

# Constructing the Reason-for-Nonparticipation Variable Using the Monthly CPS

Shigeru Fujita\*  
February 6, 2014

## Abstract

This document explains how to construct a variable that summarizes reasons for nonparticipation from the CPS micro data. The variable splits nonparticipants into five categories: (1) retired, (ii) disabled, (iii) want a job (discouraged), (iv) do not want a job: in school, and (v) do not want a job: not in school.

## 1 Introduction

In Fujita (2014), I discussed why the labor force participation rate has been falling in recent years. The summary variable (“reason for nonparticipation”) is constructed from the monthly Current Population Survey (CPS) micro data. This document describes the steps used to construct this variable.

It is important to keep in mind first that the survey itself is not designed for this type of tabulation. However, I believe that the tabulation gives us an accurate view of why individuals are not in the labor force in each month. This document presents various additional calculations to validate my final tabulation.

The first step is to download the [CPS micro data](#). Dictionary files and Stata codes for extracting the data can be found at the [NBER](#). Running the Stata do file will create a dta file with variable names and associated labels in Stata format. For example, running `cpsbmay12.do` will create the Stata dta file for May 2012.<sup>1</sup> One would want to write a separate Stata do file that loops over all monthly files and combine all monthly files into one large Stata file so that one can analyze the time series.

---

\*Shigeru Fujita is a senior economist in the Research Department at the Federal Reserve Bank of Philadelphia. Lance Liu provided excellent research assistance. The views herein do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. Comments and questions may be sent to [shigeru.fujita@phil.frb.org](mailto:shigeru.fujita@phil.frb.org). This document is available free of charge at [www.phil.frb.org/research-and-data/economists/fujita/](http://www.phil.frb.org/research-and-data/economists/fujita/).

<sup>1</sup>For the data between 2000 and 2002, one would want to use the [NBER files](#) for the reasons described on that website.

**Table 1: Labor Force Status Counts, May 2012**

Status	Recode	Unweighted Counts	Weighted (Thousands)	Share
Employed	1, 2	62,049	142,727	0.587
Unemployed	3, 4	4,867	12,271	0.051
NILF –Retired	5	17,899	39,256	0.162
NILF –Disabled	6	5,617	13,365	0.055
NILF –Other	7	13,953	35,346	0.145

Notes: Tabulated from the CPS micro data. Not seasonally adjusted. Sample: working-age population (age 16 and older).

## 2 Three Broad Reasons

Three broad reasons for nonparticipation can be constructed directly from the labor force recode variable pemlr. All specific variable names discussed below are from the May 2012 data dictionary, and thus one may need to find equivalent variables in the data sets for different months.

This variable tells the labor market status of each survey respondent: 1 and 2 mean “employed,” 3 and 4 mean “unemployed,” and 5, 6, and 7 correspond to nonparticipants. The three categories of nonparticipation correspond to the three broad reasons, with 5 being “retired,” 6 being “disabled,” and 7 being “other.” If one is interested in the contributions of retirement and disability to the total nonparticipation rate, then this variable pemlr is all one needs.

Note that this is a recoded variable, meaning that the value assigned to this variable does not come directly from an answer to one particular question. The CPS assigns each value, 1 through 7, to each survey respondent after going through a series of questions relevant to determine the respondent’s labor market status. Again, if one is interested only in the three broad reasons, the tabulation is straightforward. Summarizing this variable should give an accurate assessment of reasons for nonparticipation.

Table 1 just summarizes this variable for May 2012. The first column simply gives the counts for each category. The second column presents population counts using the “composite final weight,” and these numbers match up with those in the official releases.<sup>2</sup> The last column gives each category’s share of the working-age (16+) population. One can use these numbers to obtain the unemployment rate, the labor force participation rate, and the employment-population ratio.<sup>3</sup>

<sup>2</sup>The composite final weight is the population weight used to create the BLS’s published labor force statistics and its variable name is pwcmpwt. One can also use the so-called “final weight” (pwsswt). But my calculations are all based on the former variable.

<sup>3</sup>Note that the raw series are not seasonally adjusted. Because the series for the three broad reasons are not part of the BLS employment situation summary, one needs to adjust seasonality one’s self. The time series data presented below use the Eviews’s built-in default procedure for the seasonal adjustment.

**Table 2: Cross-Tabulation of Labor Force Recode and Desire-for-Work Question**

Do you currently want a job?	Reason for nonparticipation			Total
	Retired	Disabled	Other	
Missing	17,222	5,531	379	23,132
Nonmissing	677	86	13,574	14,337
Total	17,899	5,617	13,953	37,469

### 3 Want a Job vs. Do Not Want a Job

The next step is to split the “other” reasons group into those who want a job and those who do not want a job. This division is based on the variable called `pedwwnto`. According to the CPS interview manual (see page C4-24), those who are not in the labor force are asked whether they currently want a job. The exact question is, “Do you currently want a job, either full time or part time?” The cross-tabulation of `pemlr` and `pedwwnto`, however, reveals that not everybody in the NILF (not in the labor force) category is actually asked this question. Instead, those who answered this question mostly come from those who are classified as “other” in `pemlr`: Table 2 shows that out of all entries (37,469) of nonparticipants, only 14,337 answer this question and that almost all (13,574) valid observations are in the “other” reasons group. Based on this fact, I classify this “other” group into the want-a-job group and do-not-want-a-job group. There are three issues to be resolved at this stage.

1. There are some individuals who are recoded as retired or disabled in `pemlr` but still have valid entries to the want-a-job (desire-for-work) question. Given that my objective is to create a variable that summarizes mutually exclusive reasons for nonparticipation, I have to take a stand on which variable takes precedence in constructing the final variable. As already mentioned, I first split nonparticipants into the three broad groups based on `pemlr` and then use `pedwwnto` to split the “other” group into those who want a job and those who do not. Fortunately, the overlap between those who are “retired”/“disabled” and those who answered the want-a-job question is very small. Moreover, I find that the share of those who have valid entries for `pedwwnto` while recoded as retired or disabled in `pemlr` is roughly constant over time. These facts imply that how I treat these individuals has no material impact on the final result.<sup>4</sup>
2. Within those who are recoded as “other” in `pemlr`, there are a small number of individuals (379) who do not have a valid entry to `pedwwnto`. I treat all of these individuals

<sup>4</sup>An alternative way of creating a nonoverlapping variable is to exclude those who have a valid entry to the want-a-job question from the groups of retired and disabled individuals in `pemlr` and reclassify them based on the answers to the want-a-job question. I confirmed that this alternative classification does not change my results significantly. Allowing for this readjustment would mean that the first three broad reasons for nonparticipation cannot be constructed by the `pemlr` variable alone.

**Table 3: Answers to Want-a-Job Question**

Do you currently want a job?	Reason for nonparticipation			Total
	Retired	Disabled	Other	
Yes, maybe, it depends	249	0	2,303	2,552
No	373	0	11,220	11,593
Retired	5	0	2	57
Disabled	0	67	13	80
Unable	0	19	36	55
Nonmissing total	677	86	13,574	14,337

as those who want a job. Alternatives are to either treat all of them as not wanting a job or to split them into yes and no. Again, given that there are only a small number of them, how I treat them is largely irrelevant in the final result.

3. Lastly, the answers to the want-a-job question include choices other than yes and no. Table 3 presents a more detailed cross-tabulation of `pemlr` and `pedwwnto` among those who had nonmissing entries to the latter variable. As can be seen, individuals are given three more choices (retired, disabled, and unable) other than yes and no. It is not entirely clear why these three choices are given at this state of the survey, but it is possible that these choices (in particular, retired and disabled) are used as inputs into the labor force recode. As mentioned above, regardless of the answer to the want-a-job question, the individuals who are in either retired or disabled in `pemlr` are treated as they are. However, when the individuals in the “other” group pick an answer other than yes or no, then these individuals are treated as those who do not want a job in my final result. Again, there are very few such cases, and thus this treatment has virtually no impact on the final result.<sup>5</sup>

In summary, among the 13,953 unweighted respondents in the “other” group for the May 2012 survey, 2,682 individuals (consisting of 2,303 individuals who actually said yes to the want-a-job question and 379 individuals who did not answer this question) are labeled as those who want a job, and the remaining 11,271 individuals are labeled as those who do not want a job in my final tabulation.

---

<sup>5</sup>Note that there are two cases in which an individual is in the “other” group in `pemlr` but classified as “retired” in `pedwwnto`. Similar internally inconsistent cases (13 cases) exist for those who are classified as “disabled” in `pedwwnto`.

**CPS Table A-38: Persons Not in the Labor Force by Desire and Availability for Work, Not Seasonally Adjusted (Thousands)**

Category	Nov. 2013
Total not in the labor force	91,521
Do not want a job now <sup>(1)</sup>	86,084
Want a job <sup>(1)</sup>	5,437
Did not search for work in previous year	2,905
Searched for work in previous year but not in past 4 weeks <sup>(2)</sup>	2,532
Not available to work now	436
Marginally attached (available to work now) <sup>(3)</sup>	2,096
Discouraged over job prospects <sup>(4)</sup>	762
Reasons other than discouragement	1,334
Family responsibilities	238
In school or training	260
Ill health or disability	118
Other <sup>(5)</sup>	718

(1) Includes some persons who were not asked if they want a job.

(2) Persons who had a job in the prior 12 months must have searched since the end of that job.

(3) Persons “marginally attached to the labor force” are those who want a job, have searched for work during the prior 12 months, and were available to take a job during the reference week, but have not looked for work in the past four weeks.

(4) Discouraged workers are persons marginally attached to the labor force who did not actively look for work in the prior four weeks for reasons such as: they think no work is available, they could not find work, they lack schooling or training, employers think they are too young or old, and other types of discrimination.

(5) Includes those who did not actively look for work in the prior four weeks for such reasons as child-care and transportation problems, as well as a small number whose reasons for nonparticipation were not ascertained.

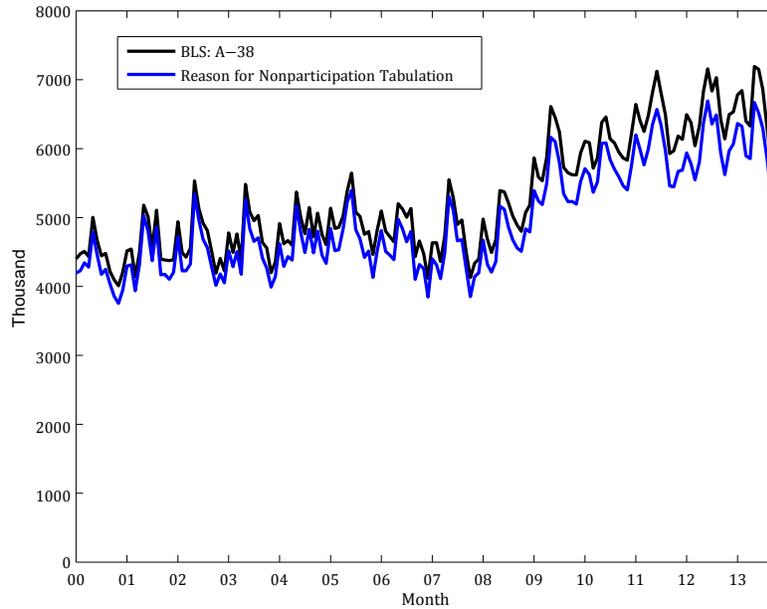
Source: Bureau of Labor Statistics

### 3.1 Relation to Table A-38

My calculation above is related to the BLS tabulated table titled “Persons not in the labor force by desire and availability for work.” This table itself is not included in the monthly employment situation release but is available from the [BLS as Table A-38](#). Note also that some of the series in “alternative measures of labor underutilization” (A-15 in the monthly employment situation by the BLS) come from Table A-38. The above table is reproduced from the November 2013 CPS release.

Note first that in Table A-38, those who do not want a job are defined as a complement of the set of those who want a job within all nonparticipants. This definition is clearly

**Figure 1: Comparison of the Number of Want-a-Job Individuals**



different from my classification. Footnote (1) to the table states that “some persons who are not asked if they want a job” are treated as “Do not want a job now.” Remember that there are a large number of individuals who are classified as retired or disabled in pemlr and did not answer the want-a-job question. Footnote (1) applies mostly to these individuals. As mentioned above, there are a small number of retired/disabled individuals who answered the want-a-job question, and these respondents are classified in my calculation accordingly based on their actual status in pemlr. This creates a discrepancy between the number of those who want a job in my data and the data in Table A-38.<sup>6</sup> Figure 1 compares the two series (not seasonally adjusted) and shows that the overall levels of the two series are close, and the seasonal as well as cyclical patterns of the two series are similar.

In my final tabulation, those who want a job are sometimes labeled as “discouraged workers.” I use this label simply because these individuals are the ones who expressed their desire to work yet were not part of the official unemployment pool (given that they did not undertake job search in the past four weeks). Note that this definition is much broader than the definition of discouraged workers in the BLS data. In particular, the U-4 series within the well-known “alternative measures of labor underutilization” refers to “total unemployed plus discouraged workers.” The discouraged workers in the U-4 series correspond exactly to “Discouraged over job prospects” in Table A-38 above. One can see that this group

<sup>6</sup>The other reason that the two series differ is that, in my calculation, all of those who were not asked the want-a-job question within the “other” group are classified as wanting a job, while only some of these are classified as such in the BLS data. This latter statement must be true, given that footnote (1) applies also to “want a job” in Table A-38.

covers a much narrower group of individuals than those who want a job (i.e., my definition of discouraged workers). There are two reasons for this discrepancy. First, as can be seen in Table A-38, in order for an individual to be part of “discouraged over job prospects,” he/she needs to have undertaken a job search in the past 12 months and is available for work now.<sup>7</sup> Second, discouraged workers in U-4 (i.e., discouraged over job prospects in Table A-38) are restricted to those who choose the five reasons listed under footnote (4) in Table A-38. However, focusing on those who chose these five reasons may not capture the full extent of discouragement. In particular, I find that the number of those who are discouraged over job prospects and the number of those who cite reasons other than discouragement are negatively correlated over the business cycles, indicating the possibility that the two groups switch their answers between economic and noneconomic reasons, depending on the stage of the business cycle. The intention here is *not* to treat individuals in “reasons other than discouragement” as a separate group (in that these individuals may pick economic reasons in the following survey). My approach is to take all of those who expressed the desire to work yet are not part of the unemployment pool as discouraged workers. Again, my definition of discouraged workers is much broader than the one in U-4.

#### 4 Schooling

Next, when individuals are in the do-not-want-a-job group (in my variable definition explained above), I further check if they are enrolled in school or not. Obviously, the purpose here is to identify those who do not want a job now because they are in school. This can be done by using the variable `prnlfsch`.<sup>8</sup> One minor issue with this variable is that the universe of this variable is individuals 16 to 24 years old, which implies that those who are older than 24 but are enrolled in school are not captured by this variable.<sup>9</sup>

Table 5 presents the cross-tabulation of `pemlr` and `prnlfsch`. The sample is restricted to those who are 16 to 24 years old. One important thing to note here is that everyone who is not in the labor force and in this age group has a valid entry to this question (thus the number of missing observations is not reported in this table), ensuring that summarizing this variable gives an accurate picture regarding school enrollment of individuals between 16 and 24 that are out of the labor force. Observe also that out of all 6,686 individuals who had valid entries to this question, almost all of them were in the “other” group in `pemlr`. One can see that there are 29 individuals who are “retired” even though they are younger than 25

---

<sup>7</sup>Those who want a job, searched for a job in the previous year but not in the past four weeks, and are available now are called “marginally attached.”

<sup>8</sup>Note that this is a recoded variable that is supposed to summarize the schooling activity of those who are out of the labor force. The key question behind this variable is in `peschenr`. One can also use this `peschenr` variable to identify those in school, and it gives very similar results as far as schooling activity of nonparticipants is concerned. But using the recoded variable should be more appropriate here.

<sup>9</sup>Starting with the January 2013 survey, the universe of this variable is expanded to individuals 16 to 54 years old. To maintain the continuity, however, my calculation restricts the sample to those 16 to 24 years old, even for the data from January 2013 on. However, the 2013 data allowed me to confirm that omitting those older than age 24 has only a minor impact on the series.

**Table 5: Cross-Tabulation of Broad Reasons for Nonparticipation and School Enrollment (16-24 years old), May 2012**

In school or not in school	pemlr			Total
	Retired	Disabled	Other	
Yes	0	2	5,445	5,447
No	29	239	971	1,239
Total	29	241	6,416	6,686

**Table 6: Cross-Tabulation of School Enrollment and Desire for Work (16-24 years old), May 2012**

Do you currently want a job?	In School	Not in school	Total
Missing	172	47	219
Yes, maybe, it depends	644	235	879
No	4,622	681	5,303
Retired	0	0	0
Disabled	1	2	3
Unable	6	6	12
Total	5,445	971	6,416

Note: Includes those who are in “other” in pemlr.

and are asked the school enrollment question. Similarly, there are 241 individuals who are “disabled” and were asked the school enrollment question. But these individuals were already treated as retired or disabled (based on pemlr) in my calculation, and the important point is that the overlap between retired/disabled in pemlr and prnlfsch is very small, as expected.

Next, Table 6 cross-tabulates the answers to the desire-for-work question and the schooling variable, focusing on those who are 16 to 24 years old and are in the “other” group in pemlr. As can be seen in the previous table, a total of 6,416 individuals 16 to 24 years old are in this category. This is the sample for Table 6.

Within this sample, those who answered “yes, maybe, it depends” are already part of the want-a-job group (discouraged workers in my definition). There are 879 such unweighted observations in the May 2012 survey. The school enrollment status of these individuals is not relevant here, and there is nothing illogical about the result that there are 644 individuals who are in school and want a job (maybe because they are graduating). Also, remember that those who miss the desire-for-work question are all included in the definition of discouraged workers in my calculation. Thus, the relevant sample consists of those who chose “no,” “retired,” “disabled,” or “unable” for the desire-for-work question. But almost all of them are from those who explicitly chose “no” to this question. (Nobody picked “retired,” in contrast).

**Table 7: Nonparticipation by Reason, May 2012, Not Seasonally Adjusted**

Category	Unweighted Counts	Weighted (Thousands)	Weighted Share
Total not in the labor force	37,469	87,968	1.000
Retired	17,899	39,256	0.446
Disabled	5,617	13,365	0.152
Other	13,953	35,346	0.402
Want a job (discouraged)	2,682	6,372	0.072
Do not want a job	11,271	28,974	0.329
In school	4,629	12,401	0.141
Not in school	6,642	16,574	0.188

The unweighted number of those who do not want a job due to school enrollment in May 2012 is therefore 4,629 (= 4,622 + 1 + 6). One can see that if one focuses on this age group, the most important reason for not wanting a job now is schooling (4,629 out of 5,318 relevant observations are in school). Also note from Table 3 that there are 11,271 relevant observations when the sample is expanded to the entire working-age population. Even then, however, schooling is a quantitatively important reason for not wanting a job.<sup>10</sup>

## 5 Summary of the Variable

Table 7 summarizes the final tabulation of “reason for nonparticipation.” The first column gives unweighted counts of the number of individuals for each category, the second column gives the number of individuals in population, calculated using the BLS composite final weight (pwcmpwgt), and the third column gives the share of each category out of the total number of nonparticipants.

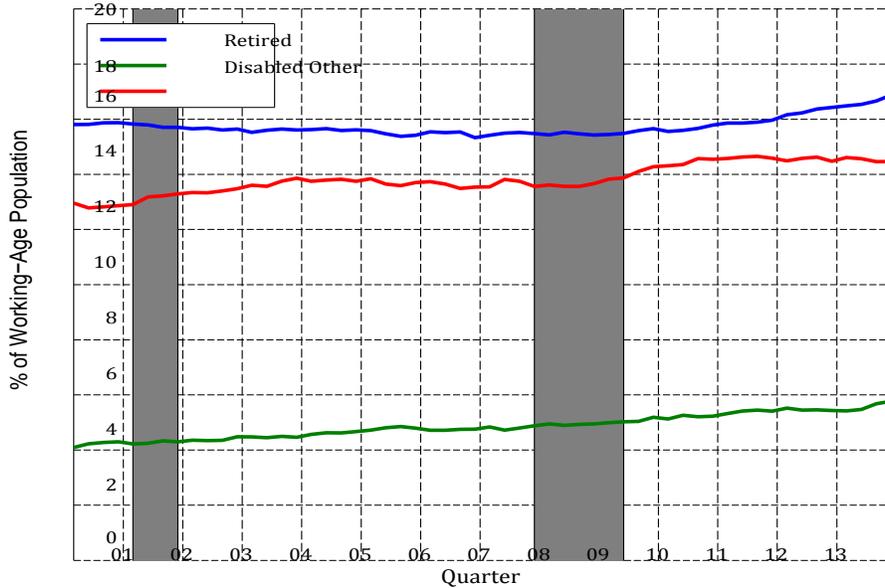
One can see that the largest share of nonparticipants is “retired.” The share of “disabled” is also significant at 15 percent. Nonparticipation due to schooling is also fairly high, even though they are all in the age group of 16-24 years old in my tabulation. More than 50 percent of those who do not want a job are in school. The share of discouraged workers is roughly 7 percent. I want to emphasize again that my definition of discouraged workers is much broader than the BLS definition of discouraged workers in the “alternative measures of labor underutilization.”

## 6 Time Series

Note that the levels of these shares are not necessarily informative about the changes of the participation rate over time. For instance, it is logically possible that nonparticipation

<sup>10</sup>More than 40 percent (4,629/11,271) of those who do not want a job are enrolled in school.

**Figure 2: Nonparticipation by Three Broad Reasons**



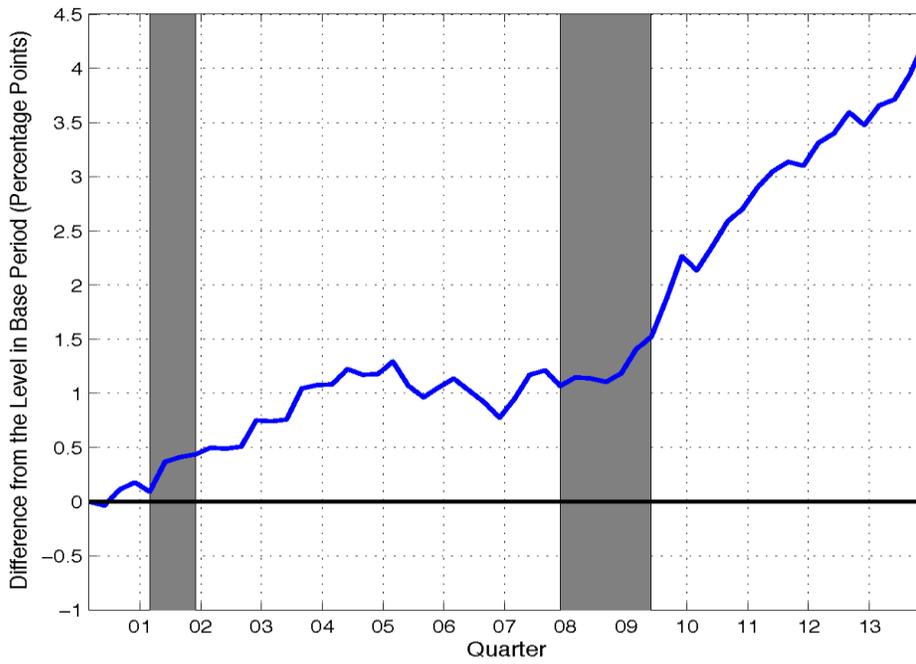
due to discouragement could be the main contributor to the decline in the participation rate over a certain period of time. For example, the share of these workers might have risen dramatically in and after the Great Recession, and it could potentially be the only contributor to the decline in the participation rate.

Figure 2 presents the number of nonparticipants by the three broad reasons, normalized by the working-age (16+) population. Summing up these three lines gives the total nonparticipation rate (which is equal to 100 minus the participation rate).

Figure 3 presents the cumulative differences in the nonparticipation rate since the first quarter of 2000. This figure shows how many percentage points the participation rate has declined at each point in time since 2000Q1. The labor force participation rate has fallen more than 4 percentage points between 2000Q1 and 2013Q4.

Figure 4 shows the contribution of each of the three broad reasons to the overall decline since that base period. Figure 5 further splits the “other” group into those who want a job (and thus are “discouraged”), those who do not want a job and are in school, and those who do not want a job but are not in school. Again, adding up these five lines equals the total decline in the participation rate since the start of the decade. These figures can be used to gauge the contribution of each reason to the changes in the total participation rate. Choosing the different base period obviously would result in different conclusions.

**Figure 3: Cumulative Differences in Nonparticipation Since 2000Q1**



**Figure 4: Cumulative Differences Since 2000Q1: Three Broad Categories**

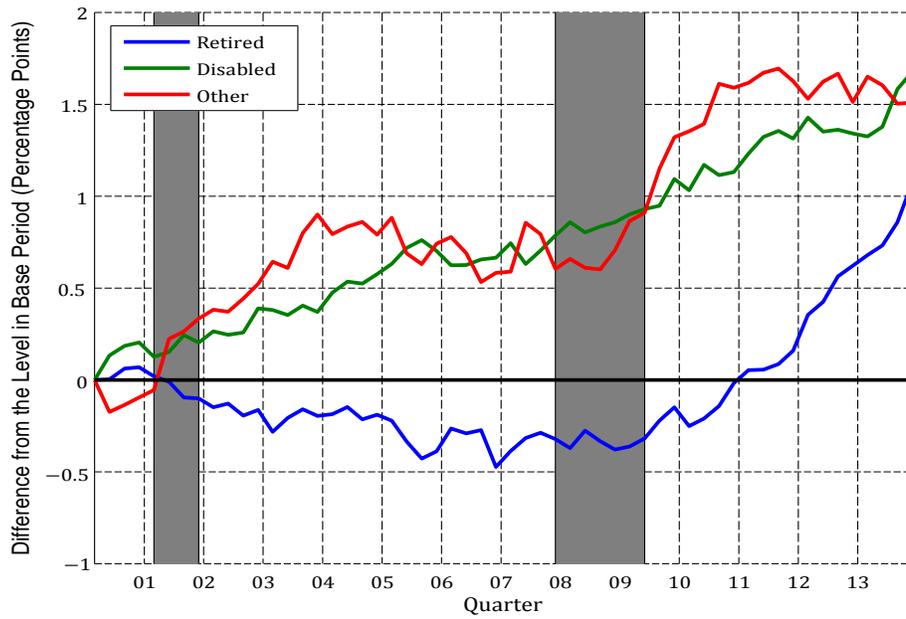


Figure 5: Cumulative Differences Since 2000Q1: Five Categories

