Simulated dataset: mode choice	Estimation subset	Holdout subset
	Individuals = 400	Individuals = 100
	Observations = 4,000	Observations = 1,000
Variable	**Description**	**Values (for estimation subset)**
ID	Unique individual ID	1 to 400
choice	choice variable	1 for car, 2 for air, 3 for rail, 4 for hsr
car_time	travel time (mins) for car	Min: 330, mean: 437.7, max: 550
car_cost	travel cost (£) for car	Min: 35, mean: 57.31, max: 80
air_time	travel time (mins) for air	Min: 45, mean: 82.85, max: 120
air_cost	travel cost (£) for air	Min: 70, mean: 120.4, max: 170
air_access	access time (mins) for air	Min: 20, mean: 55.4, max: 90
rail_time	travel time (mins) for rail	Min: 300, mean: 473.1, max: 650
rail_cost	travel cost (£) for rail	Min: 35, mean: 62.05, max: 90
rail_access	access time (mins) for rail	Min: 5, mean: 24.9, max: 45
hsr_time	travel time (mins) for high speed rail	Min: 120, mean: 206.3, max: 290
hsr_cost	travel cost (£) for high speed rail	Min: 70, mean: 115.4, max: 160
hsr_access	access time (mins) for high speed rail	Min: 10, mean: 35.3, max: 60
income	household income (£ per month after tax)	Min: 800, mean: 1738, max: 2700
female	dummy variable for female	1 for female, 0 for male