INFO 7470
Statistical Tools: Edit and Imputation
Examples of Multiple Imputation

John M. Abowd and Lars Vilhuber
April 18, 2016

Applications to Complicated Data

- Computational formulas for MI data
- Examples of building Multiply-imputed data files
  - Survey of Consumer Finances
  - SIPP "gold standard" (harmonized files underlying the SIPP Synthetic Data)
  - Quarterly Workforce Indicators

Computational Formulas

- Assume that you want to estimate something as a function of the data Q(Y)

- Formulas account for missing data contribution to variance

$Q_m(Y^m) = $ estimand from the mth implicate

$$\overline{Q} = \sum_{m=1}^{M} Q_m(Y^m) \Big/ M$$

$\overline{Q} = $ average estimand

$V_m(Y^m) = $ covariance matrix of $Q_m(Y^m)$ from the mth implicate

$$\overline{V} = \sum_{m=1}^{M} V_m(Y^m) \Big/ M$$

$\overline{V} = $ average covariance matrix

$$B = \left[\sum_{m=1}^{M} \left(Q_m(Y^m) - \overline{Q} \right) \left(Q_m(Y^m) - \overline{Q} \right)^T \right] \Big/ M$$

$B = $ between implicate variation of $Q_m(Y^m)$

$$T = \overline{V} + \left(1 + \frac{1}{M} \right) B$$

$T = $ total variance matrix of $Q(Y)$

$$\text{Missingness Ratio} = \left(1 + \frac{1}{M} \right) b_{ii} / t_{ii}$$

Survey of Consumer Finances

- [Codebook description of missing data procedures](#)
- Sensitive survey because of large very-wealthy oversample (based on IRS list of most wealthy households in the U.S.)
- The missing data and confidentiality protection procedures are both based on modeling the complex set of choices given to respondents about how much wealth information to reveal
- When a respondent does not want to provide an answer, bracketed interval choices are presented
- Missing data are multiply-imputed based on modeling the conditional distribution of the bracketed intervals, given covariates, and ignorability; see Kennickell ([2001](#))

"Gold Standard" for SIPP Synthetic Beta (SSB)

- We will discuss synthetic data in a later lecture, the synthetic SIPP data are described here: https://www.census.gov/programs-surveys/sipp/guidance/sipp-synthetic-beta-data-product.html

- The harmonized variables from the 1990-2008 SIPP panels have had all missing data imputed with a multiple imputation model that is the same one used to generate the synthetic data

- When a project from the SSB is validated (run on the confidential data by Census staff), the multiply-imputed completed harmonized files (completed Gold Standard Files) are used. These are the results released to the researcher

How are the QWIs Built?

- *Raw input files:*
  - UI wage records
  - QCEW/ES-202 report
  - Decennial census and ACS files
  - SSA-supplied administrative records
  - Census-derived administrative record household address files
  - LEHD geo-coding system
- *Processed data files:*
  - Individual characteristics
  - Employer characteristics
  - Employment history with earnings

Processing the Input Files

- Each quarter the complete history of every individual, every establishment, and every job is processed through the production system
- Missing data on the individuals are multiply imputed at the national level, posterior predictive distribution is stored
- Missing data on the employment history record are multiply imputed each quarter from fresh posterior predictive distribution
- Missing data on the employer characteristics are singly-imputed (explanation to follow)

Examples of Missing Data Problems

- Missing demographic data on the national individual file (birth date, sex, race, ethnicity, place of residence, and education)
  - Multiple imputations using information from the individual, establishment, and employment history files
  - Model estimation component updated irregularly
  - Imputations performed once for those in estimation universe, then once when a new PIK is encountered in the production system
- This process was used on the current QWI and for the S2011 snapshot

A Very Difficult Missing Data Problem

- The employment history records only code employer to the UI account level
- Establishment characteristics (industry, geo-codes) are missing for multi-unit establishments
- The establishment (within UI account) is multiply imputed using a dynamic multi-stage probability model
- Estimation of the posterior predictive distribution depends on the existence of a state with establishments coded on the UI wage record (MN)

How Is It Done?

- Every quarter the QWI processes over 7 billion employment histories (unique person-employer pair) covering 1990 to the current quarter less 9 months (2015:Q2, currently)

- Approximately 30-40% of these histories require multiple employer imputations to acquire workplace characteristics

- So, the system does more than 25 billion full information imputations every quarter

- The information used for the imputations is current, it includes all of the historical information for the person and every establishment associated with that person's UI account

Does It Work?

- Full assessment of total jobs, beginning-of-quarter employment, full-quarter employment, monthly earnings of full-quarter employed, total payroll

- Older assessment using the state that codes both (MN)

- Summary slide follows

Table 1: Summary of Total Variability of Total Employment (EmpTotal) by Table and Count

Table and EmpTotal count range	Proportion of Cells	Number of Cells	Median Count	Median Total Variation	Median Rubin Missingness Rate (Percent)	Quantiles of Coefficient of Variation			Median Approximate 90% Confidence Bounds Margin of Error		
						5th	Median	95th	Median df	Count	Percent
A: Private											
Age x Gender											
+1000	1.0000	46,480	80,832	8250.00	43.60%	0.0004	0.0011	0.0036	47	118	0.14%
Race x Ethnicity											
10-99	0.0199	695	53	49.80	96.10%	0.0836	0.1409	0.2605	9	10	19.48%
100-999	0.1306	4,553	452	411.00	95.30%	0.0242	0.0469	0.0935	9	28	6.49%
+1000	0.8495	29,612	13,614	5970.00	86.40%	0.0002	0.0044	0.0274	12	105	0.59%
Gender x Education											
+1000	1.0000	23,240	157,493	192000.00	96.60%	0.0013	0.0031	0.0086	9	606	0.43%
Industry x County											
zero measured value, after rounding	0.0045	12,955	0	0.30	94.40%	(a)	(a)	(a)	10	1	(a)
1-2	0.0001	188	1	0.42	79.35%	0.1982	0.3913	0.9553	14	1	52.73%
3-9	0.0158	45,741	7	0.50	0.00%	0.0614	0.1073	0.3800	9999	1	13.76%
10-99	0.2603	753,131	46	4.82	15.40%	0.0242	0.0511	0.1619	380	3	6.56%
100-999	0.4402	1,273,670	294	53.70	66.30%	0.0105	0.0235	0.0591	20	10	3.12%
+1000	0.2792	807,706	3,058	822.00	76.60%	0.0023	0.0081	0.0201	15	38	1.09%
Age x Gender x Industry x County											
zero measured value, after rounding	0.2019	8,888,449	0	0.21	95.10%	(a)	(a)	(a)	9	1	(a)
1-2	0.0050	220,862	2	0.35	67.20%	0.1564	0.3066	0.7280	19	1	40.71%
3-9	0.2171	9,557,988	5	0.81	61.50%	0.0885	0.1722	0.3887	23	1	22.73%
10-99	0.3796	16,713,425	27	5.35	69.70%	0.0382	0.0815	0.1800	18	3	10.84%
100-999	0.1622	7,142,409	225	55.20	75.40%	0.0142	0.0303	0.0618	15	10	4.06%
+1000	0.0343	1,511,324	1,972	502.00	75.70%	0.0041	0.0103	0.0201	15	30	1.38%
Race x Ethnicity x Industry x County											
zero measured value, after rounding	0.5839	19,047,330	0	0.20	95.20%	(a)	(a)	(a)	9	1	(a)
1-2	0.0058	190,628	2	0.69	91.70%	0.2636	0.6229	0.9257	10	1	85.47%
3-9	0.1330	4,339,494	5	2.35	88.90%	0.1325	0.3167	0.5963	11	2	43.18%
10-99	0.1607	5,240,825	26	10.10	85.30%	0.0426	0.1162	0.2704	12	4	15.76%
100-999	0.0830	2,707,617	249	75.90	79.90%	0.0137	0.0322	0.0745	14	12	4.33%
+1000	0.0336	1,094,612	2,586	766.00	79.20%	0.0031	0.0095	0.0212	14	37	1.28%
Gender x Education x Industry x County											
zero measured value, after rounding	0.0996	2,207,640	0	0.26	94.80%	(a)	(a)	(a)	10	1	(a)
1-2	0.0050	111,105	2	1.35	93.10%	0.4281	0.6538	0.9466	10	2	89.72%
3-9	0.2024	4,484,091	5	3.99	92.70%	0.2395	0.3791	0.6106	10	3	52.03%
10-99	0.4264	9,446,881	29	22.00	92.80%	0.0865	0.1624	0.2961	10	6	22.28%
100-999	0.2119	4,695,684	235	190.00	93.10%	0.0292	0.0569	0.0967	10	19	7.81%
+1000	0.0546	1,209,869	2,089	1790.00	93.50%	0.0087	0.0193	0.0321	10	58	2.65%

Table 2: Summary of Total Variability of Beginning-of-Quarter Employment (Emp) by Table and Count

Table and Emp count range	Proportion of Cells	Number of Cells	Median Count	Median Total Variation	Median Rubin Missingness Rate (Percent)	Quantiles of Coefficient of Variation			Median Approximate 90% Confidence Bounds Margin of Error		
						5th	Median	95th	Median df	Count	Percent
A: Private											
Age x Gender											
+1000	1.0000	45,712	61,308	5000.00	37.50%	0.0003	0.0011	0.0035	64	92	0.14%
Race x Ethnicity											
3-9	0.0000	1	9	3.57	92.60%	0.2099	0.2099	0.2099	10	3	28.81%
10-99	0.0282	968	47	36.80	95.90%	0.0783	0.1282	0.2729	9	8	17.73%
100-999	0.1607	5,509	466	328.00	94.70%	0.0115	0.0427	0.0828	10	25	5.86%
+1000	0.8111	27,806	11,716	4000.00	83.20%	0.0002	0.0041	0.0241	13	85	0.55%
Gender x Education											
+1000	1.0000	22,856	132,437	136000.00	96.60%	0.0013	0.0031	0.0085	9	510	0.42%
Industry x County											
zero measured value, after rounding	0.0096	27,314	0	0.29	95.50%	(a)	(a)	(a)	9	1	(a)
1-2	0.0001	313	2	0.37	66.90%	0.1425	0.3279	0.9000	20	1	43.45%
3-9	0.0242	69,274	7	0.42	0.00%	0.0561	0.1009	0.3648	9999	1	12.93%
10-99	0.2916	833,370	44	4.28	21.20%	0.0227	0.0499	0.1609	201	3	6.42%
100-999	0.4286	1,225,043	283	51.40	71.10%	0.0102	0.0235	0.0586	17	10	3.14%
+1000	0.2459	702,856	2,936	747.00	78.70%	0.0024	0.0080	0.0199	14	37	1.08%
Age x Gender x Industry x County											
zero measured value, after rounding	0.2368	10,146,295	0	0.20	95.60%	(a)	(a)	(a)	9	1	(a)
1-2	0.0051	217,085	2	0.33	69.70%	0.1409	0.2971	0.7099	18	1	39.52%
3-9	0.2243	9,610,779	5	0.72	63.10%	0.0811	0.1641	0.3815	22	1	21.69%
10-99	0.3624	15,526,526	26	4.98	72.80%	0.0365	0.0797	0.1786	17	3	10.63%
100-999	0.1432	6,136,088	222	51.50	77.80%	0.0135	0.0296	0.0610	14	10	3.97%
+1000	0.0282	1,210,018	1,931	452.00	77.40%	0.0042	0.0101	0.0196	15	29	1.35%
Race x Ethnicity x Industry x County											
zero measured value, after rounding	0.6229	20,152,114	0	0.19	95.70%	(a)	(a)	(a)	9	1	(a)
1-2	0.0052	168,667	2	0.66	92.30%	0.2579	0.6042	0.8972	10	1	82.90%
3-9	0.1222	3,951,621	5	2.16	89.60%	0.1241	0.3040	0.5810	11	2	41.45%
10-99	0.1465	4,740,254	26	9.16	85.80%	0.0395	0.1103	0.2619	12	4	14.96%
100-999	0.0746	2,411,633	246	69.90	81.40%	0.0130	0.0310	0.0716	13	11	4.18%
+1000	0.0287	926,867	2,513	687.00	80.90%	0.0031	0.0093	0.0205	13	35	1.25%
Gender x Education x Industry x County											
zero measured value, after rounding	0.1166	2,517,116	0	0.26	95.40%	(a)	(a)	(a)	9	1	(a)
1-2	0.0055	118,679	2	1.34	93.80%	0.4257	0.6500	0.9392	10	2	89.19%
3-9	0.2150	4,638,908	5	3.87	93.50%	0.2365	0.3763	0.6062	10	3	51.64%
10-99	0.4195	9,052,346	28	21.00	93.60%	0.0857	0.1620	0.2946	10	6	22.23%
100-999	0.1959	4,228,095	232	183.00	93.90%	0.0288	0.0563	0.0957	10	19	7.73%
+1000	0.0475	1,025,888	2,045	1670.00	94.20%	0.0086	0.0191	0.0315	10	56	2.62%

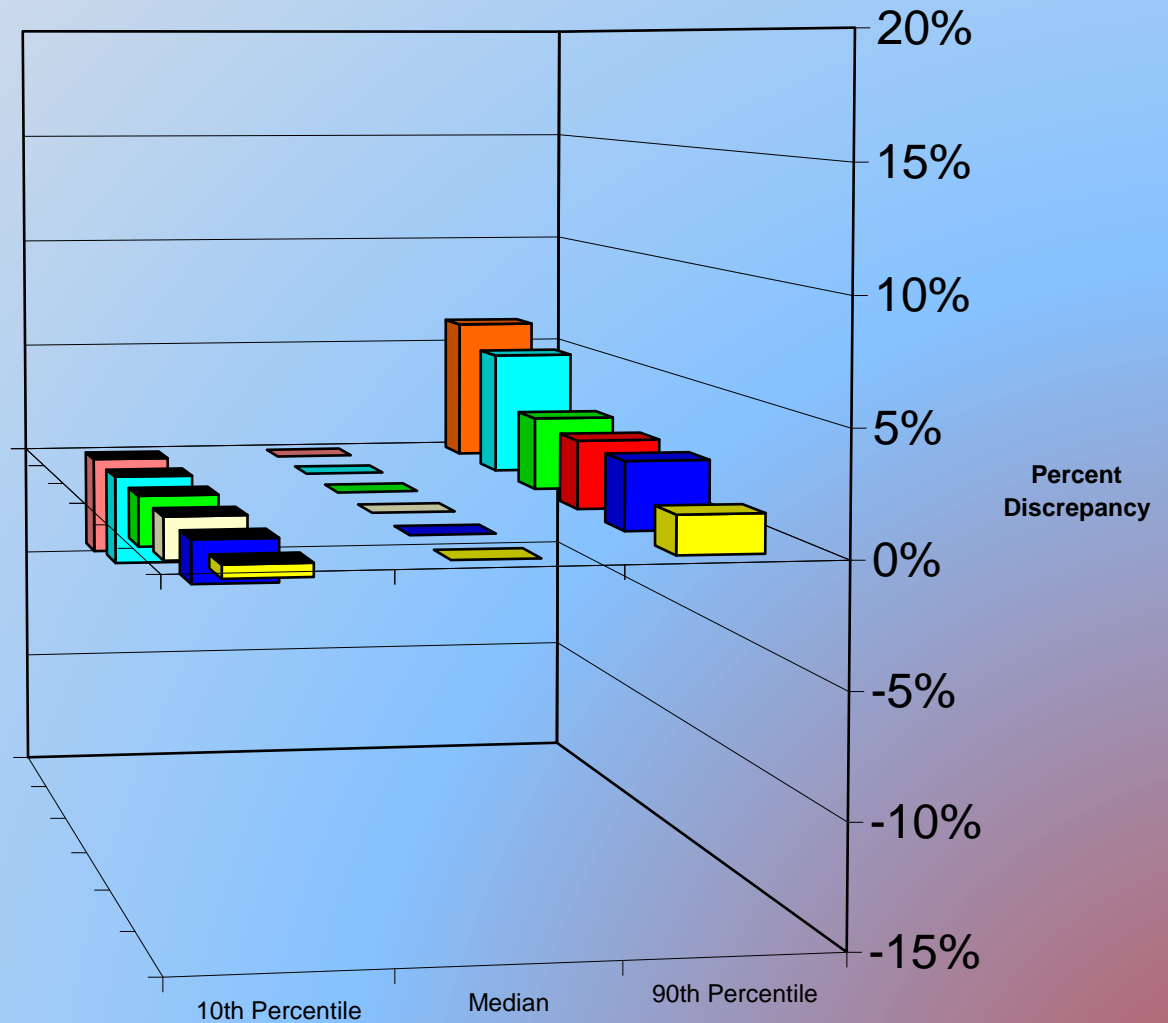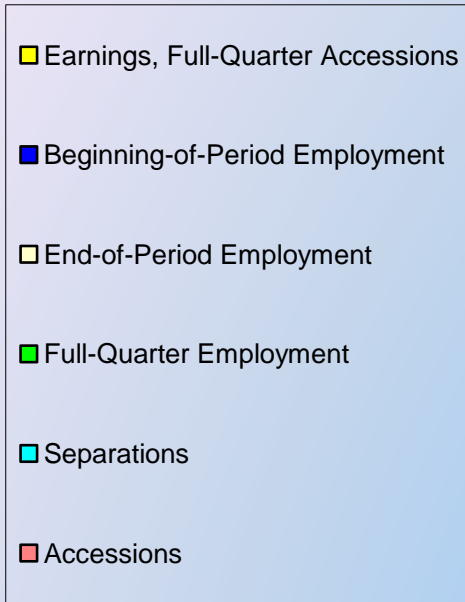| Table and EmpS Count Range	Proportion of Cells	Number of Cells	Median Count	Median Total Variation	Median Rubin Missingness Rate (Percent)	Quantiles of Coefficient of Variation			Median Approximate 90% Confidence Bounds Margin of Error		
						5th	50th	95th	Median df	Count	Percent
A: Private											
Age x Gender											
100-999	0.0001	6	965	414.00	79.70%	0.0211	0.0211	0.0211	14	27	2.84%
+1000	0.9999	44,938	50,251	3810.00	33.10%	0.0004	0.0012	0.0038	82	80	0.15%
Race x Ethnicity											
3-9	0.0005	17	9	8.14	96.60%	0.1746	0.3394	0.4180	9	4	46.94%
10-99	0.0351	1,184	46	32.70	95.10%	0.0747	0.1279	0.2935	9	8	17.69%
100-999	0.1780	6,001	452	301.00	94.40%	0.0133	0.0420	0.0856	10	24	5.76%
+1000	0.7863	26,506	10,312	3290.00	80.60%	0.0002	0.0042	0.0239	13	77	0.56%
Gender x Education											
+1000	1.0000	22,472	115,661	114000.00	96.40%	0.0014	0.0031	0.0088	9	467	0.43%
Industry x County											
zero measured value, after rounding	0.0142	40,036	0	0.28	95.60%	(a)	(a)	(a)	9	1	(a
1-2	0.0002	505	2	0.19	0.00%	0.1308	0.2518	0.8485	9999	1	32.27%
3-9	0.0327	92,014	7	0.40	0.00%	0.0571	0.1024	0.3636	9999	1	13.12%
10-99	0.3147	886,839	43	4.21	23.50%	0.0231	0.0509	0.1608	162	3	6.55%
100-999	0.4159	1,172,234	276	51.60	71.80%	0.0105	0.0241	0.0589	17	10	3.22%
+1000	0.2224	626,901	2,870	723.00	78.40%	0.0024	0.0081	0.0200	14	36	1.10%
Age x Gender x Industry x County											
zero measured value, after rounding	0.2674	11,179,951	0	0.20	95.60%	(a)	(a)	(a)	9	1	(a
1-2	0.0052	216,857	2	0.33	69.00%	0.1409	0.2958	0.7036	18	1	39.35%
3-9	0.2274	9,509,759	5	0.71	62.50%	0.0808	0.1639	0.3791	23	1	21.63%
10-99	0.3462	14,475,571	26	4.93	72.70%	0.0367	0.0805	0.1797	17	3	10.73%
100-999	0.1295	5,413,281	221	50.70	77.40%	0.0134	0.0295	0.0611	15	10	3.96%
+1000	0.0243	1,017,595	1,896	429.00	76.50%	0.0041	0.0099	0.0194	15	28	1.33%
Race x Ethnicity x Inudstry x County											
zero measured value, after rounding	0.6503	20,835,268	0	0.19	95.80%	(a)	(a)	(a)	9	1	(a
1-2	0.0049	157,195	2	0.65	92.20%	0.2579	0.6021	0.8874	10	1	82.62%
3-9	0.1144	3,664,172	5	2.08	89.10%	0.1213	0.2990	0.5761	11	2	40.76%
10-99	0.1366	4,375,170	26	8.75	84.90%	0.0385	0.1074	0.2582	12	4	14.56%
100-999	0.0685	2,195,777	244	67.70	80.80%	0.0131	0.0306	0.0705	13	11	4.14%
+1000	0.0253	810,388	2,467	663.00	80.30%	0.0031	0.0093	0.0204	13	35	1.26%
Gender x Education x Industry x County											
zero measured value, after rounding	0.1317	2,774,036	0	0.26	95.30%	(a)	(a)	(a)	9	1	(a
1-2	0.0059	124,663	2	1.32	93.80%	0.4278	0.6500	0.9368	10	2	89.19%
3-9	0.2237	4,709,531	5	3.83	93.50%	0.2365	0.3766	0.6065	10	3	51.68%
10-99	0.4122	8,679,900	27	20.60	93.60%	0.0860	0.1633	0.2954	10	6	22.40%
100-999	0.1839	3,871,290	230	181.00	93.80%	0.0288	0.0564	0.0958	10	18	7.74%
+1000	0.0426	897,250	2,011	1640.00	94.20%	0.0086	0.0192	0.0315	10	56	2.64%

Table 4: Summary of Total Payroll (Payroll) by Table and Count

Table and EmpTotal count range	Proportion of Cells	Number of Cells	Median Payroll	Median Total Variation	Median Rubin Missingness Rate (Percent)	Quantiles of Coefficient of Variation 5th	Median	95th	Median Approximate 90% Confidence Bounds Margin of Error Median df	Count	Percent
A: Private											
Age x Gender											
+1000	1.0000	46,480	375,627,224.50	3.96E+11	29.50%	0.0005	0.0016	0.0083	104	811,617.46	0.21%
Race x Ethnicity											
10-99	0.0199	695	233,792.00	2.29E+09	97.30%	0.1187	0.2083	0.4509	9	66,183.38	28.80%
100-999	0.1306	4,553	2,166,851.00	1.85E+10	96.10%	0.0334	0.0678	0.1493	9	188,112.25	9.38%
+1000	0.8495	29,612	71,561,892.50	4.67E+11	82.40%	0.0005	0.0070	0.0479	13	922,671.93	0.94%
Gender x Education											
+1000	1.0000	23,240	1,122,441,816.50	2.05E+13	96.10%	0.0018	0.0042	0.0117	9	6,261,928.94	0.57%
Industry x County											
zero measured value, after rounding	0.0040	12,955	0.00	9.45E+06	99.80%	0.0392	0.4429	1.0808	9	4,251.55	61.25%
1-2	0.1144	373,579	39,363.00	8.17E+06	0.00%	0.0000	0.0671	0.5900	9999	3,663.33	8.60%
3-9	0.0140	45,741	28,035.00	5.56E+06	0.00%	0.0302	0.0834	0.5458	9999	3,022.05	10.68%
10-99	0.2305	753,131	214,551.00	1.23E+08	9.57%	0.0198	0.0536	0.2327	984	14,222.64	6.88%
100-999	0.3899	1,273,670	1,568,970.00	2.55E+09	76.50%	0.0108	0.0307	0.0905	15	67,697.26	4.12%
+1000	0.2473	807,706	19,017,146.50	6.38E+10	80.70%	0.0035	0.0118	0.0329	13	341,035.20	1.59%
Age x Gender x Industry x County											
zero measured value, after rounding	0.1701	8,888,449	0.00	6.05E+05	100.00%	0.0000	0.2197	1.3084	9	1,075.74	30.39%
1-2	0.1618	8,454,917	4,523.00	3.34E+05	12.30%	0.0000	0.1107	0.8974	598	741.46	14.20%
3-9	0.1829	9,557,988	17,743.00	8.87E+06	83.20%	0.0453	0.1636	0.6023	13	4,021.15	22.08%
10-99	0.3198	16,713,425	113,931.00	1.16E+08	85.80%	0.0327	0.0939	0.2689	12	14,606.91	12.73%
100-999	0.1367	7,142,409	1,166,690.00	2.00E+09	87.20%	0.0144	0.0384	0.0942	11	60,974.46	5.23%
+1000	0.0289	1,511,324	14,164,728.00	3.60E+10	84.30%	0.0049	0.0135	0.0327	12	257,324.15	1.83%
Race x Ethnicity x Inudstry x County											
zero measured value, after rounding	0.4859	19,047,330	0.00	3.28E+06	100.00%	0.0115	0.4685	1.5608	9	2,504.77	64.80%
1-2	0.1727	6,771,506	4,934.00	6.96E+06	98.50%	0.0745	0.5842	1.2527	9	3,648.68	80.80%
3-9	0.1107	4,339,494	20,418.00	6.02E+07	96.70%	0.1040	0.4024	0.8577	9	10,730.73	55.65%
10-99	0.1337	5,240,825	125,020.00	3.38E+08	93.50%	0.0423	0.1488	0.3976	10	25,227.29	20.42%
100-999	0.0691	2,707,617	1,330,383.00	3.50E+09	88.60%	0.0153	0.0433	0.1145	11	80,661.63	5.90%
+1000	0.0279	1,094,612	16,504,365.00	5.75E+10	84.60%	0.0044	0.0135	0.0349	12	325,209.49	1.83%
Gender x Education x Industry x County											
zero measured value, after rounding	0.0845	2,207,640	153.00	2.35E+05	99.50%	0.0000	0.1886	2.1637	9	670.45	26.08%
1-2	0.1565	4,089,395	6,804.00	1.91E+07	98.80%	0.3127	0.6056	1.1470	9	6,044.33	83.76%
3-9	0.1716	4,484,091	25,232.00	1.38E+08	97.90%	0.2627	0.4790	0.8323	9	16,246.91	66.24%
10-99	0.3615	9,446,881	160,107.00	1.10E+09	97.20%	0.1026	0.2084	0.4189	9	45,869.87	28.83%
100-999	0.1797	4,695,684	1,522,171.00	1.37E+10	96.90%	0.0367	0.0763	0.1454	9	161,879.36	10.55%
+1000	0.0463	1,209,869	17,075,678.00	2.34E+11	96.40%	0.0118	0.0277	0.0559	9	669,020.05	3.83%

Table and EmpS count range	Proportion of Cells	Number of Cells	Median Average Monthly Earnings	Median Total Variation	Median Rubin Missingness Rate (Percent)	Quantiles of Coefficient of Variation			Median Approximate 90% Confidence Bounds Margin of Error		
						5th	Median	95th	Median df	Count	Percent
A: Private											
Age x Gender											
100-999	0.0001	6	1,686.00	13,600.00	87.00%	0.0691	0.0691	0.0691	11	159.00	9.42%
+1000	0.9999	44,938	2,146.00	6.79	22.90%	0.0004	0.0013	0.0066	171	3.35	0.17%
Race x Ethnicity											
3-9	0.0005	17	2,409.00	361,000.00	96.80%	0.1451	0.2600	0.6976	9	830.97	35.95%
10-99	0.0351	1,184	2,106.50	71,000.00	95.50%	0.0605	0.1252	0.3384	9	368.52	17.31%
100-999	0.1780	6,001	2,189.00	8,490.00	94.50%	0.0147	0.0425	0.1042	10	126.43	5.84%
+1000	0.7863	26,506	2,471.00	168.00	73.70%	0.0004	0.0052	0.0321	16	17.33	0.69%
Gender x Education											
+1000	1.0000	22,472	2,847.00	84.60	94.40%	0.0013	0.0032	0.0088	10	12.62	0.44%
Industry x County											
zero measured value, after rounding	0.0020	6,177	0.00	2,140,000.00	99.30%	(a)	(a)	(a)	9	2023.20	(a
1-2	0.1096	342,768	2,088.00	7,230.00	0.00%	0.0000	0.0545	0.2903	9999	108.98	6.98%
3-9	0.0294	92,014	1,523.00	8,540.00	0.00%	0.0202	0.0655	0.2728	9999	118.44	8.39%
10-99	0.2836	886,839	1,957.00	5,240.00	11.20%	0.0150	0.0385	0.1256	714	92.85	4.93%
100-999	0.3749	1,172,234	2,265.00	2,020.00	65.80%	0.0081	0.0207	0.0551	20	59.57	2.75%
+1000	0.2005	626,901	2,701.00	414.00	70.60%	0.0026	0.0080	0.0216	18	27.07	1.07%
Age x Gender x Industry x County											
zero measured value, after rounding	0.0025	98,832	0.00	2,690,000.00	99.60%	(a)	(a)	(a)	9	2268.34	(a
1-2	0.2269	8,953,462	1,297.00	9,140.00	0.00%	0.0000	0.0850	0.4718	9999	122.53	10.90%
3-9	0.2410	9,509,759	1,479.00	17,400.00	66.90%	0.0280	0.0932	0.2942	20	174.82	12.35%
10-99	0.3668	14,475,571	1,828.00	8,620.00	74.70%	0.0210	0.0538	0.1433	16	124.11	7.19%
100-999	0.1372	5,413,281	2,300.00	2,280.00	77.30%	0.0087	0.0224	0.0545	15	64.01	3.00%
+1000	0.0258	1,017,595	3,109.00	578.00	72.90%	0.0033	0.0084	0.0204	16	32.14	1.12%
Race x Ethnicity x Inudstry x County											
zero measured value, after rounding	0.0043	74,124	0.00	6,290,000.00	99.90%	(a)	(a)	(a)	9	3468.62	(a
1-2	0.3501	5,991,260	1,835.00	226,000.00	97.20%	0.0499	0.2686	0.7331	9	657.48	37.15%
3-9	0.2141	3,664,172	1,942.00	126,000.00	93.50%	0.0563	0.1853	0.4744	10	487.08	25.42%
10-99	0.2557	4,375,170	2,082.00	24,500.00	86.80%	0.0246	0.0757	0.2127	11	213.41	10.33%
100-999	0.1283	2,195,777	2,290.00	3,170.00	79.90%	0.0097	0.0253	0.0661	14	75.73	3.40%
+1000	0.0474	810,388	2,723.00	508.00	75.50%	0.0032	0.0087	0.0221	15	30.22	1.16%
Gender x Education x Industry x County											
zero measured value, after rounding	0.0037	83,776	0.00	3,250,000.00	98.60%	(a)	(a)	(a)	9	2493.29	(a
1-2	0.1917	4,326,849	1,800.00	424,000.00	98.10%	0.1419	0.3679	0.8320	9	900.56	50.88%
3-9	0.2087	4,709,531	1,908.00	241,000.00	95.90%	0.1254	0.2615	0.5459	9	678.95	36.16%
10-99	0.3846	8,679,900	2,210.00	63,700.00	94.40%	0.0534	0.1158	0.2588	10	346.32	15.89%
100-999	0.1715	3,871,290	2,558.00	12,000.00	94.30%	0.0203	0.0440	0.0937	10	150.32	6.04%
+1000	0.0398	897,250	3,175.00	2,870.00	94.20%	0.0076	0.0174	0.0408	10	73.51	2.38%

MN Known Unit vs. MN Imputed Unit Weighted

Legend:
- Earnings, Full-Quarter Accessions
- Beginning-of-Period Employment
- End-of-Period Employment
- Full-Quarter Employment
- Separations
- Accessions

Percent Discrepancy

10th Percentile   Median   90th Percentile

Cumulative Effect of All QWI Edits and Imputations

Entity Size*	Average Employment	Average Z-scores			Missingness Rates			Sample Size
		Beginning Employment (b)	Full-quarter Employment (f)	Earnings (z_w3)	Beginning Employment (b)	Full-quarter Employment (f)	Earnings (z_w3)	
all	437	8.79	8.09	10.15	27.1%	27.1%	31.2%	237,741
1-9	4	1.60	1.46	2.99	33.1%	33.3%	43.6%	95,520
10-99	35	4.84	4.40	6.69	24.4%	24.4%	25.7%	84,621
100-249	160	11.08	10.14	13.52	21.8%	21.6%	20.5%	21,187
250-499	354	16.66	15.29	19.37	20.9%	20.9%	19.1%	11,972
500-999	707	23.59	21.68	26.03	20.7%	20.6%	17.9%	8,787
1000+	5538	56.67	52.61	52.11	20.2%	20.1%	16.1%	15,654

*Entity is county x NAICS sector x race x ethnicity for 2008:q3.

- Z-score is the ratio of the QWI estimate to the square root of its total variation (within and between implicate components)
- Missingness rate is the ratio of the between variance to the total variance