

Wage & Employment Dynamics

THE WED PROJECT

Longitudinal attrition in ASHE



Disclaimer

"This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates."

Motivation

- ASHE sample of 1% of PAYE jobs is based on the same final digits of the NI number
- Many have used the panel dimension of ASHE to make inferences about:
 - Wage progression and its determinants (e.g. Elsby et al, 2016; Schaeffer and Singleton, 2019)
 - Job displacement or employment exit and its determinants (e.g. Dickens et al, 2015; Stokes et al, 2017)
- But ONS seemingly make little effort to preserve the longitudinal integrity of the sample: their focus is cross-sectional representativeness
- Unbiased inference requires that those who exit from the ASHE dataset between t1 and t2 are representative of those who exit PAYE employment (i.e. sample attrition is either random or ignorable)
- This assumption has – to our knowledge – never been tested
- We investigate the panel characteristics of ASHE and seek to estimate the degree of sample attrition by benchmarking to the Longitudinal Annual Population Survey (APS)

Headlines

- Employees are three times more likely to exit ASHE year-on-year than they are to exit employment
 - ASHE sample exit rate = 0.25 p.a.
 - APS employment exit rate = 0.08 p.a.
 - > Sample attrition rate = 0.17 p.a.
- Sample attrition is non-random:
 - Relatively high among those: with low wages/hours; aged 20-44; in private-sector services; in Lon/SE
 - Higher in later years of ASHE
- So sample attrition *may be non-ignorable* in any analysis of longitudinal outcomes (wage progression, job displacement, employment exit etc)
- For analysis of year-to-year changes among continuing sample members: we use our comparison of ASHE and APS to derive year-to-year weights to correct for observable attrition biases
- For analysis of changes over longer periods, or analysis of job displacement/employment exit: strong assumptions needed – proceed with caution ... until we obtain data on employment transitions for sample members

Sample retention rate across year-pairs

Table 1: Sample Retention rate in ASHE from year t to year $t+i$

Year (T)	T+1	T+2	T+3	T+4	T+5	T+6	T+7	T+8	T+9	T+10	T+11	T+12	T+13	T+14	T+15
2004	77.3	71.6	55.3	52.3	61.4	58.8	58.4	54.5	53.2	51.5	48.4	45.1	43.2	41.2	38.8
2005	78.1	59.0	55.1	64.4	61.2	60.8	56.6	55.3	53.7	50.3	46.9	44.9	42.8	40.2	
2006	64.1	58.6	67.8	64.0	63.3	58.9	57.4	55.7	52.2	48.6	46.5	44.3	41.8		
2007	75.7	71.2	66.3	65.4	60.8	59.1	57.4	53.7	50.0	47.9	45.4	42.8			
2008	77.2	70.2	68.5	63.3	61.6	59.6	55.7	52.0	49.7	47.1	44.5				
2009	77.4	73.1	67.0	64.8	62.5	58.2	54.3	51.9	49.3	46.5					
2010	78.6	70.7	67.9	65.3	60.7	56.5	54.1	51.9	48.9						
2011	76.0	71.5	68.1	63.1	58.5	55.9	53.2	50.2							
2012	77.8	72.4	66.6	61.4	58.6	55.5	52.4								
2013	78.1	70.3	64.4	61.1	57.9	54.5									
2014	75.0	67.5	63.3	59.7	56.2										
2015	73.7	67.1	62.4	58.5											
2016	73.4	66.3	61.6												
2017	72.7	65.6													
2018	71.3														
Average	75.1	68.2	64.2	61.9	60.2	57.8	55.5	53.1	51.0	48.7	46.3	44.3	42.6	40.7	38.8

Note: Shading is used to indicate the scale of sample retention: darker green for higher rates, amber for middling rates, then darker shades of red for lower rates.

Source: ASHE

Possible reasons for sample exit

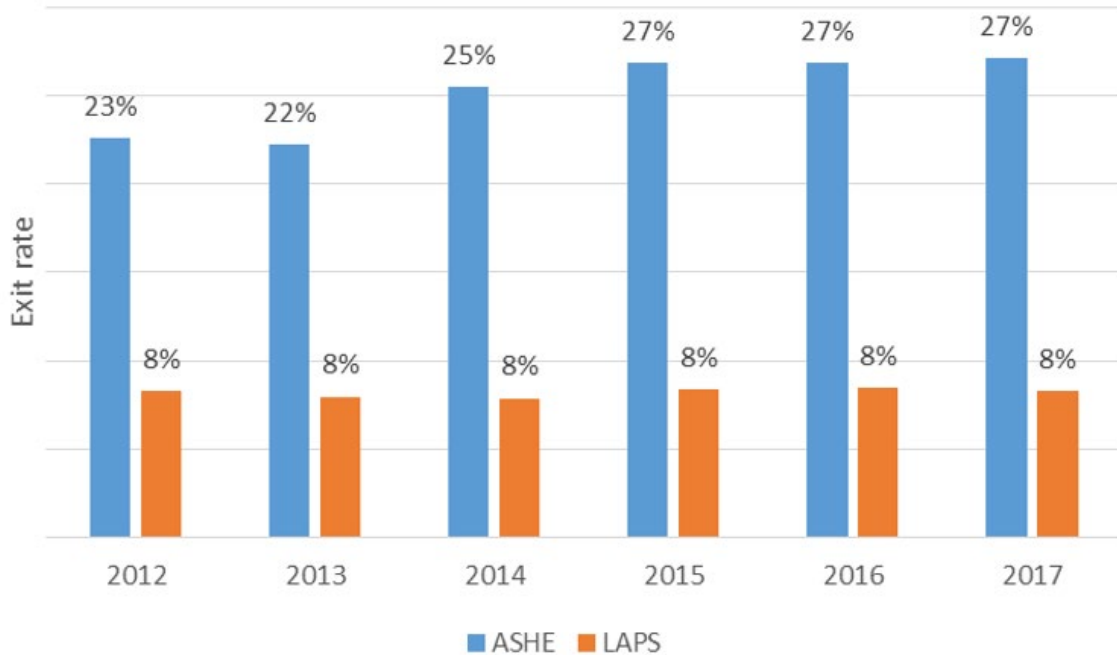
1. Sample member **moves out of scope** to the survey between $t1$ and $t2$
2. Sample member remains in scope at $t2$ but **is not sampled**
3. Sample member remains in scope and is sampled at $t2$ but **cannot be traced**
4. Sample member remains in scope, and is traced at $t2$, but **employer declines to respond**
5. Sample member remains in scope, is traced and employer responds at $t2$, but **repeated observations on this sample member are not linked in the dataset**



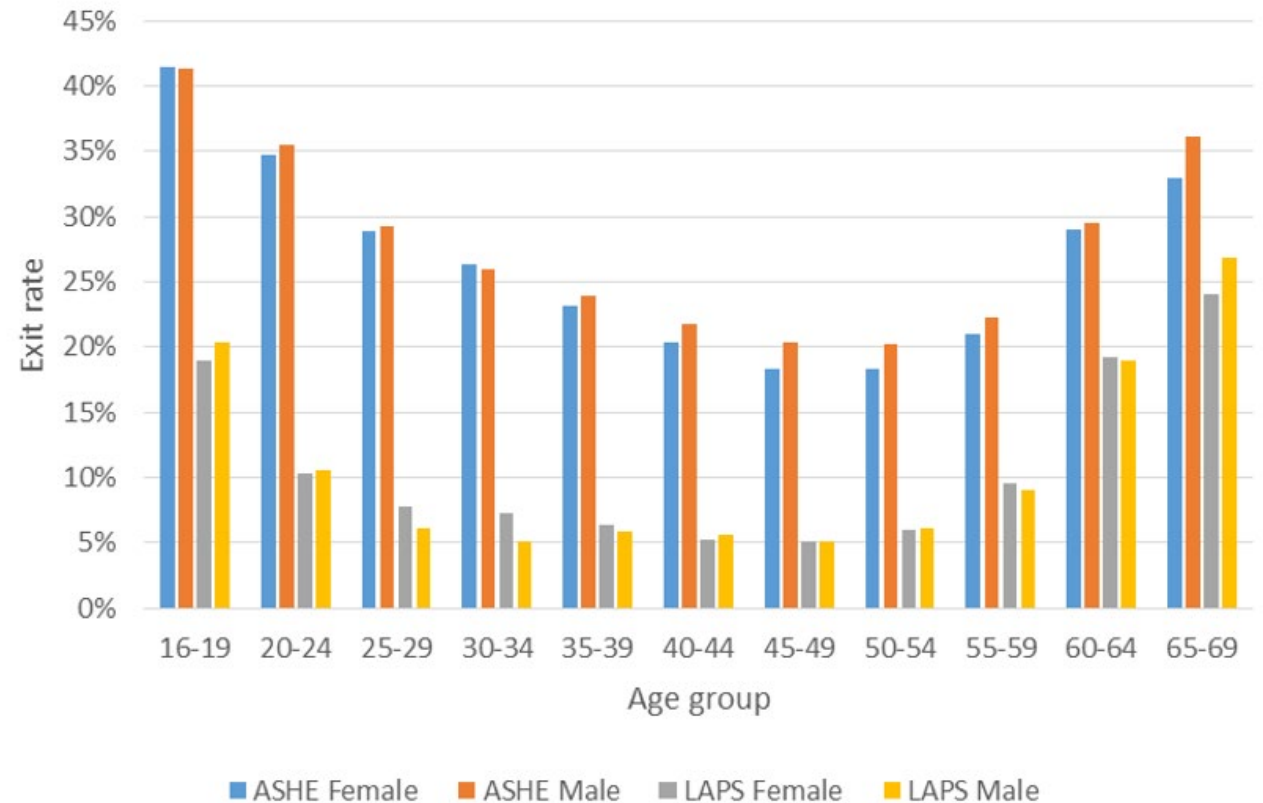
Sample attrition

1. Sample member moves out of scope

Only 8% of employees p.a. exit employment to self-employment, unemployment or inactivity (APS)



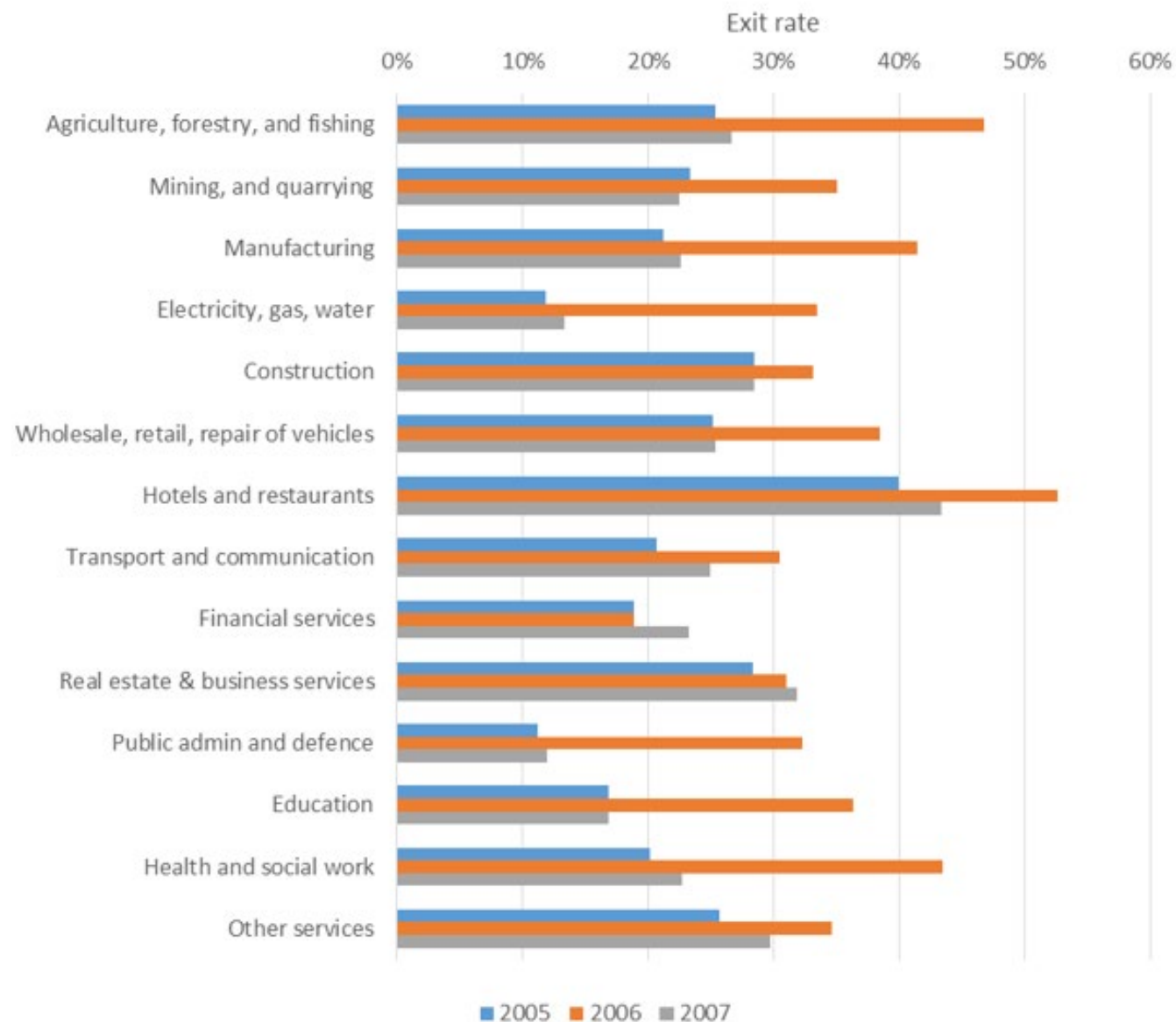
Comparison -> sample attrition disproportionately affects middle-aged in ASHE





2. Sample member not sampled at t_2

- ONS budget cuts: 20% sample reduction in 2006-7, focused on sectors with low wage variance
- Big increase in sample exit rates in 2006 in many SIC Sections – see chart
- Temporary effect: non-sampled individuals available for sampling again in 2008



3. Employer not traced at t2

- Hard to generate evidence on this point

4. Employer does not respond at t_2

- Non-response at t_2 may occur within a continuing spell of employment
 - 2007-2017: one-fifth (20%) of all sample exits followed by reappearance of the sample member in a later year, with the same employer and same employment start date (often after one year)
 - More likely in large firms, in manufacturing and in the public sector.
 - More likely among female employees and those in middle age.
- Non-response at t_2 may occur when employee moves from a 'responding employer' to a 'non-responding employer'
 - Will deflate the job exit rate among those who continue in employee status, all other things equal
 - ASHE: 8.4% of those who appear in the sample in two consecutive years switch their employer
 - APS: 10.8% of those who are in an employee job in two consecutive years switch their employer
 - Indicates further sample attrition when people move employer

5. Observations not linked over time

- Across the period 2007-2019, one-in-seven PIDENS (15%) appear only once
- Even among males aged 30-44, one in ten (11%) are observed for only one year across this 12 year period
- However, in only 1 per cent of sample exits can we identify a person in the following year of data with the same gender, firm ID (ENTREF), employment start date (EMPSTA) and workplace postcode (WPOST) as the person who exited the sample in the previous year.
- Does not suggest that linkage errors are common



Accounting for sample attrition

- Use the Longitudinal APS to run a probit regression predicting employment exit between year t and year $t+1$ as a function of characteristics measured in APS but also appearing in ASHE:

$$Emp_exit_{APS} = \alpha + \beta X_{APS} + \varepsilon \quad (\text{Eq. 1})$$

- In ASHE, run the equivalent probit regression to predict sample exit between year t and year $t+1$:

$$Sample_exit_{ASHE} = \alpha + \beta X_{ASHE} + \varepsilon \quad (\text{Eq. 2})$$

- Longitudinal weight for analysis of sample members observed in two consecutive years:

$$watt_{it} = weight_{it} * \frac{1}{(1 - \hat{p}(samp_{exit}))} * (1 - \hat{p}(emp_{exit}))$$

Impact of $watt_{it}$

Longitudinal weight:

- Boosts the representation sample of those most likely to exit ASHE between *year t* and *t+1*
- Then calibrates this adjustment to account for each individual's probability of exiting employment exit between *year t* and *t+1*

Boosts the representation of:

- Male employees
- Younger employees
- Those with low job tenure
- Private sector employees

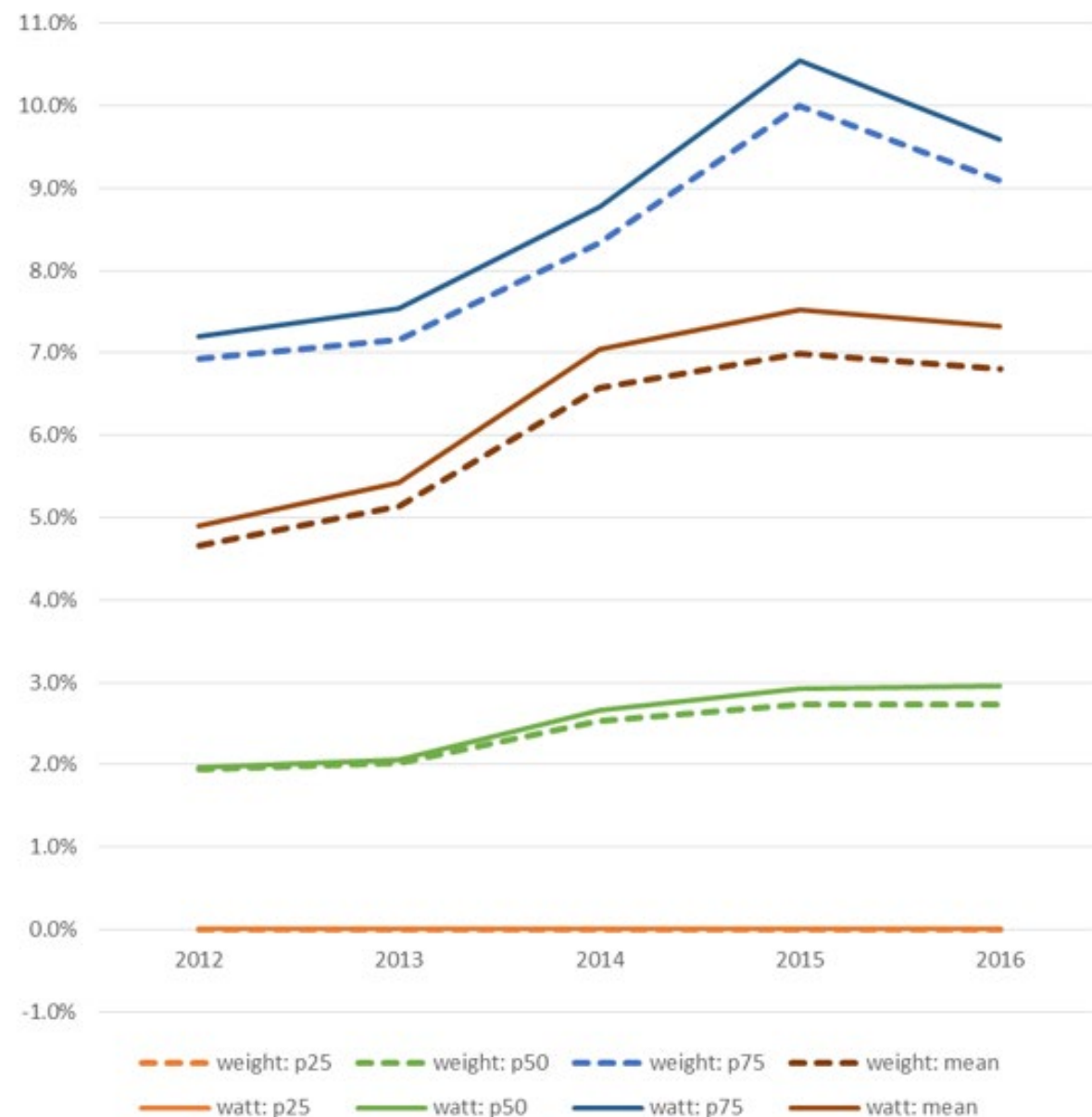
Sample	Observed in Year t and t+1	Observed in Year t and t+1	Difference: Col (2) – Col (1)
Weights	Cross-sectional	Longitudinal	
	(1)	(2)	(3)
Gender of the employee:			
Female	49.8%	49.3%	-0.5%
Male	50.2%	50.7%	0.5%
Age of the employee:			
16-19	2.9%	3.2%	0.4%
20-24	7.8%	8.8%	1.0%
25-29	11.1%	11.9%	0.7%
30-34	11.7%	12.1%	0.4%
35-39	11.4%	11.3%	0.0%
40-44	12.6%	12.3%	-0.3%
45-49	13.8%	13.3%	-0.6%
50-54	12.5%	11.8%	-0.7%
55-59	9.3%	8.8%	-0.5%
60-64	4.8%	4.6%	-0.3%
65+	2.1%	1.9%	-0.1%
Job tenure:			
<1 year	14.0%	16.1%	2.0%
1-2 years	11.6%	12.3%	0.8%
2-5 years	21.2%	21.4%	0.3%
5-9 years	21.5%	20.7%	-0.9%
10-20 years	20.1%	18.7%	-1.4%
20+ years	9.8%	8.9%	-0.9%
Missing/ invalid	1.8%	2.0%	0.1%
Sector of ownership:			
Private	73.9%	75.6%	1.7%
Public	26.1%	24.4%	-1.7%



Illustration

- We look at the distribution of annual change in nominal gross hourly wages (within-person)
- Accounting for sample attrition brings about a small widening of the distribution

Figure 8: Distribution of annual growth in nominal gross hourly wages, 2012-2017, under different weighting regimes



Recap

- Employees are three times more likely to exit ASHE year-on-year than they are to exit employment
 - ASHE sample exit rate = 0.25 p.a.
 - APS employment exit rate = 0.08 p.a.
 - -> Sample attrition rate = 0.17 p.a.
- Sample attrition is non-random:
 - Relatively high among those: with low wages/hours; aged 20-44; in private-sector services; in Lon/SE
 - Higher in later years of ASHE
- So sample attrition *may be non-ignorable* in any analysis of longitudinal outcomes (wage progression, job displacement, employment exit etc)
- For analysis of year-to-year changes among continuing sample members: we have used a comparison of ASHE and APS to derive year-to-year weights to correct for observable attrition biases
- For analysis of changes over longer periods, or analysis of job displacement/employment exit: strong assumptions needed – proceed with caution ... until we obtain data on employment transitions for sample members

Extra slides



Continuous sample retention rate

Table 2: Continuous Sample Retention rate in ASHE from year t to year $t+i$

Year (T)	T+1	T+2	T+3	T+4	T+5	T+6	T+7	T+8	T+9	T+10	T+11	T+12	T+13	T+14	T+15
2004	77.3	63.1	43.1	36.1	30.9	26.3	22.8	19.5	17.1	15.1	13.0	11.1	9.5	7.7	6.4
2005	78.1	52.4	43.1	36.5	30.8	26.6	22.6	19.9	17.4	14.9	12.7	10.9	8.8	7.3	
2006	64.1	51.3	42.8	35.8	30.7	26.0	22.7	19.8	16.9	14.3	12.2	9.9	8.1		
2007	75.7	61.5	50.8	43.3	36.3	31.5	27.4	23.2	19.6	16.6	13.5	11.1			
2008	77.2	62.3	52.5	43.6	37.6	32.5	27.5	23.1	19.5	15.9	13.1				
2009	77.4	63.7	52.2	44.4	38.0	32.0	26.7	22.4	18.3	15.1					
2010	78.6	62.8	52.6	44.7	37.2	30.9	25.8	21.0	17.3						
2011	76.0	62.1	51.9	42.7	35.1	29.1	23.4	19.2							
2012	77.8	63.6	51.6	42.0	34.6	27.7	22.6								
2013	78.1	61.9	49.5	40.3	32.1	26.1									
2014	75.0	58.4	46.6	36.9	29.7										
2015	73.7	57.1	44.6	35.4											
2016	73.4	56.1	43.9												
2017	72.7	55.2													
2018	71.3														
Average	75.1	59.4	48.1	40.1	33.9	28.9	24.6	21.0	18.0	15.3	12.9	10.7	8.8	7.5	6.4

Note: Shading is used to indicate the scale of sample retention: darker green for higher rates, amber for middling rates, then darker shades of red for lower rates.

Source: ASHE