before 2008. This could be due to selection effects in who exits employment. A large drop in the employment rates by those without deferred compensation would show up in the data as an increase in saving rates even if saving behavior among earners is unchanged. While a higher pecentage of working people were saving, the amount that they were saving decreased for the W-2 population in 2008 and 2009 while levels for the GSF population were not significantly different in 2008 but dropped from this level in 2009. Those in the GSF save at a higher rate than the overall population and save more on average when non-savers are included. Again, this could be due to either age- or earnings-related saving motivations.

Table 3: Summary Statistics For Whole Population and Gold-Standard File

Sample	Year	Number Employed	Number W-2s	Number EINs	Median Rank	Mean Earnings	Median Earnings	% Saving	Average Saving
All	2005	156,061,887	230,174,721	6,349,797	49	$38,273	$25,637	29.13%	$1,326
	2006	158,687,967	235,271,985***	6,414,233	49	$38,751***	$25,762	29.60%***	$1,369***
	Change 05-06	*1.67%*	*2.19%*	*1.01%*		*1.24%*	*0.49%*	*1.60%*	*3.19%*
	2007	160,435,190	236,155,334***	6,422,415	49	$39,522***	$26,053	30.83%***	$1,404***
	Change 06-07	*1.10%*	*0.37%*	*0.13%*		*1.97%*	*1.12%*	*4.09%*	*2.52%*
	2008	160,377,190	228,958,896***	6,407,779	49	$38,879***	$25,779	31.85%***	$1,391***
	Change 07-08	*-0.18%*	*-3.09%*	*-0.23%*		*-1.64%*	*-1.06%*	*3.26%*	*-0.93%*
	2009	155,091,971	208,917,311***	6,232,548	49	$38,751***	$25,694	31.35%***	$1,355***
	Change 08-09	*-3.20%*	*-9.16%*	*-2.77%*		*-0.33%*	*-0.33%*	*-1.59%*	*-2.62%*
	Change 07-09	*-3.39%*	*-12.26%*	*-3.00%*		*-1.97%*	*-1.39%*	*1.68%*	*-3.55%*
GSF	2005	232,314	344,680	161,351	52	$39,008	$27,455	31.82%	$1,422
	2006	232,277	347,202**	162,345	52	$40,022	$27,930	32.67%***	$1,478***
	Change 05 - 06	*-0.02%*	*0.73%*	*0.61%*		*2.57%*	*1.71%*	*2.64%*	*3.87%*
	2007	230,500	342,117***	162,607	53	$41,168	$28,634	34.45%***	$1,553***
	Change 06 - 07	*-0.77%*	*-1.48%*	*0.16%*		*2.82%*	*2.49%*	*5.28%*	*4.98%*
	2008	225,687	324,579***	156,767	54	$41,070	$28,956	35.92%***	$1,558
	Change 07 - 08	*-2.11%*	*-5.26%*	*-3.66%*		*-0.24%*	*-1.12%*	*4.20%*	*0.28%*
	2009	215,903	291,773***	144,025	54	$41,606	$29,306	35.71%***	$1,534***
	Change 08 - 09	*-4.43%*	*-10.66%*	*-8.48%*		*1.30%*	*1.20%*	*-0.59%*	*-1.54%*
	Change 07 - 09	*-6.54%*	*-15.92%*	*-12.13%*		*1.06%*	*2.32%*	*3.61%*	*-1.26%*

Year changes are calculated as logarithmic changes in values

Wald-tests were performed on estimates from regressions to determine whether population values were different from prior-year values. The columns tested were number of W-2s, mean earnings, percent of working population with deferred compensation, and mean level of deferred compensation.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Number of W-2s, Mean Earnings, % Saving, and Average Saving tested for statistical significance. Other variables were not tested.

Public Use SSB data can be found at http://www.census.gov/sipp/synth_data.html

3 Results

I examine several workforce variables to determine whether the GSF sample behaved similarly to the overall labor force during the Great Recession, including earnings, log earnings, the percent of the population with deferred compensation, average deferred compensation, the log of average deferred compensation, the average level of deferred compensation among those with deferred compensation, the number of W-2s each person had per year, the proportion entering employment in a year, and the proportion exiting employment in a year.

For each of these variables I estimated two econometric models. The first only includes a flag for whether a person was in the GSF to test for an overall difference in levels. A significant coefficient on the GSF dummy variable indicates that the GSF sample has a different level than the W-2 receiving population. The second specification includes dummy variables for the GSF sample, each year 2006 through 2009, and each year interacted with the GSF sample. The omitted year, 2005, is captured by the constant term. A series of Wald tests were performed on this regression to test the coefficients and determine whether the values for each variable differed between years. T-tests on the coefficients of the interaction variables determined whether the GSF sample differed from the overall population in a given year.

3.1 Earnings

The earnings regressions in Table 4 indicate that the GSF sample has higher earnings than the overall W-2 receiving population. The second specification shows no yearly significant difference in earnings behavior between the GSF sample and the overall population, although the GSF sample still has higher earnings across all years. Relative to the base year (2005) the W-2 population had higher earnings in every other year and the peak year for earnings was in 2007. The earnings for the GSF were not significantly different than the population in any year or from previous years for the GSF sample.

Using untransformed earnings provides coefficients which are easy to interpret, but may lead to problems in estimation because earnings cannot be negative and there is a long right tail in the earnings distribution. To help account for this I also performed the regressions using the natural log of individual yearly earnings and show the results in Table 5. This monotonic transformation may reduce the effect of very-high earnings on the estimates.

The GSF-indicator-only regression indicates that the GSF sample has statistically significantly higher earnings than the overall population. The dummy variable regression indicates that the GSF sample has higher earnings, but that they do not trend significantly differently over time. In each of the years 2006 – 2009 the W-2 population has higher earnings than in 2005, with the highest value in 2007 and declining thereafter. The GSF sample has higher earnings in 2006 than in 2005, but there is no statistically significant change after that year.

Table 4: Yearly Earnings

	(1)	(2)
GSF	727.3***	555.8*
	(6.40)	(2.19)
2006		473.6***
		(34.17)
2007		1243.8***
		(89.98)
2008		607.4***
		(43.93)
2009		300.5***
		(21.56)
GSF * 2006		200.0
		(0.56)
GSF * 2007		36.61
		(0.10)
GSF * 2008		104.9
		(0.29)
GSF * 2009		533.2
		(1.47)
Constant	38747.1***	38217.6***
	(8860.58)	(3883.36)
N	791584025	791584025

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Yearly Ln(Earnings)

	(1)	(2)
GSF	0.0548***	0.0509***
	(39.53)	(16.41)
2006		0.00853***
		(50.26)
2007		0.0120***
		(70.74)
2008		0.00254***
		(15.01)
2009		-0.00216***
		(-12.67)
GSF * 2006		0.00287
		(0.66)
GSF * 2007		0.00229
		(0.52)
GSF * 2008		0.00644
		(1.47)
GSF * 2009		0.00797
		(1.80)
Constant	9.814***	9.810***
	(183579.61)	(81358.36)
N	788425632	788425632

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.2 Deferred Compensation

Deferred compensation represents a tax-advantaged, liquidity-restrained form of retirement saving defined in a formal arrangement between employers and employees. However, not all people are eligible for retirement saving plans from their employer so we should expect that this restricted form of saving has lower levels of participation than unrestricted forms of saving. Taking money out of a retirement account before reaching age 59 1/2 triggers a penalty which the account holder must pay. There are some approved uses for the money which do not trigger the penalty, but in general deferred compensation reported on a W-2 form represents saving allocated to long-term capital growth and retirement preparedness.

I first examined the percent of the population with any deferred compensation by using a dummy variable flag for people who had both earnings and deferred compensation, versus those who had earnings but no reported deferred compensation. Results are shown in Table 6.

The GSF-flag only regression indicates that those in the GSF sample save at a higher rate than the overall population. The dummy variable regression shows that the GSF sample saves at a higher rate (about 2.5% more of the GSF sample has deferred compensation) than the overall workforce. The interaction dummies are insignificant, however, indicating that the GSF sample does not follow a different trend than the overall workforce. Each of the year dummies is significant with a peak value in 2008. Year-on-year changes in levels are significant for both the W-2 population and the GSF sample.

I examine the level of deferred compensation in several ways. First, the average level of saving among the working population (including zeros values for deferred compensation), then the level of saving among savers, and finally the natural log of the level of saving among savers (no zero values for deferred compensation).

The level of deferred compensation among all workers (Table 7) is higher for the GSF sample than for the overall population in the GSF-flag regression. The dummy variable regression indicates that the level of deferred compensation among all workers peaked in 2007, declining each year afterward. In 2007 the GSF sample increased its deferred compensation more than for the W-2 population and in 2008 the GSF sample did not decrease its deferred compensation by as much. The GSF sample saved statistically different amounts than in the prior year in 2006, 2007, and 2009. Levels for the W-2 population were statistically different from the prior year in 2006, 2007, 2008, and 2009.

The previous results indicating that the GSF sample saves more on average than the overall workforce reverse when only examining savings levels by those who saved some amount: savers in the GSF sample save less on average than savers in the overall workforce.

The GSF dummy is negative and significant in the GSF only regression indicating a lower overall level of deferred compensation by those with some deferred compensation. In the dummy variable regression the GSF dummy variable is negative but the interaction terms are not statistically significant, indicating that savers in the GSF sample behave similarly to savers in the W-2 population over time. Changes in the level of saving for the savers in the W-2 population are statistically significant for every year and changes for the GSF sample are statistically significant for 2006 and 2008.

I also examine the log of deferred compensation to help compensate for the fact that savings cannot

Table 6: Percent of Population with Deferred Compensation

	(1)	(2)
GSF	0.0254***	0.0248***
	(59.75)	(25.99)
2006		0.00470***
		(90.59)
2007		0.0171***
		(329.36)
2008		0.0273***
		(526.66)
2009		0.0223***
		(426.52)
GSF * 2006		0.000962
		(0.72)
GSF * 2007		0.00137
		(1.02)
GSF * 2008		0.000965
		(0.72)
GSF * 2009		0.000436
		(0.32)
Constant	0.306***	0.291***
	(18648.45)	(7900.45)
N	791584025	791584025

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Yearly Deferred Compensation

	(1)	(2)
GSF	93.53***	79.27***
	(28.48)	(10.79)
2006		32.57***
		(81.03)
2007		71.10***
		(177.64)
2008		57.65***
		(143.96)
2009		21.90***
		(54.26)
GSF * 2006		10.29
		(0.99)
GSF * 2007		23.52*
		(2.27)
GSF * 2008		21.31*
		(2.05)
GSF * 2009		16.84
		(1.61)
Constant	1370.5***	1333.5***
	(10827.82)	(4671.18)
N	788425632	788425632

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Yearly Deferred Compensation Among Savers

	(1)	(2)
GSF	-46.39***	-73.18***
	(-5.53)	(-3.81)
2006		40.04***
		(36.86)
2007		9.576***
		(8.93)
2008		-178.5***
		(-167.63)
2009		-222.6***
		(-206.64)
GSF * 2006		15.24
		(0.56)
GSF * 2007		29.16
		(1.09)
GSF * 2008		43.39
		(1.63)
GSF * 2009		39.73
		(1.48)
Constant	4467.3***	4540.0***
	(13293.94)	(5862.39)
N	241905485	241905485

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

be negative. The results for the log of deferred compensation indicate that savers in the GSF do not save more or less than savers in the overall population. The coefficient on the GSF dummy is positive and statistically significant in the basic GSF flag only regression, but not statistically significant once year effects are included. The dummy variable regression shows no significant difference between the GSF sample and the overall workforce over time. The coefficients for year dummies for 2006, 2007, 2008, and 2009 are all negative and statistically significant. Year-on-year changes are statistically significant in each year for both the W-2 population and the GSF sample.

3.3 Number of W-2s

Using the W-2 tax records I am able to determine how many W-2s a person received each year. These roughly correspond to the number of jobs each person held, although a person may also receive multiple W-2s because he is paid by more than one administrative unit or because there is an organizational change.

The regression results indicate that the GSF sample has statistically significantly fewer W-2 forms than the overall workforce. The dummy-flag-only regression indicates that each person in the W-2 population has an average of about 1.5 W-2s in each year and those in the GSF sample have about 0.01 fewer W-2s than this on average. The dummy variable regression again shows an average of about 1.5 W-2s per person for each year although the numbers are statistically significantly lower than the prior year in 2007, 2008, and 2009 for both the overall W-2 population and the GSF sample. The slight increase in 2006 from 2005 is statistically significant for both the population of W-2 recipients and the GSF sample. In 2008 and 2009 the average number of W-2s is not reduced as much for the GSF sample as for the overall workforce, although the statistically significant GSF-flag indicates that the GSF sample has fewer W-2s on average.

3.4 Exiting W-2 Employment

Tax records are observed on a yearly basis so it is not possible to easily discern intra-year movements between jobs or into or out of employment. However, when a person has no earnings for an entire year, I can record them as having exited W-2 employment in the prior year. For example, a person who has no W-2 reported earnings for the first time in 2008 would have exited that employment at some point during 2007. Because I require two years of observation to determine if a person in employment has exited from it, this regression does not include the year 2009. Because I am observing the W-2 universe, not having W-2 reported earnings in a year could mean that a person has exited the labor force, entered long-term unemployment, or shifted to a type of employment where their earnings are not reported using a W-2 form. I can not distinguish between these methods and only examine whether someone has exited the population of W-2 recipients.

Regression results indicate that the GSF sample is less likely to exit employment overall. The dummy variable regression indicates that the GSF sample is again statistically significantly less likely to exit employment overall and relatively more likely to exit W-2 employment than the overall W-2 population in 2006 and 2007 but relatively less likely to exit in 2008. That is, the rate of increase in the proportion of the population exiting employment increased faster for the GSF sample than the overall population in 2006 and

Table 9: Yearly Ln(Deferred Compensation)

	(1)	(2)
GSF	0.00464*	-0.00149
	(2.09)	(-0.29)
2006		-0.0258***
		(-89.70)
2007		-0.0621***
		(-218.44)
2008		-0.107***
		(-379.30)
2009		-0.157***
		(-550.43)
GSF * 2006		0.00138
		(0.19)
GSF * 2007		0.00435
		(0.61)
GSF * 2008		0.0109
		(1.55)
GSF * 2009		0.0111
		(1.56)
Constant	7.678***	7.750***
	(86179.03)	(37767.47)
N	241905485	241905485

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Number of W-2s Per Year

	(1)	(2)
GSF	-0.0102***	-0.0141***
	(-11.72)	(-7.27)
2006		0.00781***
		(73.62)
2007		-0.00329***
		(-31.13)
2008		-0.0498***
		(-470.33)
2009		-0.134***
		(-1259.60)
GSF * 2006		0.000518
		(0.19)
GSF * 2007		0.00211
		(0.77)
GSF * 2008		0.00687*
		(2.50)
GSF * 2009		0.00905**
		(3.27)
Constant	1.462***	1.497***
	(43611.24)	(19881.02)
N	791584025	791584025

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Exiting Employment

	(1)	(2)
GSF	-0.0403***	-0.0416***
	(-179.40)	(-107.18)
2006		0.00144***
		(56.71)
2007		0.00666***
		(262.60)
2008		0.0203***
		(801.27)
GSF * 2006		0.00788***
		(12.83)
GSF * 2007		0.00507***
		(8.25)
GSF * 2008		-0.00329***
		(-5.35)
Constant	0.0672***	0.0601***
	(7488.44)	(3350.95)
N	779248540	779248540

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2007, and more slowly than the W-2 population in 2008. Each year saw a statistically significant increase over the prior year in the proportion of the population exiting employment for the overall population and the GSF sample.

3.5 Entering W-2 Employment

As with the proportion of the workforce exiting employment, I examine the proportion entering into employment. The yearly nature of tax records mean that I cannot determine intra-year changes but I can observe if a person has no W-2 reported earnings in one year and has W-2 reported earnings the next year. I measure whether a person has earnings after a non-employment year for the years 2006 through 2009 and examine the proportions of the W-2 population and GSF sample which had no earnings in the prior year. Entering into the population could represent either taking on new employment or having the method the employer uses to report earnings shift to using the W-2 form.

Table 12: Entering Employment

	(1)	(2)
GSF	-0.0420***	-0.0535***
	(-188.77)	(-139.42)
2007		-0.00243***
		(-96.64)
2008		-0.00845***
		(-335.64)
2009		-0.0192***
		(-762.02)
GSF * 2007		0.0121***
		(19.93)
GSF * 2008		0.0147***
		(24.13)
GSF * 2009		0.0214***
		(35.12)
Constant	0.0662***	0.0737***
	(7428.23)	(4138.96)
N	779248539	779248539

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The GSF sample is less likely to enter employment during the observation period. Although the GSF sample is less likely to enter employment overall, the changes in the proportion of the population entering employment are less negative over time for the GSF sample than the overall W-2 workforce, as shown by the positive and statistically significant interaction terms. Both the overall W-2 workforce and the GSF sample have lower proportions entering into W-2 employment in each of the years 2007, 2008, and 2009 than in the prior year.

3.6 Multivariate Regressions

I observe that the values of several of the workforce variables vary greatly across the earnings distribution. For example, those in higher earnings percentiles exhibit higher rates of saving and lower rates of exiting employment. To account for this, I use calculated earnings percentiles from the W-2 records to control for the effect of a person's location in the earnings distribution on the workforce variables. When creating these percentiles, I do not bottom censor the earnings as the administrative data represents recorded pay.

These percentiles may not, therefore, be directly comparable to other percentiles which are calculated using additional sources of income or for households rather than individuals.

I use the demographic information in the Gold Standard file to examine the how various demographic groups experienced two of the previously examined workforce measures: the probability of having deferred compensation and the probability of exiting employment. I first describe how the workforce attributes have varied for each subpopultion in the GSF over time. I then conduct yearly regressions for demographic variables on the saving and exiting employment. Because it uses information from the survey data, the demographic analysis is only conducted for the GSF sample. I show tables of summary statistics on the workforce variables for the GSF sample population split by birth cohort (Table 13), sex (Table 14), race (Table 15), and education (Table 16).

3.6.1 Birth Cohort

To examine the effects of age on the workforce variables, I split the GSF into birth cohorts by decade of birth. Variables are tested for year-on-year differences using a regression with year dummy variables but groups are not tested against each other for intra-year differences. The GSF sample was limited to those who were 16 or older in 2005, so all of the years within each decade are included. There are also some workers who were born before the 1940s, but statistics are not reported due to their small numbers and the fact that those remaining in the workforce may not be representative of the whole group due to large retirement effects. The cross-tabulations do not distinguish between cohort effects, such as average education level of a group and the effects of being a different age and at a different point in one's working life when the recession occurred.

The oldest cohort examined is those born in the 1940s. They are at or approaching age 65 when they qualify for unreduced government retirement benefits and every year shows fewer of them working. Wages are significantly lower in 2008 while fewer of those who are working are saving for retirement in 2009. Average savings drop in 2008 and 2009 for those born in the 1940s. The drop in saving rate and amounts in 2009 were also statistically significant for those born in the 1950s, and 1980s while the drop in rates was statistically significant for those born in the 1960s.

Those born in the 1980s have the lowest estimated earnings and the lowest values in the earnings distribution, as a group. However, they have quickly risen over time with significant increases in average earnings every year, perhaps because more are completing their educations and entering full-time work. More of them also save over in each successive year and the average amount saved increases. The years after the recession saw statistically significant saving rates and amounts for this cohort.

The observed pattern is that average earnings increases with age through one's forties and fifties and then decreases. People begin saving for retirement in earnest during their thirties and increase their savings over time. It is difficult to isolate the effects of age from the effects of being a given age in a certain year. Over time, as the very oldest move down in the earnings distribution, the youngest move up.

Table 13: Summary Statistics by Birth Cohort

Birth Cohort	Year	Number Employed	Number W-2s	Median Rank	Mean Earnings	Median Earnings	% Saving	Average Saving
Born During 1940's	2005	27,604	34,413	59	$47,193	$32,930	39.24%	$2,352
	2006	26,292	32,820	57	$46,636	$31,745	38.69%	$2,368
	2007	24,719	30,527*	55	$47,224	$30,451	38.27%	$2,356
	2008	23,033	28,123*	53	$43,143***	$28,395	37.49%	$2,217**
	2009	20,915	24,885***	51	$41,799	$26,807	35.45%***	$2,088**
Born During 1950's	2005	49,368	65,525	66	$53,156	$39,373	43.68%	$2,283
	2006	48,825	64,621	66	$54,446	$39,488	43.91%	$2,408***
	2007	48,120	63,143**	66	$54,544	$39,501	45.25%***	$2,538***
	2008	47,080	60,646***	66	$53,655	$38,865	46.10%**	$2,533
	2009	45,373	56,765***	65	$52,284*	$38,157	44.74%***	$2,462*
Born During 1960's	2005	52,381	73,604	64	$49,694	$37,171	40.02%	$1,786
	2006	52,229	73,130	64	$51,337	$37,731	41.02%**	$1,867***
	2007	52,132	72,591	65	$52,723	$38,321	42.51%***	$1,971***
	2008	51,686	70,445***	65	$52,670	$38,072	43.45%**	$1,972
	2009	50,189	65,324***	65	$53,445	$38,228	42.60%**	$1,931
Born During 1970's	2005	48,617	76,113	55	$35,653	$29,939	31.75%	$971
	2006	48,245	74,981	56	$37,841***	$31,235	33.75%***	$1,083***
	2007	48,092	73,598***	58	$39,814***	$32,533	36.55%***	$1,215***
	2008	47,425	69,666***	59	$40,712***	$32,786	38.2%***	$1,265**
	2009	45,911	62,900***	59	$41,723***	$33,411	38.12%	$1,280
Born During 1980's	2005	46,050	85,301	22	$12,911	$8,261	8.3%	$112
	2006	49,242	92,663**	25	$14,909***	$9,849	11.15%***	$161***
	2007	50,759	94,476*	28	$17,230***	$12,076	14.76%***	$232***
	2008	50,547	88,817***	32	$19,328***	$14,218	18.59%***	$307***
	2009	48,441	76,109***	35	$21,377***	$15,838	20.37%***	$354***

F-tests were performed on Dummy variable regressions to determine whether values were different from prior-year values. Tested values include number of W-2s, mean earnings, percent of working population with deferred compensation, and mean level of deferred compensation.

3.6.2 Sex

The men in the GSF reported higher average earnings than women, although there was a statistically significant reduction in men's average earnings in 2008. Women reported statistically significant increases in average wages in 2006, 2007, and 2009. Saving rates for both men and women peaked in 2008 with a statistically significant decline in 2009. Both men and women saw a statistically significant drop in average saving amounts in 2009 after seeing statistically significant increased in 2006 and 2007. The observed median percentile rank for women has been rising over time while the median percentile rank for men has held steady. Both men and women saw a statistically significant reduction in the number of jobs worked in both 2008 and 2009. The decrease in number of W-2s in 2007 was not statistically significant for women, but was for men.

Table 14: Summary Statistics by Sex

Sex	Year	Number Employed	Number W-2s	Median Rank	Mean Earnings	Median Earnings	% Saving	Average Saving
Women	2005	115,655	171,535	45	$29,571	$22,495	30.46%	$1,095
	2006	115,706	173,193**	45	$30,271**	$22,910	31.3%**	$1,139**
	2007	114,836	170,825	46	$31,140**	$23,587	33.11%***	$1,199***
	2008	112,474	161,958***	47	$31,505	$24,074	34.5%***	$1,213
	2009	108,056	146,534***	48	$32,456***	$24,887	34.27%**	$1,196*
Men	2005	116,659	173,145	60	$48,363	$34,123	33.17%	$1,745
	2006	116,571	174,009	61	$49,701	$34,585	34.04%**	$1,815**
	2007	115,664	171,292**	61	$51,124	$35,274	35.77%***	$1,905***
	2008	113,213	162,621***	62	$50,573*	$35,278	37.34%***	$1,901
	2009	107,847	145,239***	61	$50,773	$35,119	37.16%**	$1,872**

F-tests were performed on Dummy variable regressions to determine whether values were different from prior-year values. Tested values include number of W-2s, mean earnings, percent of working population with deferred compensation, and mean level of deferred compensation.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Number of W-2s, Mean Earnings, % Saving, and Average Saving tested for statistical significance. Other variables were not tested.

Public Use SSB data can be found at http://www.census.gov/sipp/synth_data.html

3.6.3 Race

Table 15 gives workforce statistics broken down by racial category. The three racial categories recorded and standardized across the SIPP panels are black, other race, and white.

There were no significant differences by year for wages for those reporting black or other for race. Blacks saw a significant increase in saving rates in 2007 and 2008. None of the changes in the average amount saved were statistically significant for blacks and only 2006 saw a statistically significant increase in the amount saved for those reporting a race besides white or black. Whites saw statistically significant increases in the amount of deferred compensation in 2006 and 2007 and a statistically significant drop in average earnings in 2009. Each racial category reported statistically significantly fewer W-2s in both 2008 and 2009 and the drop in number of W-2s for whites was also statistically significant in 2007. The saving rate among whites was statistically different in each year, peaking in 2008 with a drop in 2009.

Table 15: Summary Statistics by Race

Race	Year	Number Employed	Number W-2s	Median Rank	Mean Earnings	Median Earnings	% Saving	Average Saving
Black	2005	27,450	44,559	41	$27,366	$20,426	24.67%	$645
	2006	27,750	45,996***	42	$27,604	$20,574	25.4%	$659
	2007	27,560	45,268*	42	$28,343	$21,166	27.21%***	$697
	2008	26,939	42,263***	43	$28,712	$21,629	29.08%***	$709
	2009	25,389	36,156***	44	$30,021	$22,148	29.18%	$699
Other Race	2005	10,995	16,328	52	$40,720	$27,640	34%	$1,851
	2006	11,048	16,471	53	$42,446	$28,709	35.64%*	$2,006*
	2007	11,043	16,308	54	$43,972	$29,540	36.96%	$2,116
	2008	10,834	15,412***	55	$44,749	$29,946	38.95%*	$2,161
	2009	10,314	13,898***	56	$50,924	$30,762	38.79%	$2,201
White	2005	193,869	283,793	53	$40,559	$28,603	32.71%	$1,507
	2006	193,479	284,735	54	$41,664**	$29,140	33.55%***	$1,565***
	2007	191,897	280,541**	55	$42,849*	$29,843	35.34%***	$1,644***
	2008	187,914	266,904***	55	$42,630*	$30,128	36.73%***	$1,645
	2009	180,200	241,719***	56	$42,704	$30,419	36.46%***	$1,613***

F-tests were performed on Dummy variable regressions to determine whether values were different from prior-year values. Tested values include number of W-2s, mean earnings, percent of working population with deferred compensation, and mean level of deferred compensation.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Number of W-2s, Mean Earnings, % Saving, and Average Saving tested for statistical significance. Other variables were not tested.

Public Use SSB data can be found at http://www.census.gov/sipp/synth_data.html

3.6.4 Education

Table 16 shows the statistics for the five education categories examined: those without a high school diploma, high school graduates, those who have attended some college, those who graduated from college (and obtained a bachelor's degree) and those who have a degree beyond a bachelor's degree. These categories were created from a person's responses to SIPP survey questions. Education is observed during that sample period and not during out-of-sample periods so some respondents may since have increased their education level from the level recorded illustrating one of the issues of using GSF data for analysis outside of the sample time period. The most populous group is those who have some college education, but not a degree.

Each educational level earns significantly more on average than groups with lower levels of education. In addition, those with more education are more likely to be preparing for retirement, although the difference in saving rates in not statistically significant for those with a college or graduate degree in 2008, and 2009. While the average amount of deferred compensation for those without a high school diploma was $580 in 2009, it was $5,093 for those with a graduate degree. Fewer than 25% of those without a high school diploma are saving in any given year while about half of college graduates are.

Most groups saw relatively low numbers of people saving in 2009 compared to other years the drop in saving rates from 2008 to 2009 was statistically significant for every group except those with no high school diploma. The least educated group saw statistically significant increases in saving rates in 2006, 2007, and 2008, although this increase may be due to the fact that it is contingent on a respondent remaining in the workforce – the number of jobs reported by this group had statistically significant reductions in both 2008 and 2009.

Every education group saw significant reductions in the number of W-2s between 2008 and 2009 and every group except graduate degree holders saw a significant drop between 2007 and 2008. Every group also saw a statistically significant increase in the level of deferred compensation in 2007 and high school graduates, those with some college, and those with college degrees saw a significant drop in 2009. The average amounts saved are higher for those with graduate degrees than those with college degrees in every year, perhaps due to their higher average earnings.

Table 16: Summary Statistics by Education

Education	Year	Number Employed	Number W-2s	Median Rank	Mean Earnings	Median Earnings	% Saving	Average Saving
No High School Diploma	2005	27,751	43,540	38	$23,454	$18,111	18.71%	$435
	2006	27,674	43,957*	39	$24,534***	$18,919	19.94%***	$476**
	2007	27,475	43,297	40	$25,686***	$19,630	22.11%***	$544***
	2008	26,571	39,949***	41	$26,358**	$20,225	24.09%***	$578*
	2009	25,037	34,841***	42	$26,894*	$20,605	24.44%	$580
High School Graduate	2005	49,106	68,385	51	$27,067	$36,452	30.76%	$907
	2006	48,152	67,198	51	$27,415*	$52,560	31.70%**	$930
	2007	47,141	65,047**	52	$27,663	$25,219	33.30%***	$977**
	2008	45,725	61,653***	52	$27,466	$38,892	34.33%***	$982
	2009	43,399	55,725***	51	$27,187	$36,180	33.69%*	$921***
Some College	2005	52,739	74,286	61	$41,118	$34,567	39.22%	$1,534
	2006	52,020	72,938	61	$42,205***	$35,135	40.37%***	$1,608***
	2007	51,192	70,086	62	$43,109***	$35,879	42.23%***	$1,699***
	2008	50,142	68,084***	62	$42,920	$35,707	43.27%***	$1,693
	2009	48,181	62,826***	62	$42,989	$35,960	42.58%*	$1,633**
College Graduate	2005	27,039	35,397	77	$69,123	$51,352	52.10%	$3,399
	2006	26,728	34,919	77	$71,324	$52,049	52.38%	$3,544**
	2007	26,300	34,135	78	$72,575	$52,884	53.37%*	$3,686**
	2008	25,879	33,155**	78	$71,788	$52,771	53.81%	$3,633
	2009	25,203	31,563***	78	$72,197	$53,024	51.94%***	$3,526*
Graduate Degree	2005	12,778	16,329	85	$95,206	$65,936	54.43%	$4,824
	2006	12,543	16,055	85	$97,651	$66,317	53.96%	$5,039*
	2007	12,271	15,627	85	$102,941	$66,910	54.44%	$5,242*
	2008	11,924	15,020	85	$97,300	$66,791	54.24%	$5,124
	2009	11,550	14,294**	85	$95,610	$68,083	52.10%**	$5,093

F-tests were performed on Dummy variable regressions to determine whether values were different from prior-year values. Tested values include number

3.6.5 Exiting Employment

The first workforce variable I examine in a multivariate framework is the chance of exiting employment. For the W-2 population and GSF sample, I examine whether or not a person exited the workforce in a given year dependent on a person's percentile rank in the earnings distribution and also the square of that rank. I also include whether or not a person has deferred compensation and an interaction term to measure interaction the effect of both being in a given rank and having deferred compensation. I include dummy variables to measure the effect of being in the GSF.

The regression in Table 17 shows that the chance of exiting employment drops as a person is higher up in the earnings distribution, a statistically significant result for all years. The quadratic term is positive, indicating that the reduced chance of exiting employment levels off. Those who save are more likely to exit employment, but controlling for being in a given earnings percentile, being a saver makes one less likely to exit employment. The GSF sample behaves differently in some years. In 2005, those in high ranks in the GSF were less likely to exit the workforce than those similarly ranked in the overall population. In 2007 and 2008 those in higher ranks in the GSF were more likely to exit employment with no statistically significant difference for the GSF sample across the earnings distribution in 2006. Savers in the GSF were also more likely to exit than savers in the population of W-2 recipients in all years, although the effect is reduced for higher ranked savers in the GSF.

In table 18 I include the demographic variables from the GSF in a regression on exiting employment to examine which groups were affected. The ommitted dummy variable categories are female, with a high school diploma, white, and born before 1950. Several of the demographic variables are significant. Men are less likely exit employment in every year. Those who are neither white nor black are more likely to exit employment, as are blacks. The likelihood of exiting employment depends significantly on one's level of education. Those with no high school diploma are more likely to exit employment. Those who have attended some college or have a college or graduate degree are less likely to exit employment. All of the younger age groups are less likely to exit employment than the ommitted group of those born before 1950.

3.6.6 Deferred Compensation

I also examine the effect of earnings percentile on the probability of having deferred compensation. In this regression I include earnings percentile and the square of earnings percentile as controls. I also include interaction terms to test for differences between those in the GSF sample and the population of W-2 recipients. The likelihood of having deferred compensation increases with one's relative location in the income distribution. The quadratic term is also positive and significant. Those who are in the GSF sample are less likely to save in a given year overall although the effect is not significant in 2008 and 2009. When controlling for location in the earnings distribution, however, those in the GSF sample are more likely to have deferred compensation. The negative quadratic interaction term indicates that the rate does not increase as quickly for the GSF sample as for the W-2 population as earnings percentile rises.

When I examine the effect of demographic variables on the likelihood of having deferred compensation I find that men are more likely to have deferreed compensation than women. After 2005, those who

Table 17: Exiting Employment - Multivariate Regressions - Earnings Distribution

	(1) 2005	(2) 2006	(3) 2007	(4) 2007
GSF	-0.00982***	-0.0116***	-0.0222***	-0.0286***
	(-6.24)	(-7.40)	(-13.89)	(-16.55)
Earnings Pctl	-0.00803***	-0.00830***	-0.00950***	-0.0109***
	(-2738.49)	(-2850.93)	(-3150.97)	(-3338.51)
Earnings Pctl^2	0.0000594***	0.0000616***	0.0000711***	0.0000799***
	(1884.70)	(1972.40)	(2207.59)	(2301.64)
Any Saving	0.0765***	0.0696***	0.0679***	0.0706***
	(459.95)	(430.19)	(423.34)	(421.12)
Any Saving * Pctl	-0.00114***	-0.00106***	-0.00107***	-0.00113***
	(-466.91)	(-446.69)	(-449.68)	(-450.73)
GSF * Pctl	-0.000206**	-0.000152	0.000224**	0.000425***
	(-2.62)	(-1.94)	(2.81)	(4.91)
(GSF * Pctl)^2	0.00000391***	0.00000327***	0.000000456	-0.00000105
	(4.70)	(3.93)	(0.54)	(-1.14)
GSF * Any Saving	0.0184***	0.0191***	0.0302***	0.0210***
	(4.39)	(4.64)	(7.39)	(4.86)
GSF * Any Saving * Pctl	-0.000232***	-0.000218***	-0.000367***	-0.000246***
	(-3.76)	(-3.59)	(-6.02)	(-3.79)
Constant	0.278***	0.285***	0.320***	0.375***
	(4810.14)	(4960.83)	(5358.68)	(5822.08)
N	156294830	158923137	160671623	160374616

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: Exiting Employment - Multivariate Regressions - Demographic Variables

	(1) 2005	(2) 2006	(3) 2007	(4) 2008
Male	-0.00419***	-0.00256**	-0.00283**	0.00417***
	(-4.86)	(-2.97)	(-3.15)	(4.27)
Black	0.00475***	0.00743***	0.0101***	0.0157***
	(3.59)	(5.61)	(7.28)	(10.49)
Other Race	0.00602**	0.00591**	0.00762***	0.0120***
	(3.00)	(2.94)	(3.63)	(5.25)
No HS Degree	0.0184***	0.0162***	0.0165***	0.0125***
	(13.05)	(11.50)	(11.26)	(7.82)
Some College	-0.00353**	-0.00341**	-0.00975***	-0.0116***
	(-3.04)	(-2.93)	(-8.04)	(-8.77)
College Graduate	-0.0122***	-0.0106***	-0.0184***	-0.0257***
	(-8.09)	(-7.00)	(-11.71)	(-15.05)
Graduate Degree	-0.0137***	-0.0125***	-0.0167***	-0.0280***
	(-6.61)	(-6.04)	(-7.74)	(-11.92)
Born in 1980s	-0.0279***	-0.0316***	-0.0333***	-0.0296***
	(-20.12)	(-22.74)	(-23.01)	(-18.81)
Born in 1970s	-0.0325***	-0.0392***	-0.0432***	-0.0445***
	(-23.35)	(-28.13)	(-29.75)	(-28.19)
Born in 1960s	-0.0343***	-0.0413***	-0.0469***	-0.0470***
	(-25.23)	(-30.39)	(-33.12)	(-30.51)
Born in 1950s	-0.0331***	-0.0369***	-0.0409***	-0.0439***
	(-23.84)	(-26.59)	(-28.28)	(-27.92)
Constant	0.0816***	0.0849***	0.0951***	0.105***
	(70.48)	(73.23)	(78.75)	(79.74)
N	276458	276457	276457	276458

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: Deferred Compensation - Multivariate Regressions - Earnings Distribution

	(1) 2005	(2) 2006	(3) 2007	(4) 2008	(5) 2009
GSF	-0.00706**	-0.00787**	-0.00538*	-0.00381	-0.00375
	(-2.95)	(-3.27)	(-2.22)	(-1.54)	(-1.49)
Earnings Pctl	0.00233***	0.00282***	0.00362***	0.00421***	0.00373***
	(562.18)	(679.35)	(852.53)	(980.67)	(852.10)
Earnings Pctl^2	0.0000595***	0.0000543***	0.0000473***	0.0000419***	0.0000444***
	(1461.18)	(1331.84)	(1138.29)	(995.95)	(1036.13)
GSF * Pctl	0.000845***	0.000896***	0.000869***	0.000712***	0.000707***
	(7.65)	(8.07)	(7.77)	(6.27)	(6.09)
(GSF * Pctl)^2	-0.00000650***	-0.00000684***	-0.00000674***	-0.00000518***	-0.00000520***
	(-6.06)	(-6.34)	(-6.20)	(-4.69)	(-4.61)
Constant	-0.0166***	-0.0193***	-0.0257***	-0.0270***	-0.0167***
	(-189.13)	(-219.19)	(-283.07)	(-293.79)	(-178.07)
N	156294830	158923137	160671623	160374616	155319819

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

are neither white nor black are significantly more likely to save than whites, while blacks are less likely than whites to have deffered compensation in all years. Different educational groups also have significant differences. Those with no high school diploma are less likely to save than high school graduates while those who attended some college, have a college degree, or a graduate degree are all more likely to save than high school graduates. Those born in the 1980s are less likely to save than those born before 1950 in 2005, 2006, and 2007 but more likely to save in 2008 and 2009. Those born in the 1970s, 1960s, and 1950s are more likely to save in all years than those born before 1950.

4 Conclusion

The behavior of people in the GSF sample differs from the overall population in several important ways. They have higher earnings and are more likely to save. When they do save they save lower amounts compared to other savers. If one only compares them to the overall population without controlling for whether a person is saving then they appear to save higher amounts, perhaps due to their higher saving rates. The people in the GSF sample work fewer jobs, but the recession has not caused them to reduce the number of W-2s reported as quickly as those overall population of W-2 recipients. Those in the GSF sample are less likely to exit employment and their chances of exit have increased less than the overall workforce

Table 20: Deferred Compensation - Multivariate Regressions - Demographic Variables

	(1) 2005	(2) 2006	(3) 2007	(4) 2008	(5) 2009
Male	0.0270***	0.0279***	0.0276***	0.0297***	0.0296***
	(14.79)	(15.21)	(14.91)	(15.82)	(15.53)
Black	-0.0480***	-0.0488***	-0.0504***	-0.0476***	-0.0469***
	(-16.89)	(-17.17)	(-17.60)	(-16.38)	(-15.80)
Other Race	0.00794	0.0153***	0.00971*	0.0144**	0.0144**
	(1.85)	(3.54)	(2.24)	(3.28)	(3.21)
No HS Degree	-0.0624***	-0.0587***	-0.0536***	-0.0491***	-0.0468***
	(-20.98)	(-19.60)	(-17.66)	(-15.77)	(-14.64)
Some College	0.0916***	0.0992***	0.108***	0.112***	0.111***
	(38.01)	(40.68)	(43.66)	(44.31)	(43.18)
College Graduate	0.205***	0.206***	0.209***	0.209***	0.198***
	(66.25)	(65.78)	(65.60)	(64.58)	(60.37)
Graduate Degree	0.230***	0.228***	0.232***	0.230***	0.219***
	(54.49)	(53.16)	(52.98)	(51.43)	(48.23)
Born in 1980s	-0.165***	-0.116***	-0.0544***	0.00657*	0.0452***
	(-51.81)	(-36.56)	(-17.29)	(2.09)	(14.19)
Born in 1970s	0.0390***	0.0777***	0.129***	0.167***	0.187***
	(12.71)	(25.04)	(41.26)	(53.00)	(58.74)
Born in 1960s	0.0812***	0.110***	0.147***	0.179***	0.192***
	(27.20)	(36.34)	(48.45)	(58.19)	(61.91)
Born in 1950s	0.108***	0.128***	0.164***	0.194***	0.204***
	(35.70)	(42.02)	(52.98)	(61.99)	(64.14)
Constant	0.242***	0.220***	0.194***	0.169***	0.147***
	(87.94)	(79.96)	(71.40)	(62.42)	(54.52)
N	233999	235993	237549	235771	228274

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

in 2008 after having a greater increase in 2006 and 2007. They are also less likely to enter employment, although their likelihood has not decreased as quickly as the likelihood of the overall W-2 population. It is important to know and understand these differences in the GSF sample when conducting research using the Gold Standard research File. The GSF sample has responded during the recession differently than the overall workforce in some ways.

There are several paths of future research suggested by these results. Having determined some ways in which the Gold-Standard research file sample differs from the population of W-2 recipients it would be valuable to examine the Census Bureau's other data products based on the SIPP. One is the SIPP Synthetic Beta, a public-use file which uses data synthesis methods to maintain the statistical properties of the GSF while preserving confidentiality. Examining this file would be particularly valuable as it would help validate the imputations performed and provide insight as to how well the public use files, which have much lower costs to use for outside researchers, can replicate results from the confidential data file. Another test would be to compare the GSF once the 2008 data is incorporated by adding in the observations from the most recent SIPP panel and examining their behavior separately as well as pooled with the GSF.

Having determined that different portions of the earnings distribution respond differently, I would like to incorporate regressions at different earnings percentiles to examine workforce effects during the Great Recession on those at different relative levels of earnings. This could help to determine how much one's location in the earnings distribution matters in determining employment and deferred compensation effects.

I would also like to make use of the earnings histories to isolate those who experienced earnings shocks to determine whether those who had a negative or positive shock change their saving behavior. I could then investigate whether the GSF sample population was as likely to have an earnings shock and whether they responded in a comparable manner.

5 References

Abowd, John, Gary Benedetto, and Martha Sinson (2007) *The Covariance of Earnings and Hours Revisited*

Autor, David (2010) *U.S. Labor Market Challenges over the Longer Term*

Benedetto, Gary, Graton Gathright and Martha Stinson (2010) *The Earnings Impact of Graduating from College during a Recession*

Elsby, Michael, Bart Hobijn, and Aysegul Sahin (2010) *The Labor Market in the Great Recession*

Hurd, Michael and Susann Rohwedder (2010) *Effects of the Financial Crisis and Great Recession on American Households*

Munnell, Alicia, Dan Muldoon, and Steven Sass (2009) *Recessions and Older Workers*

Rutledge, Matthew (2011) *Long-Run Earnings Volatility and Health Insurance Coverage: Evidence from the SIPP Gold Standard File*

Sahin, Aysegul, Joseph Song, and Bart Hobijn (2009) *The Unemployment Gender Gap During the Current Recession*

Shaefer, Luke (2011) Recessions and the Annual Labor Market Earnings of Vulnerable Workers: New Evidence using Longitudinal Administrative Earnings Data