

UI Imputation

Introduction

Unemployment Insurance (UI) is intended to provide a safety net for workers who have been displaced from work due to no fault of their own. These benefits are primarily distributed through the State UI program, Unemployment Compensation for Federal Employees (UCFE) program, or the Unemployment Compensation for Ex-servicemembers (UCX) program. We also include the Workshare (STC) UI program, which allows reduced-hour employees to receive a fraction of the State UI program benefits while working. We do not consider the Extended Benefits (EB) program, which extends the length of coverage of the first three programs, because its program totals (both recipient and benefit) are less than 0.1% of the regular and Workshare UI totals, and these benefits were mostly phased out by 2014.

The Current Population Survey (CPS) provides UI micro-data in its March Supplement with its "UC_YN" and "UC_VAL" variables: these variables are the total number of UI (all programs) recipients as well as their overall weighted-sum benefits respectively. Both UI benefits and UI recipients were underreported in the 2015 CPS compared to the administrative totals from US Department of Labor (DOL). More specifically, CPS unemployment benefits were underreported by 13 billion, with individuals reporting 22.6 billion in claimed benefits, and recipients were underreported by 5.3 million, with 4.6 million individuals receiving UI benefits. CPS underreporting typically occurs with government benefit programs.

We augmented the total number of recipients and their dollar amount of benefits from CPS 2015 March Supplement to match the 2014 UI state-by-state reports provided by DOL. We used the 2015 CPS because the UC_YN and UC_VAL variables report the respondents' unemployment compensation and recipiency for the calendar year 2014, rather than 2015.

In addition to matching UI recipient numbers and total benefits, we tried to maintain the original CPS micro-data distribution by utilizing DOL current distribution and UI eligibility rules. However, since CPS data is insufficient in many ways (and partially DOL data), a number of assumptions are made in order to augment the data reasonably. This report details those assumptions and explains our imputation procedure.

Since DOL data on UI is measured by first payments, last payments, and continued weeks claimed, we made assumptions about those receiving any UI benefits in 2014. Specifically, since the 2015 CPS reports the total amount of those receiving any UI benefits at any time in 2014, and since the UI programs benefit duration can last up to 26 weeks, we needed to include not only first payments, but also continued weeks claimed from 2013 that spill into 2014. To do this, we considered only the continued weeks that were claimed on the first week of January using DOL's ETA 539 data. The ETA 539's number of continued weeks claimed the first week of January represents the number of people receiving spill-over benefits from 2013 into 2014. Thus, we add this number to the number of first

payments in 2014 from the ETA 5159, to represent the total number of individuals receiving UI compensation at any point in 2014.

CPS micro-data and UI targets

In the Source of Income – Unemployment Compensation section of CPS March Supplement, the CPS contains the total unemployment benefit amount for each respondent in the "UC_VAL" variable and whether or not each respondent received any unemployment compensation in its "UC_YN" variable.

Targets for imputation

The targets for imputation come from DOL official data. The State UI, UCFE, UCX and Workshare data (both recipient and benefit amounts) come from the DOL's "[ETA 5159](#)" report. The UI spillover recipients from 2013 come from DOL's "[ETA 539](#)" report, since this report contains weekly UI administrative data. We use first payments to calculate recipients, and then we add on the continued recipients from what we calculated above. The benefits paid come from the continued weeks claimed, amount section. This report contains the target aggregate administrative recipient amounts for all these programs by state, and their corresponding outlays, which serve as targets for UC_YN, and UC_VAL respectively.

We use calendar year administrative totals because the CPS UI compensation is reported according to calendar year. In the calendar year 2014, administrative data suggests that approximately 10 million individuals claimed roughly 35.7 billion dollars in UI benefits. The annual average combined benefit for each recipient is approximately \$3,410 with significant variation across states (Table 2). Thus, for our imputation we partition the CPS March Supplement data by state for better measurements.

Below we provide a summary of CPS UI outlays, and the corresponding DOL's administrative target outlays for each state before our imputation. See Appendix 1 for total recipients by state.

State	CPS total benefits (annually)	Admin total benefits (annually)
Alabama	171383501	245421919
Alaska	89125636	146016667
Arizona	251467393	345259585
Arkansas	99541699	257556579
California	3127716603	6116890032
Colorado	478016831	534094365
Connecticut	340396135	795497360
Delaware	46138337	88302050
District of Columbia	59327462	141909671
Florida	587008596	813042057
Georgia	337220885	529038155
Hawaii	118805756	197493199
Idaho	135421700	124983302
Illinois	1430732836	2086738785
Indiana	736729267	436288948
Iowa	124647303	408334806
Kansas	117350049	304176059
Kentucky	185701489	378758469

Louisiana	154544223	172393444
Maine	85577862	144246245
Maryland	765133694	681425221
Massachusetts	620231421	1709641940
Michigan	740373160	990796645
Minnesota	638097774	823387177
Mississippi	90758957	147608399
Missouri	249949708	404770589
Montana	80612429	110738135
Nebraska	44279082	101882012
Nevada	236809146	386389895
New Hampshire	114286779	83149353
New Jersey	1094387027	2320488278
New Mexico	71658767	211422452
New York	1520839098	2680416901
North Carolina	359666349	456524204
North Dakota	38958947	91119864
Ohio	1196167972	1055418694
Oklahoma	137982083	232906702
Oregon	222943813	577828157
Pennsylvania	1699397447	2423256821
Rhode Island	136959950	187686838
South Carolina	272321746	178014215
South Dakota	32425292	28053864
Tennessee	259315419	325473455
Texas	1831168125	2311647908
Utah	94486444	194299089
Vermont	44203868	81156651
Virginia	198727906	514513150
Washington	674551487	1094705163
West Virginia	118625464	210988260
Wisconsin	369535718	703948490
Wyoming	45189548	65768702

Imputation Procedure

We follow a similar two-step procedure as the SSI imputation for augmentation. First, we add up the individual weights for each CPS respondent in the recipient pool to see if we reach the administrative level for each state. If we don't, then we augment by including the most likely recipients from the non-recipient pool, using probabilities derived from both a logistic regression and a Random Forest classifier model, until we match administrative totals for each state. Second, we obtain an adjust ratio for the benefit amount that allows us to match the administrative dollar benefit totals for each state.

Step I: Recipient Imputation w/ Probabilities From Two Models

Model I:

We propose a basic logistic regression model for analyzing the likelihood of being a UI recipient. Following UI eligibility rules, we use the available corresponding CPS information to create independent variables for whether or not individuals are currently employed, how many weeks they spent looking for work, whether they were unemployed due to their own fault, how long were the stretches that they were looking for work, their income amounts (being above a threshold that yielded

the highest r-squared), whether or not they were disabled (must be able to work), and the market value of food stamps received. These are all statistically significant independent variables for determining UI eligibility, and important variables listed in the Random Forest's feature importances.¹

UI eligibility rules that we include:

- + According to DOL, individuals must establish a significant "base period" of earnings to be eligible for UI compensation; thus our income threshold identifies individuals who don't make enough for this qualification. We use *ptotval* as a proxy for income as it was the most statistically significant income variable in the regression.
- + Eligible UI individuals must be actively searching for a job. This aspect is captured by both the stretch of job searching and the number of total weeks spent searching.
- + If individuals lose their job due to their own fault, then they are not eligible for UI compensation. This is captured in the CPS "pruntype" variable.
- + Disability disqualifies individuals from UI compensation
- + Market value of food stamps is included since it was a significant feature importance for our random forests, and upon implementation, statistically significant in the logistic regression.

Below we give our proposed logistic regression model for predicting the likelihood of receiving UI compensation.

$$UC_YN = \alpha + weuemp * \beta_1 + ptotval * \beta_2 + pruntype * \beta_3 + a_explf * \beta_4 + lkweeks * \beta_5 + lkstrch * \beta_6 + f_mv_fs * \beta_7 + disability * \beta_8 + \epsilon$$

This yielded a pseudo r-squared value of .3.

We then run the model on the CPS dataset. After, we use the fitted coefficients to produce a vector of probabilities for UI recipients. We then rank all recipients according to their fitted probability. For each state sub-group, we aggregate the recipient weights, and add extra non-recipients by likelihood until the weights reach administrative level.

Model II:

We use a Random Forest Classifier (RFC) model to determine UI recipient likelihood. Random Forests performed much better than the logistic regression model, with an accuracy/score of .98, compared to the logistic regression's pseudo r-squared of 0.3.

To train the RFC model we used all of the CPS variables except those that approximately identified those receiving UI benefits. To create feasible variables for the training, we converted all variables

1 The code and results for the feature importances can be found in *Rf_probs.py*

containing categorical strings into numerical categorical variables, and created proxy variables for many columns with missing data (Not in Universe, None, etc.).

After training the Random Forest on a training set (80% of the data), we computed the probability that each CPS respondent received UI compensation. Then, we ranked the probabilities as we did above, and imputed recipients until the recipient state totals matched the administrative state totals ².

Step II: Benefit imputation

For each imputed/augmented recipient, we assign the average benefit amount for the corresponding state. We then calculate the new total outlays for each state, and compare these outlays with DOL administrative state outlays. We calculate the adjustment ratios for each state by dividing administrative outlays by the new outlays. Most adjustment ratios close to 1, but some are significantly larger. We use these adjustment ratios to augment individual's benefits to match the state administrative totals.

² The code and score results for the Random Forest Classifier model can be found in C-TAM's github documentation

Appendix

Table 1: Annual UI recipient numbers by state for CPS and administration before augmentation

State	CPS total recipients	Admin total recipients
Alabama	45706	116799
Alaska	23119	51183
Arizona	78377	135643
Arkansas	37349	98513
California	645270	1652222
Colorado	75168	139813
Connecticut	86005	199901
Delaware	12110	28549
District of Columbia	11060	38160
Florida	154408	341525
Georgia	89764	271273
Hawaii	21537	39330
Idaho	33927	55603
Illinois	241127	535903
Indiana	84859	168305
Iowa	43086	134286
Kansas	28637	91902
Kentucky	36784	98841
Louisiana	31452	79686
Maine	22767	51447
Maryland	68088	188452
Massachusetts	80692	321769
Michigan	180590	400007
Minnesota	108347	211250
Mississippi	27925	74202
Missouri	79140	191873
Montana	19337	40375
Nebraska	14936	42583
Nevada	49376	113168
New Hampshire	26376	31452
New Jersey	186470	476309
New Mexico	16120	54212
New York	310449	775502
North Carolina	100746	219841
North Dakota	11106	24425
Ohio	187176	321693
Oklahoma	35649	65475
Oregon	80023	162572
Pennsylvania	354761	618861
Rhode Island	26515	50080
South Carolina	44081	99627
South Dakota	9466	10601
Tennessee	59888	148008
Texas	287193	598590
Utah	27510	62151
Vermont	13763	28122
Virginia	64415	156310
Washington	151500	275864
West Virginia	33011	69914
Wisconsin	110959	272104
Wyoming	8764	18863

Table 2: Average UI outlays by state

State	Average Benefit amount
Alabama	2341.1866
Alaska	3339.6612
Arizona	2618.3799
Arkansas	2872.816
California	3702.8725
Colorado	3883.5035
Connecticut	4307.6696
Delaware	3313.0248
District of Columbia	4823.5782
Florida	2428.9271
Georgia	2176.064
Hawaii	5200.0631
Idaho	2449.2602
Illinois	4097.6062
Indiana	2790.3027
Iowa	3484.2934
Kansas	3463.7885
Kentucky	3862.2024
Louisiana	2259.1199
Maine	2929.1551
Maryland	3968.0495
Massachusetts	5489.686
Michigan	2567.8551
Minnesota	4163.3994
Mississippi	2104.6024
Missouri	2251.7653
Montana	2956.8805
Nebraska	2626.5696
Nevada	3608.6922
New Hampshire	2812.6155
New Jersey	5035.4977
New Mexico	4117.6031
New York	3661.3115
North Carolina	2197.8711
North Dakota	4091.9644
Ohio	3582.1226
Oklahoma	3807.7182
Oregon	3852.4702
Pennsylvania	4089.8642
Rhode Island	3834.4128
South Carolina	2032.8219
South Dakota	2830.2929
Tennessee	2344.1158
Texas	4045.1613
Utah	3390.909
Vermont	3157.9692
Virginia	3479.4023
Washington	4235.7538
West Virginia	3184.3439
Wisconsin	2728.1441
Wyoming	4059.5458

Table 3: Adjustment ratios of outlays by state

State	Imputed	Admin	adjust ratio
Alabama	309609312	245421919	0.7926
Alaska	157326561	146016667	0.9281
Arizona	391208129	345259585	0.8825
Arkansas	248725971	257556579	1.0355
California	6851211399	6116890032	0.8928
Colorado	722213279	534094365	0.7395
Connecticut	763753429	795497360	1.0415
Delaware	94822540	88302050	0.9312
District of Columbia	147132214	141909671	0.9645
Florida	1021483830	813042057	0.7959
Georgia	670529848	529038155	0.7889
Hawaii	202421002	197493199	0.9756
Idaho	176702682	124983302	0.7073
Illinois	2531235128	2086738785	0.8243
Indiana	934613209	436288948	0.4668
Iowa	384165977	408334806	1.0629
Kansas	320839796	304176059	0.948
Kentucky	420502848	378758469	0.9007
Louisiana	256025041	172393444	0.6733
Maine	162158923	144246245	0.8895
Maryland	1176714351	681425221	0.579
Massachusetts	1886617165	1709641940	0.9061
Michigan	1268719875	990796645	0.7809
Minnesota	1011530344	823387177	0.814
Mississippi	179236421	147608399	0.8235
Missouri	474993293	404770589	0.8521
Montana	133981786	110738135	0.8265
Nebraska	107074898	101882012	0.9515
Nevada	445640302	386389895	0.867
New Hampshire	123652789	83149353	0.6724
New Jersey	2468953170	2320488278	0.9398
New Mexico	216673007	211422452	0.9757
New York	3067750309	2680416901	0.8737
North Carolina	594089017	456524204	0.7684
North Dakota	83619587	91119864	1.0896
Ohio	1581350892	1055418694	0.6674
Oklahoma	234855422	232906702	0.9917
Oregon	494184875	577828157	1.1692
Pennsylvania	2669539451	2423256821	0.9077
Rhode Island	224020907	187686838	0.8378
South Carolina	361895238	178014215	0.4918
South Dakota	33123327	28053864	0.8469
Tennessee	440881252	325473455	0.7382
Texas	2979815826	2311647908	0.7757
Utah	194460614	194299089	0.9991
Vermont	81519980	81156651	0.9955
Virginia	489735789	514513150	1.0505
Washington	1130544573	1094705163	0.9682
West Virginia	223490813	210988260	0.944
Wisconsin	769973173	703948490	0.9142
Wyoming	75862705	65768702	0.8669

Table 4: Administrative and CPS totals after augmentation

State	post augment CPS total benefits	post augment CPS total recipients	Admin total benefits (annual)	Admin total recipients
-------	---------------------------------------	---	-------------------------------------	---------------------------

	(annual)			
Alabama	245421919	117214	245421919	116799
Alaska	146016667	51153	146016667	51183
Arizona	345259585	134341	345259585	135643
Arkansas	257556578	98590	257556579	98513
California	6116890031	1650840	6116890032	1652222
Colorado	534094365	138048	534094365	139813
Connecticut	795497359	199708	795497360	199901
Delaware	88302049	28652	88302050	28549
District of Columbia	141909671	38060	141909671	38160
Florida	813042057	341992	813042057	341525
Georgia	529038154	272395	529038155	271273
Hawaii	197493199	39288	197493199	39330
Idaho	124983301	55669	124983302	55603
Illinois	2086738785	537177	2086738785	535903
Indiana	436288948	167738	436288948	168305
Iowa	408334805	133694	408334806	134286
Kansas	304176059	92637	304176059	91902
Kentucky	378758469	98990	378758469	98841
Louisiana	172393444	78719	172393444	79686
Maine	144246245	51614	144246245	51447
Maryland	681425221	188861	681425221	188452
Massachusetts	1709641940	321282	1709641940	321769
Michigan	990796645	399662	990796645	400007
Minnesota	823387177	212026	823387177	211250
Mississippi	147608398	74137	147608399	74202
Missouri	404770588	191562	404770589	191873
Montana	110738135	40173	110738135	40375
Nebraska	101882011	42576	101882012	42583
Nevada	386389895	113614	386389895	113168
New Hampshire	83149352	31429	83149353	31452
New Jersey	2320488278	476905	2320488278	476309
New Mexico	211422452	54213	211422452	54212
New York	2680416901	775467	2680416901	775502
North Carolina	456524204	220200	456524204	219841
North Dakota	91119864	24571	91119864	24425
Ohio	1055418694	320535	1055418694	321693
Oklahoma	232906702	65558	232906702	65475
Oregon	577828156	162700	577828157	162572
Pennsylvania	2423256820	619899	2423256821	618861
Rhode Island	187686838	49703	187686838	50080
South Carolina	178014215	99346	178014215	99627
South Dakota	28053864	10858	28053864	10601
Tennessee	325473455	146692	325473455	148008
Texas	2311647907	598127	2311647908	598590
Utah	194299089	62292	194299089	62151
Vermont	81156650	28229	81156651	28122
Virginia	514513150	156950	514513150	156310
Washington	1094705163	275503	1094705163	275864
West Virginia	210988259	69577	210988260	69914
Wisconsin	703948490	271908	703948490	272104
Wyoming	65768702	18848	65768702	18863