Variable and Weight Description for the Yale Labor Survey

This version: December 14, 2020

This document describes the main labor market series that are generated by the Yale Labor Survey (YLS). A paper describing the methodology of the YLS is "The US Employment Situation Using the Yale Labor Survey," published in July 2020 by Yale's Cowles Foundation.[1]

The first section of this document provides some basic information about the YLS. The second section describes the variables, and the third section describes an updated system of sample-weight construction that was adopted in early December 2020.

1 Basic Information

- All YLS variables are not seasonally adjusted. Comparison data from the Current Population Survey (CPS) and the Census Bureau's Household Pulse Survey are also not seasonally adjusted.
- All YLS results refer to the US population aged 20 and older, as do comparison data from the CPS and Household Pulse Survey.
- Variables with the `_adjonline` suffix are adjusted by –1.4 percentage point to reflect the fact that some people answered the YLS thinking that filling out online surveys alone qualifies as working for pay. This adjustment was quantified using questions that were added to the YLS in early July. A 1.4-percentage-point subtraction was then imposed on the work-for-pay and employment series throughout the sample period. The YLS unemployment rate and labor force participation rate were not adjusted, due to the difficulty of quantifying the appropriate adjustments for those series.
- YLS and comparison data are published in a Microsoft Excel file that includes one or more individual spreadsheets.
  - The name of the Excel file is `YLS_comparisons_[ddMonYYYY].xlsx`, where [ddMonYYYY] is the date that the Excel file was produced. For example, the file created on August 18, 2020 is `YLS_comparisons_18Aug2020.xlsx`.
  - Individual spreadsheets within the Excel file include YLS variables constructed with a particular weight. For example, YLS variables that use the `wt_empr_feb_20` weight are listed in the spreadsheet titled `weight_empr_feb_20`. The importance of sample weights for YLS results is discussed at length in the methodology paper published by the Cowles Foundation.

[1] The paper is Cowles Foundation Discussion Paper No. 2243 and is available at https://cowles.yale.edu/sites/default/files/files/pub/d22/d2243.pdf. The paper's authors are Christopher L. Foote, William D. Nordhaus, and Douglas Rivers.

– When the Excel file is immediately opened, data are formatted as rounded to the nearest whole number. This is purely a formatting issue; YLS data are stored to a much higher precision (typically several places after the decimal point).

2 Variable Descriptions

- `refweek_date`. This "reference week date" indicates the end of the relevant reference week (or reference period). For YLS and CPS data, this date always refers to a Saturday, indicating that the data cover the Sunday-to-Saturday week than ends on that day. Importantly, these are *reference weeks* and not *survey weeks*, with the latter designation indicating the week in which the data were collected. Respondents to both the CPS and the YLS are asked about their labor market activity *last week*, so survey weeks are always one week later than corresponding reference weeks. Note that the only reference weeks for which CPS data are available are those weeks containing the 12th of the month. In the Household Pulse Survey, respondents were asked whether they worked for pay or profit *in the last seven days* rather than last week. Consequently, the `refweek_date` for each Household Pulse data point is the last day of the survey period.[2]

2.1 Work-for-Pay Ratio

The following data refer to the measure of persons working for pay or profit during the reference week (or reference period). These "persons at work" are categorized in the CPS and YPS as employed, along with persons who are temporarily absent from their jobs for reasons such as illness, vacation, etc.

- `WFPonlyYLS_adjonline`. This YLS variable is the 20+ population share calculated only with the YLS's `workforpay` question. As noted in the methodology paper (and the next bullet point), the YLS's baseline measure of work-for-pay also includes people who answered a separate question `worksituation_working` to indicate that they worked either at their usual place of work or at a different place (for example, at home). Typically, the work-for-pay rate is increased by about two percentage points by adding respondents who answer "no" to the `workforpay` question, but who indicate in `worksituation_working` that they did in fact work. As noted above, the `_adjonline` suffix on this and other variables indicates that it reflects an adjustment of –1.4 percentage points. This adjustment accounts for the fact that some YLS respondents thought that filling out online surveys alone was working for pay.

- `WFPbaseYLS_adjonline`. **This is the baseline measure of the work-for-pay rate in the YLS.** It includes people who answered "yes" to the `workforpay` question as

[2]For example, because Week 12 of the Household Pulse Survey covered July 16 to July 21, the `refweek_date` for the corresponding data point is 7/21/2020.

well as anyone who indicated in the `worksituation_working` question that he or she worked for pay.

- `WFP_uperror_adjonline`. This is the upper bound of the 95% confidence interval for the baseline work-for-pay rate `WFPbaseYLS_adjonline`. It is calculated by adding 1.96 times the appropriate standard error to `WFPbaseYLS_adjonline`.

- `WFP_loerror_adjonline`. The lower bound of the 95% confidence interval for `WFPbaseYLS_adjonline`.

- `WFP20microCPS`. The work-for-pay rate for the 20+ population in the CPS, calculated with CPS microdata.

- `WFP20Xpub`. The published work-for-pay rate for the 20+ population *excluding agricultural industries*. This information is published along with other labor market data early each month, when the Bureau of Labor Statistics (BLS) releases the Employment Situation report.

- `WFPpulse`. The share of the 20+ population that worked within the past seven days, as calculated with microdata from the Household Pulse Survey.

- `WFPpulse18pub`. The share of the 18+ population that worked within the past seven days, according to published data from Household Pulse. During Phase 1 of the Household Pulse project (late April–late July), the 20+ work-for-pay ratio was on average 0.15 percentage point higher than the corresponding 18+ rate. Because this difference is small, and because the published 18+ data are available sooner than the microdata needed to calculate the 20+ rate, the 18+ rate is included in YLS charts.

2.2 Employment-to-Population Ratio (EPR)

- `EPRyls_adjonline`. The baseline employment-to-population ratio in the YLS. It includes respondents who are working for pay and respondents who indicate that they are on *paid* absences through their answers to the `worksituation_working` question.

- `EPR_uperror_adjonline`. The upper bound of the 95% confidence interval for `EPRyls_adjonline`.

- `EPR_loerror_adjonline`. The lower bound of the 95% confidence interval for `EPRyls_adjonline`.

- `EPR20pub`. The employment-to-population ratio for the 20+ population from the CPS, as published by the BLS.

- `EPRalt20microCPS`. The 20+ employment-to-population ratio from CPS microdata, after removing "other-reasons absences" from the employed population. **In YLS graphs, this variable is called "EPR-alt."** As discussed in the methodology paper, many CPS respondents who were displaced by the pandemic were incorrectly recorded in the CPS as employed but absent from work. In line with a method suggested by the BLS, this variable excludes employed-but-absent workers who report that they missed work for "other reasons," that is, reasons besides the usual ones such as vacation, illness, bad weather, etc. See the methodology paper for details.

2.3 Unemployment Rate (UR)

- `URyls`. The baseline unemployment rate in the YLS. As in the CPS, the unemployment rate is defined as the share of unemployed persons in the labor force, where the labor force is defined as the unemployed plus the employed.
- `UR_uperror`. The upper bound of the 95% confidence interval for `URyls`.
- `UR_loerror`. The lower bound of the 95% confidence interval for `URyls`.
- `UR20pub`. The 20+ unemployment rate from the CPS, as published by the BLS.
- `URalt20microCPS`. The 20+ unemployment rate calculated from CPS microdata, after *adding* other-reasons absences to the stock of unemployed workers. **In YLS graphs, this variable is called "U3-alt."** The U3 designation refers to the official name of the CPS's headline unemployment rate series (U3).

2.4 Labor Force Participation Rate (LFPRs)

- `LFPRyls`. The baseline labor force participation rate in the YLS.
- `LFPR_uperror`. The upper bound of the 95% confidence interval for `LFPRyls`.
- `LFPR_loerror`. The lower bound of the 95% confidence interval for `LFPRyls`.
- `LFPR20pub`. The 20+ participation rate from the CPS, as published by the BLS. Note that there is no "alt" version of the LFPR, because recent misclassification errors involving employment and unemployment do not affect the size of the labor force. (Employment and unemployment are both components of the labor force.)

3 December 2020 Update for Construction of Sample Weights

As described in the background documentation referenced at the start of this document, a novel feature of the YLS is that its sample weights balance the sample with respect to the labor force attachment of the population as well as demographic characteristics. This balance

is required because the YLS sample is self-selected—only persons who have previously agreed to answer YouGov questionnaires can be included in the YLS. As a result, the labor force attachment of YLS respondents differs from that of the overall US population. In particular, YLS respondents are less likely to be employed.

To bring the labor market attachment of YLS respondents into balance with that of the overall population, sample weights in the YLS incorporate information regarding YLS respondents' previous labor market statuses (that is, employment, unemployment, and participation) in addition to demographic information (such as race and gender). Before the December 2020 update, the weights used labor market information only from February. As a result, in all weekly YLS surveys, the (weighted) February labor market status of YLS respondents matched the official percentages of persons employed, unemployed, and out-of-the-labor force in the February CPS.

As time progressed, the February-based weights did less well at bringing current YLS estimates of employment, unemployment, and participation into line with new CPS data. Accordingly, in December 2020 the YLS introduced a new weighting system, which updates the months used for weight-construction on a rolling basis. The table on the next page shows how updated information is incorporated into the new scheme. Starting in early July, the YLS added questions on labor market status in June as well as February. It then became possible to construct a June-based weight as well as a February-based weight. Questions were added in future surveys so that weights based on other months could be constructed as well. The new system constructs "final weights" that are averages of these month-specific weights. As an example, in the new weighting system the final weight for the survey for July 11 is a weighted average of the February-based weight and the June-based weight. The table shows that the percentages used to construct the weighted average are 75% for the February weight and 25% for the June weight. For the next week, July 18, a weighted average of the February and June weights is also used, but the relative percentages change to 50 percent each. Introducing information from new months on a gradual basis reduces the possibility that shifting from one weight to the next results in a spurious shift in a YLS series.

Research has indicated that the new weighting system causes the YLS survey to better approximate the CPS for employment-related measures (the work-for-pay rate and the employment-to-population ratio). Yet the new strategy still generates in sizeable gaps in rates of unemployment and participation. These gaps remain an active area of research.

Reference Week Date	Feb-based weight	Jun-based weight	Jul-based weight	Aug-based weight	Sep-based weight	Oct-based weight	Nov-based weight
11-Apr-20	1.00						
18-Apr-20	1.00						
25-Apr-20	1.00						
2-May-20	1.00						
9-May-20	1.00						
16-May-20	1.00						
23-May-20	1.00						
30-May-20	1.00						
6-Jun-20	1.00						
13-Jun-20	1.00						
20-Jun-20	1.00						
27-Jun-20	1.00						
4-Jul-20	1.00						
11-Jul-20	0.75	0.25					
18-Jul-20	0.50	0.50					
25-Jul-20	0.25	0.75					
1-Aug-20		1.00					
8-Aug-20		0.75	0.25				
15-Aug-20		0.50	0.50				
22-Aug-20		0.25	0.75				
29-Aug-20			1.00				
5-Sep-20			0.75	0.25			
12-Sep-20			0.50	0.50			
19-Sep-20			0.25	0.75			
26-Sep-20				1.00			
3-Oct-20				0.75	0.25		
10-Oct-20				0.50	0.50		
17-Oct-20				0.25	0.75		
24-Oct-20					1.00		
31-Oct-20					1.00		
7-Nov-20					0.75	0.25	
14-Nov-20					0.50	0.50	
21-Nov-20					0.25	0.75	
28-Nov-20						1.00	
5-Dec-20						0.75	0.25
12-Dec-20						0.50	0.50
19-Dec-20						0.25	0.75
26-Dec-20							1.00

Table 1. Importance of Month-Specific Weights for Final Weights in December 2020 Update. This table shows how updated information on labor market status in prior months is incorporated into the "final weights" used in the December 2020 weighting update. The column for the February-based weight refers to the weight that balances the YLS respondents' labor market attachment in February to official CPS data for February. The other columns refer to weights that balance labor market status in June through November. The final weight for any given YLS week is a weighted average of two month-specific weights, with the importance of each month in this average reported in the appropriate row.