Appendices

A Supplementary tables and figures

Figure A1: Creating the common-coded CPS-ASEC
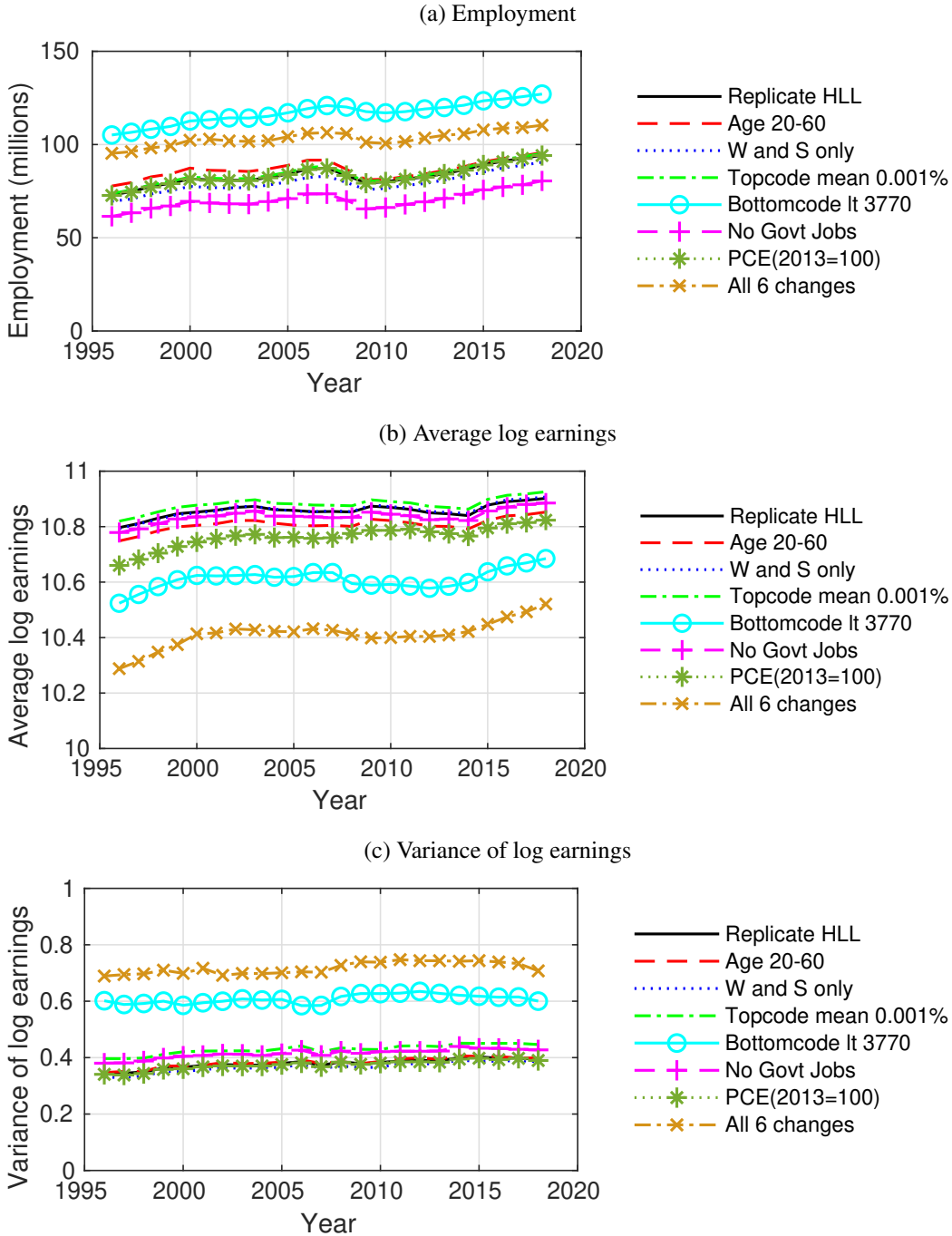
(a) Employment

(b) Average log earnings

(c) Variance of log earnings

Notes: Authors' tabulations of HLL CPS-ASEC data downloaded from *Journal of Economic Perspectives* website. Pooled males and females. See Table 4 for the criteria applied to select a "common coded" sample.

Figure A2: Mean, variance, and employment of HLL CPS-ASEC, HHS LEHD, common-coded CPS-ASEC, and common-coded LEHD earnings, by year

(a) Employment

(b) Average log earnings

(c) Variance of log earnings
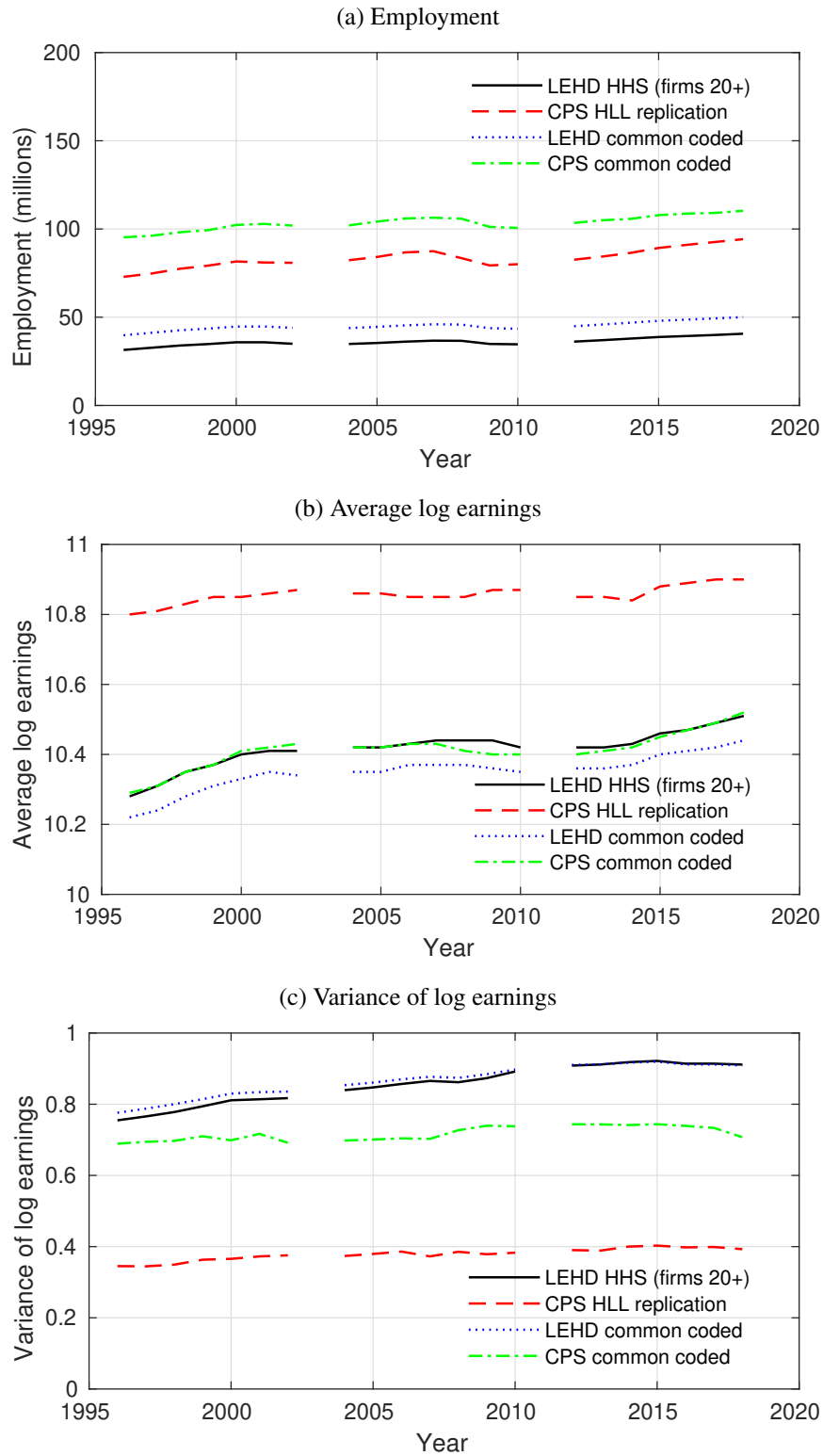
Notes: "CPS HLL replication" shows authors' tabulations of HLL CPS-ASEC data downloaded from *Journal of Economic Perspectives* website. "LEHD HHS (firms 20+)" shows authors' tabulations of LEHD administrative records as utilized in HHS, which considers only people who work at firms that employ at least twenty people. "Common coding" applies a consistent set of sample selection criteria to the HLL CPS-ASEC data downloaded from *Journal of Economic Perspectives* website and the HHS LEHD administrative records dataset. See Table 4 for further definitions applied to select the sample in each dataset.

Figure A3: Linking common-coded CPS-ASEC and common-coded LEHD data

(a) Observations
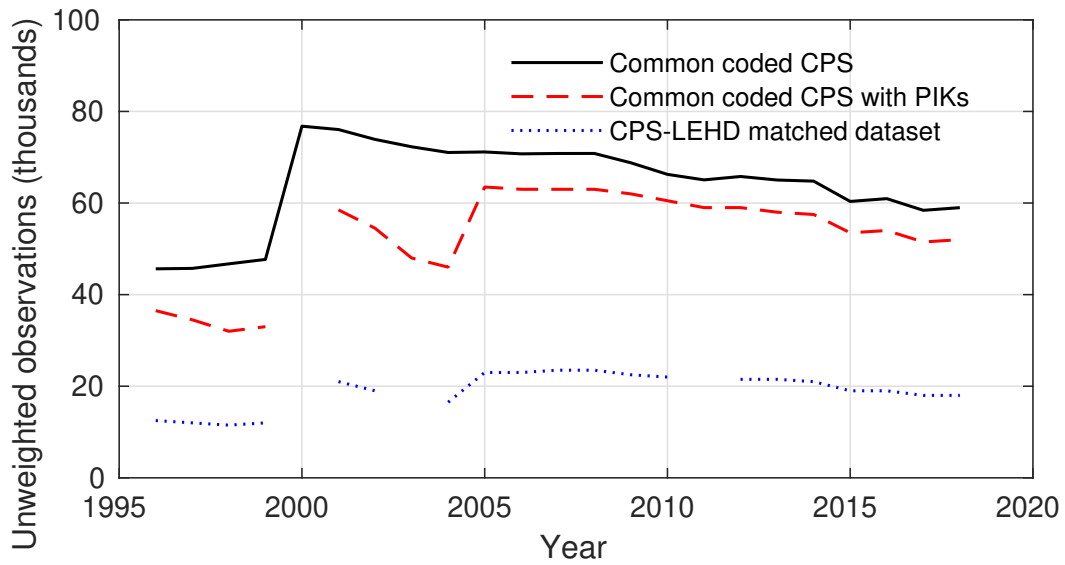
(b) Match rates
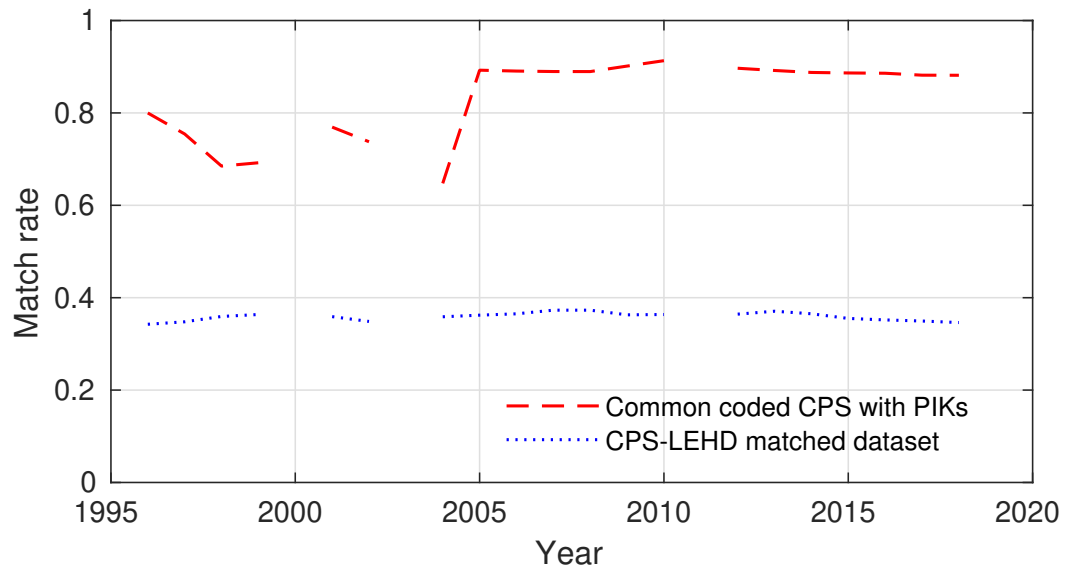
Notes: See Table 4 for the criteria applied to select a "common coded" sample. Further details about the construction of the CPS-LEHD matched dataset are provided in Section 4.1.

Figure A4: Earnings distributions of linked CPS-ASEC and LEHD

Notes: Results pool all years. See Table 4 for the criteria applied to select a "common coded" sample. Further details about the construction of the CPS-LEHD matched dataset are provided in Section 4.1.

Figure A5: PDF of earnings difference, linked CPS-ASEC and LEHD data
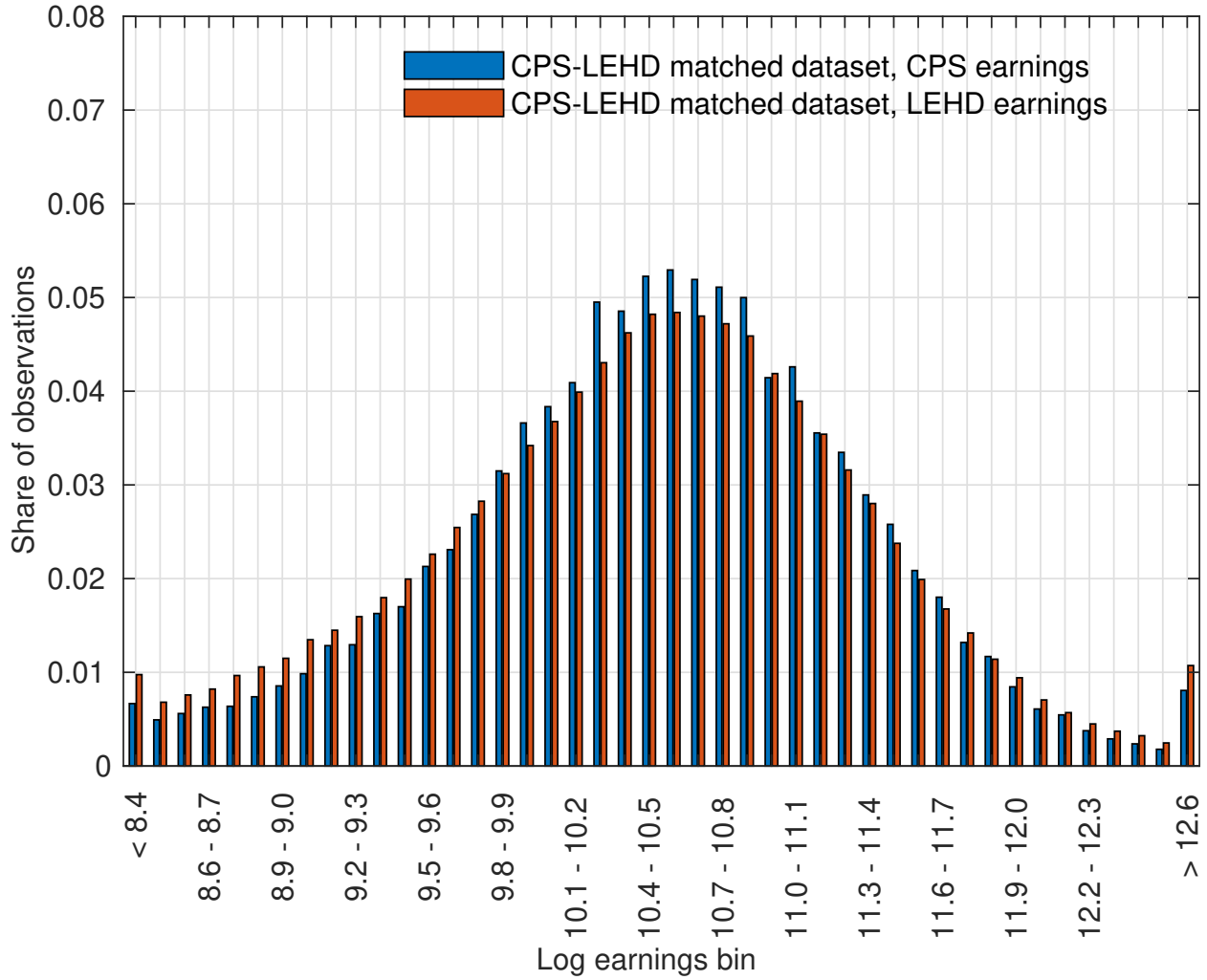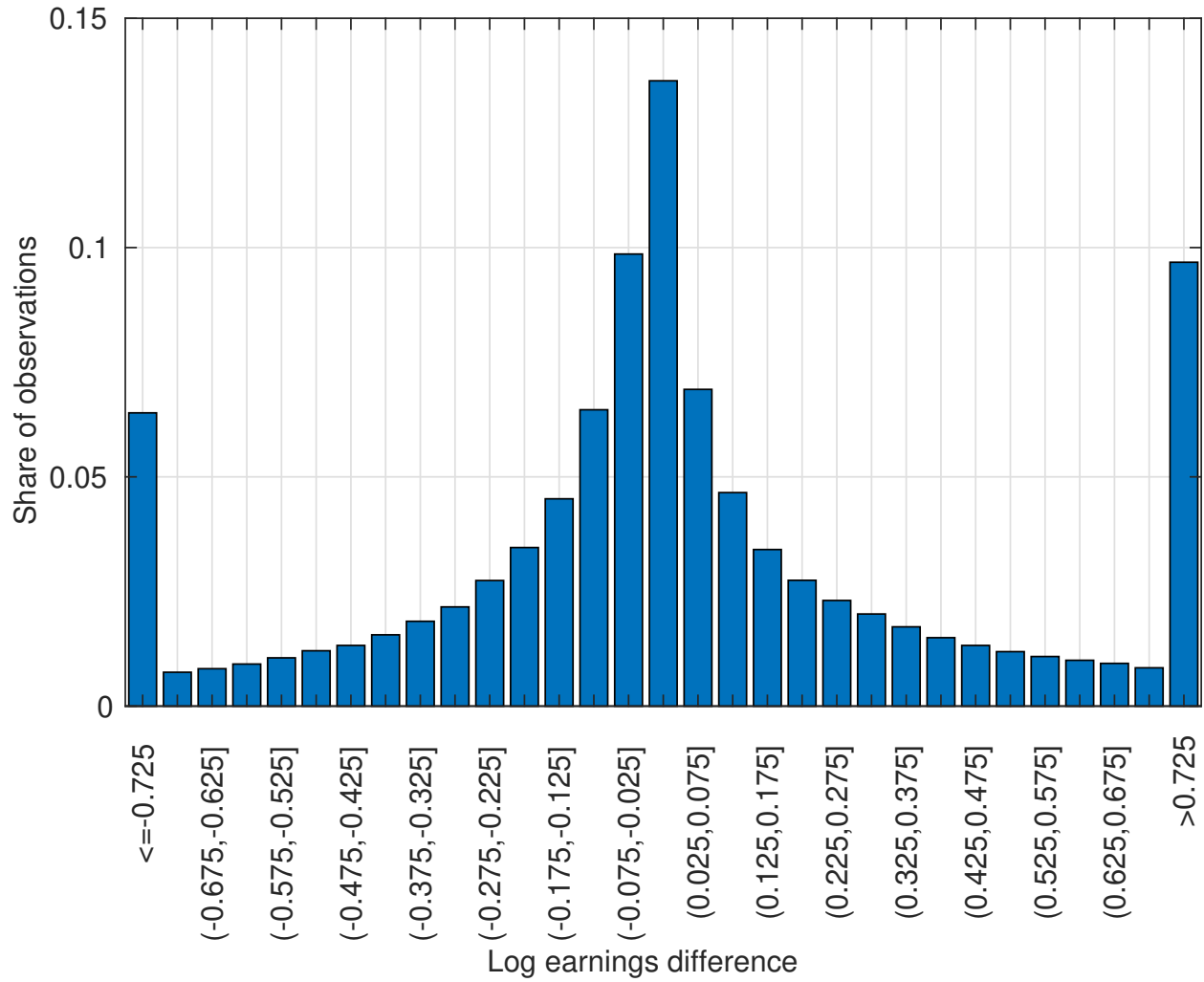

Notes: See Table 4 for the criteria applied to select a "common coded" sample. Further details about the construction of the CPS-LEHD matched dataset are provided in Section 4.1.

Table A1: Estimation of the human capital earnings equation using CPS ASEC data and 5-year intervals

	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010-2014	2015-2018	Growth 1975-79 to 2015-18
Earnings variance	0.283	0.292	0.310	0.332	0.349	0.372	0.380	0.390	0.398	0.115

Contributions to total variance, in levels:

	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010-2014	2015-2018	Growth 1975-79 to 2015-18
Age and education	0.045	0.049	0.060	0.071	0.079	0.088	0.092	0.094	0.093	0.048
Occupation	0.023	0.022	0.023	0.021	0.023	0.024	0.026	0.026	0.027	0.004
Industry	0.017	0.017	0.016	0.015	0.014	0.012	0.011	0.011	0.011	-0.006
Residual	0.198	0.205	0.220	0.225	0.234	0.249	0.251	0.259	0.267	0.069

Contributions to total variance, in percentages:

	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010-2014	2015-2018	Growth 1975-79 to 2015-18
Age and education	15.9%	16.8%	18.8%	21.5%	22.6%	23.5%	24.2%	24.1%	23.5%	41.9%
Occupation	8.3%	7.4%	7.1%	6.3%	6.6%	6.4%	6.9%	6.8%	6.8%	3.3%
Industry	5.9%	5.7%	5.1%	4.4%	3.9%	3.1%	2.8%	2.8%	2.7%	-5.3%
Residual	69.9%	70.1%	69.0%	67.8%	66.8%	66.9%	66.0%	66.4%	67.0%	60.1%

Notes: We downloaded the HLL CPS-ASEC data from the *Journal of Economic Perspectives* website. Our earnings variable is the natural log of real annual labor earnings. Our regression specification is based on HLL Figure 4, except we use labor earnings instead of total income, and we pool male and females. "Age and education" is the fraction of the variance of labor earnings explained by equation (2). "Occupation" is the marginal contribution of including occupation, obtained by subtracting the percentage of the variance explained by equation (2) from that of equation (3). "Industry" is the marginal contribution of industry, obtained by subtracting the percentage of variance explained by equation (3) from that of equation (4). Industry is defined using 12 SIC categories. "Residual" is the fraction of the variance that is unexplained when estimating equation (4).

Table A2: First and second panels replicate HLL Figure 4, third panel is HLL with pooled genders, fourth panel is pooled genders with labor earnings rather than total income (Table A1 of this paper)

	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010-2014	2015-2018	Growth 1975-79 to 2015-18
HLL Figure 4, males										
Total income variance	0.262	0.286	0.329	0.358	0.389	0.416	0.427	0.444	0.457	0.195
Age and education	0.052	0.061	0.076	0.091	0.102	0.112	0.119	0.124	0.124	0.072
Occupation	0.009	0.010	0.013	0.016	0.021	0.023	0.025	0.026	0.028	0.019
Industry	0.009	0.010	0.011	0.011	0.010	0.009	0.008	0.008	0.008	-0.001
Residual	0.193	0.206	0.229	0.240	0.256	0.272	0.274	0.286	0.297	0.105
HLL Figure 4, females										
Total income variance	0.190	0.208	0.250	0.273	0.301	0.311	0.333	0.351	0.374	0.185
Age and education	0.038	0.041	0.054	0.066	0.077	0.081	0.089	0.096	0.102	0.065
Occupation	0.012	0.013	0.017	0.019	0.019	0.020	0.023	0.023	0.025	0.013
Industry	0.007	0.008	0.010	0.010	0.009	0.007	0.006	0.006	0.006	-0.001
Residual	0.133	0.147	0.170	0.179	0.195	0.204	0.215	0.226	0.241	0.108
HLL Figure 4, pooled										
Total income variance	0.298	0.305	0.337	0.352	0.380	0.396	0.407	0.421	0.436	0.138
Age and education	0.050	0.057	0.069	0.081	0.091	0.100	0.102	0.105	0.106	0.056
Occupation	0.025	0.022	0.023	0.022	0.024	0.025	0.028	0.028	0.029	0.005
Industry (SIC 12)	0.016	0.016	0.016	0.015	0.014	0.012	0.011	0.011	0.011	-0.005
Residual	0.207	0.211	0.230	0.235	0.250	0.262	0.266	0.277	0.289	0.083
Table A1 this paper, pooled										
Labor earnings variance	0.283	0.292	0.318	0.332	0.349	0.372	0.380	0.390	0.398	0.115
Age and education	0.045	0.049	0.060	0.071	0.079	0.088	0.092	0.094	0.093	0.048
Occupation	0.023	0.022	0.023	0.021	0.023	0.024	0.026	0.026	0.027	0.004
Industry (SIC 12)	0.017	0.017	0.016	0.015	0.014	0.011	0.011	0.011	0.011	-0.006
Residual	0.198	0.205	0.220	0.225	0.234	0.249	0.251	0.259	0.267	0.069

Notes: The top two panels replicate columns {B, E, H, P, W, X} of HLL's figure_4.xlsx downloaded from the *Journal of Economic Perspectives* website.

Table A3: Variance decomposition of the human capital earnings equation, CPS-ASEC data

	1975-1981	1982-1988	1989-1995	1996-2002	2004-2010	2012-2018	Growth 1975-81 to 2012-18	Growth 1996-02 to 2012-18
Earnings variance	0.283	0.310	0.333	0.360	0.380	0.397	0.113	0.037
Using 12 SIC industries								
Within-industry:	95.2%	95.2%	95.4%	95.8%	95.3%	94.9%	94.0%	86.2%
Age, educ., & occ.	24.9%	25.7%	27.5%	28.8%	29.3%	28.0%	35.8%	20.5%
Residual	70.3%	69.5%	67.8%	67.0%	66.0%	66.9%	58.2%	65.7%
Between-industry:	4.8%	4.8%	4.6%	4.2%	4.7%	5.1%	6.0%	13.8%
Segregation	2.2%	2.4%	2.8%	3.0%	3.5%	4.0%	8.5%	13.3%
Pay premia	7.9%	6.7%	5.5%	4.5%	3.5%	3.4%	-7.8%	-6.9%
Sorting	-5.3%	-4.3%	-3.7%	-3.3%	-2.4%	-2.3%	5.2%	7.4%
Using 18 NAICS industries								
Within-industry:	93.2%	93.3%	93.8%	94.1%	93.6%	92.5%	90.8%	76.9%
Age, educ., & occ.	24.7%	25.9%	27.6%	28.5%	29.1%	27.5%	34.6%	18.2%
Residual	68.5%	67.5%	66.2%	65.6%	64.5%	65.0%	56.2%	58.8%
Between-industry:	6.8%	6.7%	6.2%	5.9%	6.4%	7.5%	9.2%	23.1%
Segregation	2.9%	2.9%	3.2%	3.3%	3.8%	4.4%	8.1%	14.8%
Pay premia	10.2%	9.0%	7.3%	6.0%	5.4%	5.5%	-6.2%	1.0%
Sorting	-6.3%	-5.2%	-4.3%	-3.4%	-2.8%	-2.4%	7.3%	7.3%

Notes: Authors' tabulations of HLL CPS-ASEC data downloaded from *Journal of Economic Perspectives* website. Pooled males and females. The year 2000 is deleted. Earnings is natural log of real annual labor earnings. The 12 SIC aggregate industries are defined following the Standard Industrial Classification system. The 18 NAICS aggregate industries are defined following the North American Industrial Classification System. Coding of CPS industry data (indly) into NAICS industries follows Table C-5 of Pollard (2019). See equation (10) for definitions.

Table A4: Industry sector agreement rate in the CPS-LEHD linked data

CPS NAICS	CPS NAICS industry distribution	% LEHD with same industry
Agriculture	1.0%	58.8%
Mining	0.9%	51.5%
Construction	6.9%	70.2%
Manufacturing	15.7%	66.7%
Wholesale Trade	4.1%	**46.9%**
Retail Trade	13.3%	71.3%
Transportation & Warehousing	4.2%	64.4%
Utilities	0.9%	53.6%
Information	6.1%	**46.6%**
Finance & Insurance	6.5%	*77.0%*
Real Estate & Rental	1.9%	50.9%
Professional &+ Business Services	10.3%	67.1%
Educational Services	2.8%	**44.2%**
HealthCare & Social Assistance	12.8%	*83.4%*
Arts, Entertainment, & Recreation	1.5%	51.5%
Accommodation & Food Services	6.7%	*77.9%*
Other Services	3.2%	**48.0%**
Unknown (2nd job)	1.2%	**0.0%**
Total	100%	66.1%

Notes: See Table 4 for the criteria applied to select a "common coded" sample. Further details about the construction of the CPS-LEHD matched dataset are provided in Section 4.1. *Italics* shows agreement >75%. **Bold** shows agreement <50%.

Table A5: Variance decomposition of the human capital earnings equation

	(1)	(2)	(3)	(4)	(5)	(6)
Data	CPS	CPS	Linked CPS-LEHD	Linked CPS-LEHD	Linked CPS-LEHD	Linked CPS-LEHD
Sample	HLL JEP	Common coded	Common coded	Common coded	Common coded	Common coded
Earnings measure	CPS	CPS	CPS	CPS	CPS	LEHD
Industry measure	CPS 18	CPS 18	CPS 18	LEHD 18	LEHD 299	LEHD 299
Variance level 1996-2002						
Earnings variance	0.360	0.703	0.667	0.667	0.667	0.746
Within-industry:	94.1%	86.9%	87.6%	91.7%	82.5%	78.3%
Age, education & occupation	28.5%	24.7%	27.8%	28.1%	22.4%	18.8%
Age and education	13.0%	12.3%	14.0%	14.3%	11.3%	10.3%
Occupation	7.2%	6.1%	6.5%	6.6%	5.6%	4.3%
Covariance: age+educ. & occ.	8.2%	6.3%	7.4%	7.2%	5.5%	4.2%
Residual	65.6%	62.2%	59.8%	63.6%	60.1%	59.5%
Between-industry:	5.9%	13.1%	12.4%	8.3%	17.5%	21.7%
Segregation:	3.3%	2.2%	2.2%	2.1%	4.2%	3.1%
Age and education	1.8%	1.0%	1.1%	1.1%	1.8%	1.5%
Occupation	0.5%	0.4%	0.4%	0.3%	0.8%	0.5%
Covariance: age+educ. & occ.	1.1%	0.8%	0.8%	0.7%	1.6%	1.1%
Pay Premia	6.0%	11.0%	8.6%	4.3%	8.1%	11.2%
Sorting:	-3.4%	-0.1%	1.5%	1.8%	5.2%	7.4%
Covariance: age+educ. & ind.	-2.2%	-0.2%	0.9%	0.9%	3.3%	5.4%
Covariance: industry & occ.	-1.2%	0.1%	0.6%	0.9%	1.9%	1.9%

Notes: The rows titled "Data" and "Sample" indicate the data used for the variance decomposition (see text for description). "Earnings measure" indicates whether CPS or LEHD earnings is used in the decomposition. "Industry measure:" "CPS 18" refers to 18 NAICS sectors from the CPS-ASEC (recoding CPS-ASEC variable indly following Table C-5 of Pollard (2019)), "LEHD 18" refers to NAICS sectors from the LEHD, and "LEHD 299" refers to 299 4-digit NAICS industries from the LEHD. See equation (10) for definitions.

Table A6: Variance decomposition of the human capital earnings equation (continued)

	(1)	(2)	(3)	(4)	(5)	(6)
Data	CPS	CPS	Linked CPS-LEHD	Linked CPS-LEHD	Linked CPS-LEHD	Linked CPS-LEHD
Sample	HLL JEP	Common coded	Common coded	Common coded	Common coded	Common coded
Earnings measure	CPS	CPS	CPS	CPS	CPS	LEHD
Industry measure	CPS 18	CPS 18	CPS 18	LEHD 18	LEHD 299	LEHD 299
Variance level 2012-2018						
Earnings variance	0.397	0.738	0.717	0.717	0.717	0.845
Within-industry:	92.5%	86.1%	85.3%	88.6%	79.1%	73.1%
Age, education, & occupation:	27.5%	24.9%	27.2%	27.8%	21.4%	18.2%
Age and education	12.9%	12.9%	14.3%	14.5%	11.3%	10.9%
Occupation	6.9%	5.7%	5.9%	6.3%	5.1%	3.7%
Covariance: age+educ. & occ.	7.7%	6.3%	6.9%	7.0%	4.9%	3.6%
Residual	65.0%	61.2%	58.1%	60.8%	57.8%	54.9%
Between-industry:	7.5%	13.9%	14.7%	11.4%	20.9%	26.9%
Segregation:	4.4%	3.9%	4.4%	4.1%	6.4%	4.6%
Age and education	2.0%	1.5%	1.7%	1.8%	2.6%	2.0%
Occupation	0.7%	0.7%	0.7%	0.6%	1.0%	0.7%
Covariance: age+educ. & occ.	1.8%	1.7%	2.0%	1.7%	2.7%	1.8%
Pay premia	5.5%	8.3%	6.9%	3.7%	7.3%	13.2%
Sorting:	-2.4%	1.7%	3.4%	3.5%	7.2%	9.1%
Covariance: age+educ. & ind.	-1.8%	0.8%	2.0%	1.9%	4.5%	6.2%
Covariance: industry & occ.	-0.6%	1.0%	1.5%	1.6%	2.7%	2.9%

Notes: The rows titled "Data" and "Sample" indicate the data used for the variance decomposition (see text for description). "Earnings measure" indicates whether CPS or LEHD earnings is used in the decomposition. "Industry measure:" "CPS 18" refers to 18 NAICS sectors from the CPS-ASEC (recoding CPS-ASEC variable indly following Table C-5 of Pollard (2019)), "LEHD 18" refers to NAICS sectors from the LEHD, and "LEHD 299" refers to 299 4-digit NAICS industries from the LEHD. See equation (10) for definitions.

Table A7: Variance decomposition of the human capital earnings equation

	(1)	(2)	(3)	(4)	(5)	(6)
Data	CPS	CPS	Linked CPS-LEHD	Linked CPS-LEHD	Linked CPS-LEHD	Linked CPS-LEHD
Sample	HLL JEP	Common coded	Common coded	Common coded	Common coded	Common coded
Earnings measure	CPS	CPS	CPS	CPS	CPS	LEHD
Industry measure	CPS 18	CPS 18	CPS 18	LEHD 18	LEHD 299	LEHD 299
Change from 1996-02 to 2012-18						
Variance growth	0.037	0.035	0.050	0.050	0.050	0.100
Within-industry:	76.9%	70.7%	54.0%	47.8%	34.5%	33.8%
Age, education & occupation:	18.2%	30.0%	17.9%	23.7%	7.8%	13.6%
Age and education	11.8%	25.4%	18.3%	17.7%	12.2%	14.9%
Occupation	4.1%	1.3%	-1.4%	1.4%	-1.8%	-0.4%
Covariance: age+educ. & occ.	2.2%	6.9%	1.0%	4.6%	-2.6%	-0.9%
Residual	58.8%	40.7%	36.1%	24.1%	26.7%	20.3%
Between-industry:	23.1%	29.3%	46.0%	52.2%	65.5%	66.2%
Segregation	14.8%	36.9%	33.7%	30.3%	34.9%	15.3%
Age and education	3.8%	11.7%	11.0%	11.2%	12.7%	6.2%
Occupation	2.5%	5.7%	5.4%	4.0%	4.6%	1.9%
Covariance: age+educ. & occ:	8.4%	19.4%	17.3%	15.1%	17.5%	7.1%
Pay premia	1.0%	-47.0%	-16.1%	-2.6%	-3.4%	21.4%
Sorting	7.3%	39.4%	28.5%	24.5%	34.1%	29.5%
Covariance: age+educ. & ind.	2.0%	20.6%	15.7%	14.1%	20.5%	19.4%
Covariance: industry & occ.	5.3%	18.9%	12.7%	10.4%	13.5%	10.1%

Notes: The rows titled "Data" and "Sample" indicate the data used for the variance decomposition (see text for description). "Earnings measure" indicates whether CPS or LEHD earnings is used in the decomposition. "Industry measure:" "CPS 18" refers to 18 NAICS sectors from the CPS-ASEC (recoding CPS-ASEC variable indly following Table C-5 of Pollard (2019)), "LEHD 18" refers to NAICS sectors from the LEHD, and "LEHD 299" refers to 299 4-digit NAICS industries from the LEHD. See equation (10) for definitions.