Online Appendix to Accompany:

**Hurry Up and Weight, etc.:
Auxiliary Population Samples to Correct for
Systematic Selection in Citizen Science Data**

Contents

List of Figures

List of Tables

A Appendix: Ordered-Probit eBird Engagement-Level Models for all Relevant Subsets of Regressors (qBus Sample)

Table A1: qBus sample: Model 1-3 (of 24) to accommodate eBird missing values

	Model 1	Model 2	Model 3
engagement			
travel1mile			1.087***
			(0.0967)
empl_part		-0.0259	
		(0.0641)	
empl_look		-0.0910	
		(0.0933)	
empl_unem		-0.342***	
		(0.0696)	
empl_retired		-0.885***	
		(0.0750)	
ed_hs		0.132**	
		(0.0638)	
ed_somecoll		-0.124**	
		(0.0585)	
ed_master		0.358***	
		(0.0724)	
ed_doctoral		0.376***	
		(0.110)	
cut1	0.850***	0.719***	1.832***
	(0.0222)	(0.0458)	(0.0940)
cut2	1.203***	1.092***	2.201***
	(0.0255)	(0.0476)	(0.0955)

Continued on next page

Table A1 – continued from previous page

cut3	1.384***	1.286***	2.389***
	(0.0280)	(0.0490)	(0.0966)
cut4	1.604***	1.523***	2.617***
	(0.0319)	(0.0515)	(0.0982)
cut5	1.977***	1.921***	2.998***
	(0.0420)	(0.0588)	(0.103)
Observations	4161	4161	4161
Max. log-likelihood	-3287.24	-3166.85	-3198.14

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: qBus sample: Model 4-6 (of 24) to accommodate eBird missing values

	Model 4	Model 5	Model 6
engagement			
empl_part		-0.0851	-0.0513
		(0.0641)	(0.0651)
empl_look		-0.134	-0.105
		(0.0939)	(0.0953)
empl_unem		-0.383***	-0.349***
		(0.0681)	(0.0709)
empl_retired		-0.888***	-0.898***
		(0.0758)	(0.0766)
travel1mile	1.085***	1.099***	1.095***
	(0.0970)	(0.0987)	(0.0989)
ed_hs	0.0485		0.120*
	(0.0619)		(0.0648)
ed_somecoll	-0.166***		-0.132**
	(0.0580)		(0.0596)
ed_master	0.328***		0.355***
	(0.0725)		(0.0739)
ed_doctoral	0.327***		0.360***
	(0.110)		(0.112)
cut1	1.842***	1.645***	1.697***
	(0.100)	(0.0973)	(0.103)
cut2	2.216***	2.028***	2.086***
	(0.102)	(0.0988)	(0.105)
cut3	2.408***	2.225***	2.287***
	(0.103)	(0.0999)	(0.106)
cut4	2.642***	2.464***	2.533***
	(0.105)	(0.101)	(0.107)

Continued on next page

Table A2 – continued from previous page

cut5	3.034***	2.863***	2.942***
	(0.109)	(0.106)	(0.112)
Observations	4161	4161	4161
Max. log-likelihood	-3168.54	-3110.82	-3081.53

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3: qBus sample: Model 7-9 (of 24) to accommodate eBird missing values

	Model 7	Model 8	Model 9
engagement			
age_16_24	0.712***	0.753***	0.728***
	(0.0909)	(0.0918)	(0.0895)
age_25_34	0.710***	0.711***	0.716***
	(0.0797)	(0.0803)	(0.0798)
age_35_44	0.549***	0.540***	0.541***
	(0.0821)	(0.0827)	(0.0820)
age_55_64	-0.394***	-0.398***	-0.424***
	(0.103)	(0.104)	(0.101)
age_65up	-0.380***	-0.453***	-0.545***
	(0.124)	(0.126)	(0.110)
travel1mile	1.124***	1.127***	1.124***
	(0.102)	(0.103)	(0.102)
empl_part	-0.168**	-0.118*	
	(0.0684)	(0.0694)	
empl_look	-0.286***	-0.238**	
	(0.0970)	(0.0983)	
empl_unem	-0.470***	-0.410***	
	(0.0708)	(0.0734)	
empl_retired	-0.232**	-0.198**	
	(0.0990)	(0.101)	
ed_hs		0.0198	-0.0185
		(0.0676)	(0.0694)
ed_somecoll		-0.197***	-0.210***
		(0.0618)	(0.0627)
ed_master		0.383***	0.390***
		(0.0759)	(0.0769)
ed_doctoral		0.344***	0.359***

Continued on next page

Table A3 – continued from previous page

		(0.115)	(0.116)
inc_lt25			-0.123
			(0.0777)
inc_25_50			-0.0243
			(0.0727)
inc_75_100			0.101
			(0.0803)
inc_100up			0.0175
			(0.0731)
cut1	2.060***	2.086***	2.162***
	(0.120)	(0.125)	(0.134)
cut2	2.470***	2.503***	2.578***
	(0.122)	(0.127)	(0.136)
cut3	2.681***	2.720***	2.793***
	(0.123)	(0.128)	(0.137)
cut4	2.935***	2.982***	3.053***
	(0.124)	(0.130)	(0.139)
cut5	3.354***	3.414***	3.479***
	(0.129)	(0.134)	(0.143)
Observations	4161	4161	4161
Max. log-likelihood	-2979.69	-2945.35	-2959.28

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A4: qBus sample: Model 10-12 (of 24) to accommodate eBird missing values

	Model 10	Model 11	Model 12
engagement			
age_16_24	0.761***		
	(0.0924)		
age_25_34	0.716***		
	(0.0809)		
age_35_44	0.539***		
	(0.0828)		
age_55_64	-0.396***		
	(0.104)		
age_65up	-0.455***		
	(0.127)		
travel1mile	1.128***	1.091***	1.087***
	(0.103)	(0.0969)	(0.0970)
empl_part	-0.107		
	(0.0706)		
empl_look	-0.225**		
	(0.0997)		
empl_unem	-0.395***		
	(0.0756)		
empl_retired	-0.189*		
	(0.101)		
ed_hs	0.0384		0.0643
	(0.0704)		(0.0621)
ed_somecoll	-0.185***		-0.146**
	(0.0631)		(0.0582)
ed_master	0.381***		0.314***
	(0.0771)		(0.0727)
ed_doctoral	0.350***		0.285***

Continued on next page

Table A4 – continued from previous page

	(0.117)		(0.110)
inc_lt25	-0.0273		
	(0.0802)		
inc_25_50	-0.00603		
	(0.0732)		
inc_75_100	0.0949		
	(0.0805)		
inc_100up	0.0113		
	(0.0734)		
male		0.244^{***}	0.207^{***}
		(0.0438)	(0.0445)
cut1	2.112^{***}	1.960^{***}	1.956^{***}
	(0.135)	(0.0972)	(0.104)
cut2	2.530^{***}	2.331^{***}	2.331^{***}
	(0.137)	(0.0988)	(0.105)
cut3	2.747^{***}	2.521^{***}	2.525^{***}
	(0.138)	(0.0999)	(0.106)
cut4	3.010^{***}	2.752^{***}	2.761^{***}
	(0.140)	(0.102)	(0.108)
cut5	3.442^{***}	3.139^{***}	3.157^{***}
	(0.144)	(0.106)	(0.113)
Observations	4161	4161	4161
Max. log-likelihood	-2944.15	-3182.55	-3157.73

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5: qBus sample: Model 13-15 (of 24) to accommodate eBird missing values

	Model 13	Model 14	Model 15
engagement			
male	0.236***	0.206***	0.325***
	(0.0457)	(0.0462)	(0.0459)
travel1mile	1.095***	1.090***	
	(0.0988)	(0.0989)	
empl_part	-0.0402	-0.0149	
	(0.0649)	(0.0658)	
empl_look	-0.0910	-0.0696	
	(0.0945)	(0.0959)	
empl_unem	-0.311***	-0.290***	
	(0.0696)	(0.0722)	
empl_retired	-0.884***	-0.894***	
	(0.0758)	(0.0765)	
ed_hs		0.120*	-0.0419
		(0.0649)	(0.0647)
ed_somecoll		-0.119**	-0.201***
		(0.0597)	(0.0602)
ed_master		0.346***	0.382***
		(0.0740)	(0.0746)
ed_doctoral		0.323***	0.315***
		(0.112)	(0.113)
age_16_24			0.713***
			(0.0875)
age_25_34			0.722***
			(0.0783)
age_35_44			0.532***
			(0.0808)
age_55_64			-0.442***

Continued on next page

Table A5 – continued from previous page

			(0.0998)
age_65up			-0.583***
			(0.108)
/			
cut1	1.780***	1.813***	1.303***
	(0.101)	(0.107)	(0.0811)
cut2	2.166***	2.204***	1.705***
	(0.103)	(0.108)	(0.0828)
cut3	2.365***	2.407***	1.913***
	(0.104)	(0.109)	(0.0841)
cut4	2.607***	2.654***	2.168***
	(0.106)	(0.111)	(0.0864)
cut5	3.010***	3.067***	2.590***
	(0.110)	(0.116)	(0.0923)
Observations	4161	4161	4161
Max. log-likelihood	-3097.45	-3071.59	-3021.49

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A6: qBus sample: Model 16-18 (of 24) to accommodate eBird missing values

	Model 16	Model 17	Model 18
engagement			
male	0.288***	0.288***	0.360***
	(0.0468)	(0.0470)	(0.0461)
age_16_24	0.740***	0.740***	0.659***
	(0.0906)	(0.0911)	(0.0877)
age_25_34	0.720***	0.717***	0.724***
	(0.0792)	(0.0798)	(0.0791)
age_35_44	0.525***	0.525***	0.557***
	(0.0815)	(0.0815)	(0.0818)
age_55_64	-0.423***	-0.419***	-0.439***
	(0.102)	(0.102)	(0.101)
age_65up	-0.505***	-0.505***	-0.515***
	(0.125)	(0.125)	(0.108)
empl_part	-0.0381	-0.0369	
	(0.0693)	(0.0703)	
empl_look	-0.178*	-0.176*	
	(0.0970)	(0.0983)	
empl_unem	-0.329***	-0.326***	
	(0.0735)	(0.0756)	
empl_retired	-0.160	-0.161	
	(0.0992)	(0.1000)	
ed_hs	0.0297	0.0310	
	(0.0668)	(0.0697)	
ed_somecoll	-0.172***	-0.172***	
	(0.0609)	(0.0622)	
ed_master	0.374***	0.382***	
	(0.0747)	(0.0758)	
ed_doctoral	0.309***	0.328***	

Continued on next page

Table A6 – continued from previous page

	(0.113)	(0.115)	
inc_lt25		-0.00307	
		(0.0796)	
inc_25_50		0.00679	
		(0.0722)	
inc_75_100		0.0658	
		(0.0789)	
inc_100up		-0.0290	
		(0.0721)	
travel1mile			1.110***
			(0.102)
/			
cut1	1.237***	1.242***	2.346***
	(0.0829)	(0.0956)	(0.121)
cut2	1.641***	1.646***	2.758***
	(0.0846)	(0.0970)	(0.123)
cut3	1.850***	1.856***	2.969***
	(0.0858)	(0.0981)	(0.125)
cut4	2.107***	2.113***	3.224***
	(0.0880)	(0.100)	(0.127)
cut5	2.533***	2.539***	3.644***
	(0.0938)	(0.106)	(0.131)
Observations	4161	4161	4161
Max. log-likelihood	-3009.86	-3009.03	-2974.66

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A7: qBus sample: Model 19-21 (of 24) to accommodate eBird missing values

	Model 19	Model 20	Model 21
engagement			
male	0.319***	0.308***	0.279***
	(0.0467)	(0.0471)	(0.0476)
age_16_24	0.729***	0.726***	0.765***
	(0.0891)	(0.0914)	(0.0923)
age_25_34	0.726***	0.730***	0.730***
	(0.0797)	(0.0802)	(0.0807)
age_35_44	0.546***	0.553***	0.544***
	(0.0824)	(0.0825)	(0.0831)
age_55_64	-0.440***	-0.413***	-0.416***
	(0.102)	(0.104)	(0.105)
age_65up	-0.580***	-0.435***	-0.501***
	(0.110)	(0.125)	(0.127)
travel1mile	1.113***	1.116***	1.117***
	(0.102)	(0.103)	(0.103)
ed_hs	-0.0637		0.0106
	(0.0657)		(0.0678)
ed_somecoll	-0.215***		-0.185***
	(0.0613)		(0.0620)
ed_master	0.381***		0.372***
	(0.0760)		(0.0761)
ed_doctoral	0.299***		0.294**
	(0.115)		(0.115)
empl_part		-0.113	-0.0700
		(0.0693)	(0.0702)
empl_look		-0.239**	-0.196**
		(0.0978)	(0.0990)
empl_unem		-0.384***	-0.334***

Continued on next page

Table A7 – continued from previous page

		(0.0722)	(0.0747)
empl_retired		-0.190*	-0.160
		(0.0995)	(0.101)
/			
cut1	2.304***	2.236***	2.240***
	(0.127)	(0.124)	(0.128)
cut2	2.724***	2.651***	2.661***
	(0.129)	(0.125)	(0.130)
cut3	2.941***	2.864***	2.880***
	(0.130)	(0.127)	(0.132)
cut4	3.205***	3.123***	3.147***
	(0.132)	(0.129)	(0.133)
cut5	3.639***	3.549***	3.584***
	(0.137)	(0.133)	(0.138)
Observations	4161	4161	4161
Max. log-likelihood	-2939.44	-2958.28	-2928.12

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A8: qBus sample: Model 22-24 (of 24) to accommodate eBird missing values

	Model 22	Model 23	Model 24
engagement			
male	0.333***	0.310***	0.297***
	(0.0467)	(0.0471)	(0.0474)
age_16_24	0.714***	0.746***	0.748***
	(0.0892)	(0.0900)	(0.0921)
age_25_34	0.752***	0.733***	0.747***
	(0.0796)	(0.0802)	(0.0806)
age_35_44	0.554***	0.546***	0.551***
	(0.0819)	(0.0825)	(0.0826)
age_55_64	-0.444***	-0.437***	-0.421***
	(0.101)	(0.102)	(0.104)
age_65up	-0.523***	-0.580***	-0.451***
	(0.108)	(0.110)	(0.125)
travel1mile	1.121***	1.114***	1.123***
	(0.102)	(0.102)	(0.103)
inc_lt25	-0.0912	-0.0792	0.00449
	(0.0761)	(0.0782)	(0.0792)
inc_25_50	-0.0313	-0.0132	-0.00989
	(0.0726)	(0.0730)	(0.0730)
inc_75_100	0.148*	0.0854	0.140*
	(0.0796)	(0.0805)	(0.0798)
inc_100up	0.140**	-0.00488	0.127*
	(0.0701)	(0.0735)	(0.0703)
ed_hs		-0.0298	
		(0.0698)	
ed_somecoll		-0.198***	
		(0.0629)	
ed_master		0.381***	

Continued on next page

Table A8 – continued from previous page

		(0.0772)	
ed_doctoral		0.308***	
		(0.117)	
empl_part			-0.0892
			(0.0708)
empl_look			-0.208**
			(0.0997)
empl_unem			-0.354***
			(0.0755)
empl_retired			-0.162
			(0.100)
/			
cut1	2.390***	2.312***	2.308***
	(0.131)	(0.136)	(0.133)
cut2	2.803***	2.732***	2.724***
	(0.133)	(0.138)	(0.135)
cut3	3.016***	2.950***	2.938***
	(0.134)	(0.139)	(0.136)
cut4	3.274***	3.215***	3.199***
	(0.136)	(0.141)	(0.138)
cut5	3.698***	3.648***	3.627***
	(0.141)	(0.146)	(0.142)
Observations	4161	4161	4161
Max. log-likelihood	-2966.74	-2937.53	-2954.50

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B Appendix: Ordered-Probit eBird Engagement-Level Models for all Relevant Subsets of Regressors (eBird Sample)

Table A9: eBird sample: Model 1-3 (of 24) to accommodate missing values

	Model 1	Model 2	Model 3
/			
cut1	-0.260***	0.0773	0.806*
	(0.0370)	(0.196)	(0.433)
cut2	0.448***	0.808***	1.546***
	(0.0379)	(0.197)	(0.433)
cut3	1.046***	1.421***	2.158***
	(0.0448)	(0.199)	(0.435)
engagement			
hav_empl		0.278***	
		(0.0987)	
empl_part		-0.194	
		(0.135)	
empl_look		-0.553	
		(0.376)	
empl_unem		-0.133	
		(0.154)	
empl_retired		-0.462***	
		(0.0777)	
hav_ed		0.287	
		(0.199)	
ed_hs		0.0525	
		(0.190)	
ed_somecoll		-0.173*	
		(0.103)	
ed_master		0.0961	

Continued on next page

Table A9 – continued from previous page

		(0.0801)	
ed_doctoral		0.303***	
		(0.111)	
hav_travel			-0.443
			(0.476)
travel1mile			1.600***
			(0.202)
Observations	1177	1177	1177
Max. log-likelihood	-1544.88	-1515.63	-1499.10

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A10: eBird sample: Model 4-6 (of 24) to accommodate missing values

	Model 4	Model 5	Model 6
engagement			
hav_travel	-0.571	-0.477	-0.592
	(0.482)	(0.476)	(0.481)
travel1mile	1.624***	1.544***	1.563***
	(0.204)	(0.203)	(0.205)
hav_ed	0.181		0.191
	(0.202)		(0.204)
ed_hs	0.290		0.239
	(0.195)		(0.197)
ed_somecoll	-0.173*		-0.155
	(0.104)		(0.105)
ed_master	0.0940		0.106
	(0.0801)		(0.0808)
ed_doctoral	0.326***		0.310***
	(0.111)		(0.111)
hav_empl		0.314***	0.271***
		(0.0985)	(0.0994)
empl_part		-0.232*	-0.199
		(0.135)	(0.135)
empl_look		-0.608	-0.584
		(0.374)	(0.376)
empl_unem		-0.127	-0.0683
		(0.155)	(0.157)
empl_retired		-0.415***	-0.398***
		(0.0782)	(0.0787)
/			
cut1	0.931**	0.780*	0.903**
	(0.453)	(0.435)	(0.454)
cut2	1.679***	1.533***	1.663***

Continued on next page

Table A10 – continued from previous page

	(0.454)	(0.435)	(0.455)
cut3	2.298***	2.154***	2.290***
	(0.455)	(0.437)	(0.456)
Observations	1177	1177	1177
Max. log-likelihood	-1488.93	-1483.72	-1474.82

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A11: eBird sample: Model 7-9 (of 24) to accommodate missing values

	Model 7	Model 8	Model 9
engagement			
hav_age	0.754**	0.810**	0.781**
	(0.303)	(0.325)	(0.326)
age_16_24	0.496*	0.578**	0.541**
	(0.266)	(0.272)	(0.270)
age_25_34	0.291*	0.362**	0.337**
	(0.149)	(0.151)	(0.153)
age_35_44	0.105	0.120	0.0926
	(0.135)	(0.136)	(0.136)
age_55_64	-0.177*	-0.197*	-0.237**
	(0.106)	(0.107)	(0.103)
age_65up	-0.159	-0.215*	-0.324***
	(0.127)	(0.128)	(0.103)
hav_travel	-0.732	-0.782	-0.789
	(0.485)	(0.490)	(0.492)
travel1mile	1.515***	1.522***	1.483***
	(0.203)	(0.205)	(0.205)
hav_empl	0.254**	0.198*	
	(0.103)	(0.104)	
empl_part	-0.235*	-0.182	
	(0.138)	(0.139)	
empl_look	-0.756**	-0.745*	
	(0.380)	(0.383)	
empl_unem	-0.116	-0.0333	
	(0.155)	(0.158)	
empl_retired	-0.265**	-0.196*	
	(0.107)	(0.109)	
hav_ed		-0.112	-0.172

Continued on next page

Table A11 – continued from previous page

		(0.222)	(0.221)
ed_hs		0.215	0.252
		(0.202)	(0.202)
ed_somecoll		-0.155	-0.140
		(0.106)	(0.106)
ed_master		0.170**	0.196**
		(0.0822)	(0.0822)
ed_doctoral		0.397***	0.420***
		(0.113)	(0.115)
hav_inc			0.183*
			(0.109)
inc_lt25			-0.0895
			(0.158)
inc_25_50			0.137
			(0.112)
inc_75_100			-0.149
			(0.117)
inc_100up			0.101
			(0.101)
/			
cut1	1.171**	1.135**	1.066**
	(0.468)	(0.478)	(0.479)
cut2	1.936***	1.911***	1.846***
	(0.469)	(0.479)	(0.480)
cut3	2.562***	2.544***	2.479***
	(0.470)	(0.480)	(0.482)
Observations	1177	1177	1177
Max. log-likelihood	-1471.72	-1460.29	-1457.42

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A12: eBird sample: Model 10-12 (of 24) to accommodate missing values

	Model 10	Model 11	Model 12
engagement			
hav_age	0.659**		
	(0.332)		
age_16_24	0.606**		
	(0.273)		
age_25_34	0.360**		
	(0.154)		
age_35_44	0.0972		
	(0.136)		
age_55_64	-0.190*		
	(0.107)		
age_65up	-0.205		
	(0.129)		
hav_travel	-0.805	-0.417	-0.502
	(0.490)	(0.492)	(0.495)
travel1mile	1.492***	1.562***	1.583***
	(0.205)	(0.204)	(0.206)
hav_inc	0.186*		
	(0.109)		
inc_lt25	-0.0527		
	(0.159)		
inc_25_50	0.137		
	(0.113)		
inc_75_100	-0.170		
	(0.118)		
inc_100up	0.0744		
	(0.102)		
hav_empl	0.188*		

Continued on next page

Table A12 – continued from previous page

	(0.105)		
empl_part	-0.189		
	(0.140)		
empl_look	-0.725*		
	(0.389)		
empl_unem	-0.0115		
	(0.159)		
empl_retired	-0.189*		
	(0.110)		
hav_ed	-0.128		0.237
	(0.222)		(0.212)
ed_hs	0.220		0.198
	(0.203)		(0.195)
ed_somecoll	-0.149		-0.196*
	(0.107)		(0.104)
ed_master	0.179**		0.0724
	(0.0825)		(0.0804)
ed_doctoral	0.400***		0.243**
	(0.115)		(0.112)
hav_male		-0.198	-0.294
		(0.229)	(0.241)
male		0.407***	0.384***
		(0.0665)	(0.0673)
/			
cut1	1.084**	0.764*	0.855*
	(0.477)	(0.459)	(0.469)
cut2	1.867***	1.518***	1.617***
	(0.478)	(0.459)	(0.470)
cut3	2.504***	2.147***	2.250***

Continued on next page

Table A12 – continued from previous page

	(0.479)	(0.460)	(0.471)
Observations	1177	1177	1177
Max. log-likelihood	-1453.55	-1480.29	-1472.43

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A13: eBird sample: Model 13-15 (of 24) to accommodate missing values

	Model 13	Model 14	Model 15
engagement			
hav_male	-0.215	-0.306	-0.629**
	(0.232)	(0.244)	(0.287)
male	0.395***	0.378***	0.431***
	(0.0676)	(0.0683)	(0.0670)
hav_travel	-0.446	-0.519	
	(0.493)	(0.496)	
travel1mile	1.513***	1.526***	
	(0.204)	(0.206)	
hav_empl	0.279***	0.242**	
	(0.0996)	(0.100)	
empl_part	-0.144	-0.120	
	(0.136)	(0.136)	
empl_look	-0.494	-0.483	
	(0.374)	(0.376)	
empl_unem	-0.00182	0.0512	
	(0.156)	(0.159)	
empl_retired	-0.382***	-0.369***	
	(0.0787)	(0.0790)	
hav_ed		0.244	-0.0884
		(0.213)	(0.220)
ed_hs		0.138	-0.0454
		(0.198)	(0.194)
ed_somecoll		-0.182*	-0.196*
		(0.105)	(0.105)
ed_master		0.0887	0.168**
		(0.0811)	(0.0814)
ed_doctoral		0.238**	0.347***

Continued on next page

Table A13 – continued from previous page

		(0.113)	(0.113)
hav_age			1.412***
			(0.377)
age_16_24			0.562**
			(0.269)
age_25_34			0.345**
			(0.149)
age_35_44			0.143
			(0.135)
age_55_64			-0.255**
			(0.102)
age_65up			-0.416***
			(0.100)
/			
cut1	0.726	0.820*	0.473
	(0.458)	(0.469)	(0.292)
cut2	1.492***	1.593***	1.237***
	(0.459)	(0.470)	(0.293)
cut3	2.128***	2.234***	1.872***
	(0.460)	(0.471)	(0.295)
Observations	1177	1177	1177
Max. log-likelihood	-1466.60	-1459.31	-1479.87

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A14: eBird sample: Model 16-18 (of 24) to accommodate missing values

	Model 16	Model 17	Model 18
engagement			
hav_male	-0.621**	-0.643**	-0.745**
	(0.287)	(0.289)	(0.291)
male	0.422***	0.405***	0.424***
	(0.0682)	(0.0688)	(0.0668)
hav_age	1.301***	1.142***	1.310***
	(0.382)	(0.388)	(0.368)
age_16_24	0.609**	0.641**	0.353
	(0.272)	(0.273)	(0.264)
age_25_34	0.372**	0.370**	0.256*
	(0.151)	(0.154)	(0.148)
age_35_44	0.154	0.131	0.102
	(0.136)	(0.136)	(0.135)
age_55_64	-0.212**	-0.202*	-0.244**
	(0.106)	(0.107)	(0.102)
age_65up	-0.311**	-0.293**	-0.347***
	(0.128)	(0.129)	(0.101)
hav_empl	0.141	0.133	
	(0.104)	(0.105)	
empl_part	-0.0657	-0.0756	
	(0.140)	(0.140)	
empl_look	-0.610	-0.594	
	(0.384)	(0.390)	
empl_unem	0.0480	0.0678	
	(0.157)	(0.158)	
empl_retired	-0.147	-0.141	
	(0.108)	(0.109)	
hav_ed	-0.0622	-0.0793	

Continued on next page

Table A14 – continued from previous page

	(0.222)	(0.222)	
ed_hs	-0.0695	-0.0547	
	(0.195)	(0.197)	
ed_somecoll	-0.201*	-0.195*	
	(0.105)	(0.106)	
ed_master	0.156*	0.166**	
	(0.0818)	(0.0822)	
ed_doctoral	0.330***	0.337***	
	(0.114)	(0.115)	
hav_inc		0.220**	
		(0.108)	
inc_lt25		-0.0559	
		(0.157)	
inc_25_50		0.117	
		(0.112)	
inc_75_100		-0.174	
		(0.117)	
inc_100up		0.0632	
		(0.101)	
hav_travel			-0.670
			(0.497)
travel1mile			1.484***
			(0.204)
/			
cut1	0.489*	0.477	1.021**
	(0.293)	(0.293)	(0.476)
cut2	1.255***	1.248***	1.800***
	(0.294)	(0.294)	(0.477)
cut3	1.892***	1.889***	2.440***

Continued on next page

Table A14 – continued from previous page

	(0.295)	(0.296)	(0.478)
Observations	1177	1177	1177
Max. log-likelihood	-1477.17	-1469.96	-1455.64

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A15: eBird sample: Model 19-21 (of 24) to accommodate missing values

	Model 19	Model 20	Model 21
engagement			
hav_male	-0.767***	-0.737**	-0.763***
	(0.295)	(0.291)	(0.295)
male	0.397***	0.408***	0.390***
	(0.0676)	(0.0681)	(0.0688)
hav_age	1.365***	1.182***	1.256***
	(0.384)	(0.372)	(0.389)
age_16_24	0.465*	0.424	0.523*
	(0.270)	(0.267)	(0.273)
age_25_34	0.322**	0.292*	0.355**
	(0.150)	(0.150)	(0.151)
age_35_44	0.109	0.113	0.123
	(0.136)	(0.136)	(0.136)
age_55_64	-0.251**	-0.196*	-0.213**
	(0.103)	(0.107)	(0.107)
age_65up	-0.364***	-0.224*	-0.273**
	(0.101)	(0.128)	(0.129)
hav_travel	-0.701	-0.686	-0.720
	(0.502)	(0.495)	(0.500)
travel1mile	1.481***	1.491***	1.489***
	(0.206)	(0.204)	(0.206)
hav_ed	-0.0950		-0.0742
	(0.222)		(0.224)
ed_hs	0.158		0.128
	(0.201)		(0.203)
ed_somecoll	-0.170		-0.178*
	(0.106)		(0.106)
ed_master	0.173**		0.160*

Continued on next page

Table A15 – continued from previous page

	(0.0821)		(0.0825)
ed_doctoral	0.349***		0.336***
	(0.114)		(0.114)
hav_empl		0.204**	0.153
		(0.104)	(0.104)
empl_part		-0.128	-0.0843
		(0.139)	(0.140)
empl_look		-0.642*	-0.638*
		(0.381)	(0.384)
empl_unem		0.0173	0.0943
		(0.157)	(0.160)
empl_retired		-0.184*	-0.125
		(0.108)	(0.110)
/			
cut1	0.997**	1.031**	1.010**
	(0.485)	(0.474)	(0.483)
cut2	1.787***	1.813***	1.803***
	(0.486)	(0.475)	(0.484)
cut3	2.433***	2.456***	2.452***
	(0.487)	(0.476)	(0.485)
Observations	1177	1177	1177
Max. log-likelihood	-1445.06	-1451.87	-1442.05

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A16: eBird sample: Model 22-24 (of 24) to accommodate missing values

	Model 22	Model 23	Model 24
engagement			
hav_male	-0.773***	-0.779***	-0.759***
	(0.293)	(0.297)	(0.293)
male	0.403***	0.382***	0.392***
	(0.0676)	(0.0683)	(0.0686)
hav_age	1.186***	1.238***	1.062***
	(0.375)	(0.390)	(0.378)
age_16_24	0.393	0.489*	0.456*
	(0.265)	(0.271)	(0.268)
age_25_34	0.287*	0.321**	0.310**
	(0.152)	(0.153)	(0.153)
age_35_44	0.0889	0.0881	0.0983
	(0.136)	(0.136)	(0.136)
age_55_64	-0.231**	-0.243**	-0.185*
	(0.103)	(0.103)	(0.107)
age_65up	-0.321***	-0.353***	-0.206
	(0.102)	(0.103)	(0.128)
hav_travel	-0.683	-0.720	-0.702
	(0.495)	(0.501)	(0.494)
travel1mile	1.451***	1.456***	1.463***
	(0.205)	(0.207)	(0.205)
hav_inc	0.171	0.171	0.176
	(0.109)	(0.110)	(0.109)
inc_lt25	-0.126	-0.0557	-0.0958
	(0.157)	(0.159)	(0.158)
inc_25_50	0.0650	0.118	0.0626
	(0.111)	(0.113)	(0.111)
inc_75_100	-0.173	-0.173	-0.195*

Continued on next page

Table A16 – continued from previous page

	(0.117)	(0.117)	(0.118)
inc_100up	0.105	0.0622	0.0809
	(0.100)	(0.101)	(0.101)
hav_ed		-0.107	
		(0.223)	
ed_hs		0.167	
		(0.202)	
ed_somecoll		-0.162	
		(0.107)	
ed_master		0.183**	
		(0.0825)	
ed_doctoral		0.355***	
		(0.116)	
hav_empl			0.195*
			(0.105)
empl_part			-0.130
			(0.140)
empl_look			-0.606
			(0.386)
empl_unem			0.0395
			(0.158)
empl_retired			-0.174
			(0.109)
/			
cut1	0.972**	0.954**	0.987**
	(0.475)	(0.484)	(0.473)
cut2	1.757***	1.749***	1.774***
	(0.475)	(0.485)	(0.474)
cut3	2.401***	2.398***	2.421***

Continued on next page

Table A16 – continued from previous page

	(0.477)	(0.486)	(0.475)
Observations	1177	1177	1177
Max. log-likelihood	-1449.53	-1439.37	-1446.09

t in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C Additional Figures

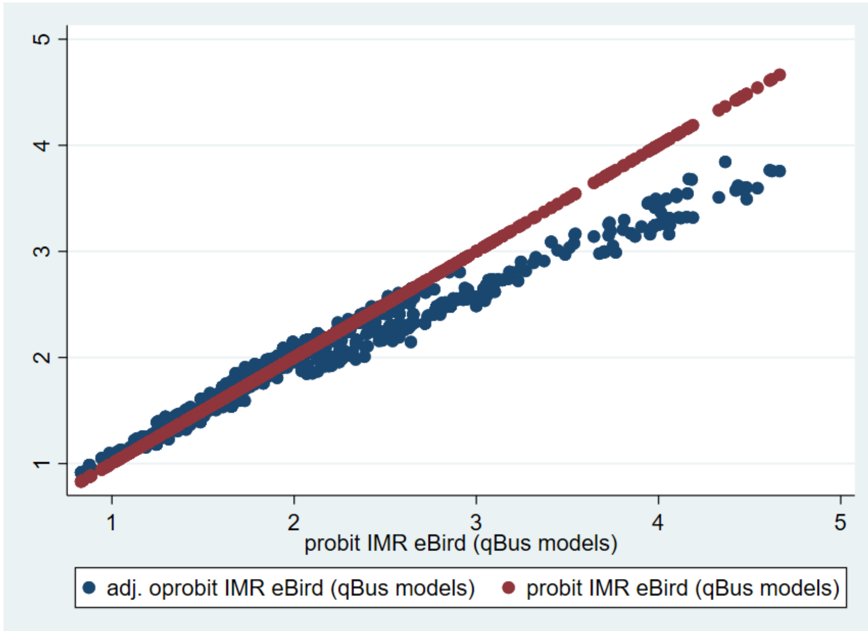

Figure A1: eBird sample: joint distribution of binary-probit-based IMRs and ordered-probit-based IMRs, both calculated from qBus parameters for the most-detailed specification consistent with any missing Z_j data in the eBird sample.

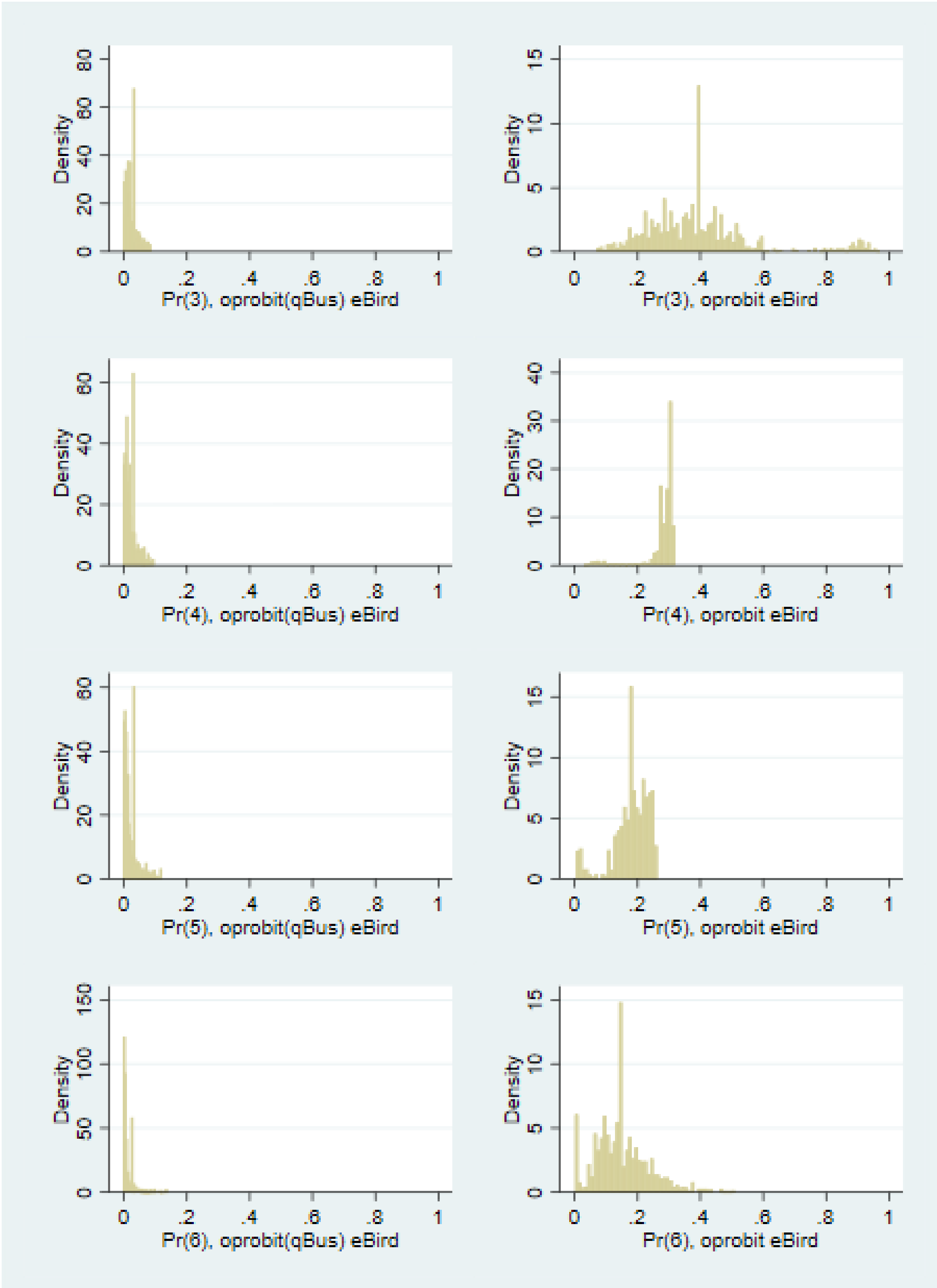

Figure A2: Compare eBird sample expected probabilities based on qBus models to fitted probabilities based on eBird models.

D Additional Tables of Results

Table A17: Market Extent Models with Intensity Weights and Sample Selection Corrections

ln(max distance)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1(Working)	0.00383	-0.0807	-0.187**	-0.218***	-0.169**	-0.197***	-0.170**	-0.206**
	(0.0754)	(0.0840)	(0.0809)	(0.0780)	(0.0775)	(0.0749)	(0.0820)	(0.0823)
1(Have self-reported income bracket)	-0.514***	-0.514***	-0.460***	-0.418**	-0.472***	-0.415**	-0.652***	-0.629***
	(0.158)	(0.187)	(0.155)	(0.169)	(0.153)	(0.167)	(0.177)	(0.198)
ln(Income in 10,000s)* 1(Have Income)	0.311***	0.307***	0.252***	0.234***	0.248***	0.227***	0.315***	0.303***
	(0.0629)	(0.0739)	(0.0623)	(0.0649)	(0.0616)	(0.0638)	(0.0805)	(0.0863)
1(Not Interested in Waterfowl)	-0.587***	-0.591**	-0.487**	-0.533**	-0.445**	-0.507**	-0.586***	-0.612***
	(0.210)	(0.243)	(0.206)	(0.235)	(0.204)	(0.241)	(0.206)	(0.235)
1(Not Interested in Wading Birds)	-0.185	-0.205	-0.0988	-0.126	-0.0962	-0.119	-0.152	-0.171
	(0.226)	(0.253)	(0.221)	(0.256)	(0.219)	(0.250)	(0.222)	(0.249)
1(Not Interested in Birds of Prey)	-0.220	-0.159	-0.0927	-0.0822	-0.0300	-0.0368	-0.145	-0.0906
	(0.240)	(0.250)	(0.236)	(0.238)	(0.235)	(0.235)	(0.236)	(0.251)
1(Not Interested in Perching Birds)	-0.460**	-0.451**	-0.439**	-0.406*	-0.463**	-0.417*	-0.400*	-0.389*
	(0.213)	(0.223)	(0.208)	(0.220)	(0.206)	(0.219)	(0.210)	(0.220)
1(Not Interested in Other Game Birds)	-0.345**	-0.351**	-0.375***	-0.372***	-0.380***	-0.380***	-0.379***	-0.381***
	(0.140)	(0.138)	(0.137)	(0.137)	(0.136)	(0.137)	(0.138)	(0.135)
IMR probit; for normal dep var in eBird sample			-0.361***	-0.318***				

Continued on next page

			(0.0632)	(0.0560)				
IMR oprobit; for normal dep var in eBird sample					-1.129***	-1.018***		
					(0.170)	(0.141)		
1(have inc data) * demeaned sel. prop.							0.578***	0.557***
							(0.208)	(0.203)
ln(Income in 10,000s)* 1(Have Income)* demeaned sel. prop.							-0.100	-0.110
							(0.108)	(0.107)
Constant	3.924***	3.968***	4.838***	4.753***	4.417***	4.382***	3.951***	3.988***
	(0.0946)	(0.105)	(0.185)	(0.173)	(0.118)	(0.118)	(0.0929)	(0.104)
lnsigma								
Constant	-0.0929***	-0.106***	-0.117***	-0.127***	-0.126***	-0.134***	-0.113***	-0.126***
	(0.0296)	(0.0323)	(0.0296)	(0.0337)	(0.0296)	(0.0340)	(0.0296)	(0.0333)
Observations	691	691	691	691	691	691	691	691
Log Likelihood	-1536.35	-1533.51	-1520.34	-1520.01	-1514.75	-1515.45	-1522.27	-1520.38
AIC	3092.70	3087.02	3062.69	3062.02	3051.50	3052.90	3068.54	3064.75
BIC	3138.08	3132.40	3112.61	3111.94	3101.41	3102.82	3123.00	3119.21
Weighted?	No	Yes	No	Yes	No	Yes	No	Yes

Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A18: Market Extent Models with Intensity Weights, Sample Selection Correction and Endogeneity Corrections

ln(max distance)	(1)	(2)	(3)	(4)	(5)	(6)
1(Working)	0.734	1.443	0.0860	0.134	0.183	0.272
	(1.163)	(1.814)	(0.209)	(0.348)	(0.308)	(0.663)
1(Have self-reported income bracket)	1.788	3.540	0.0810	0.277	-0.0337	0.0417
	(2.782)	(4.376)	(0.446)	(0.727)	(0.575)	(0.959)
ln(Income in 10,000s)* 1(Have Income)	-0.720	-1.473	0.00837	-0.0656	0.0316	0.00258
	(1.207)	(1.836)	(0.194)	(0.322)	(0.263)	(0.434)
1(Not Interested in Waterfowl)	-1.653	-2.643	-0.727**	-0.858*	-1.098**	-1.286
	(1.684)	(2.742)	(0.345)	(0.470)	(0.532)	(0.955)
1(Not Interested in Wading Birds)	2.591	5.161	0.573	0.767	0.844	1.197
	(3.433)	(5.876)	(0.586)	(1.025)	(0.860)	(1.918)
1(Not Interested in Birds of Prey)	-0.927	-1.995	-0.220	-0.350	-0.505	-0.616
	(1.436)	(2.113)	(0.344)	(0.428)	(0.494)	(0.783)
1(Not Interested in Perching Birds)	0.864	1.409	-0.140	-0.0556	0.0464	0.103
	(1.795)	(2.497)	(0.361)	(0.426)	(0.533)	(0.711)
1(Not Interested in Other Game Birds)	-1.288	-2.134	-0.611**	-0.702*	-0.703**	-0.862
	(1.248)	(1.999)	(0.245)	(0.360)	(0.354)	(0.655)
IMR probit; for normal dep var in eBird sample	-1.103	-1.835				
	(0.927)	(1.613)				

Continued on next page

IMR oprobit; for normal dep var in eBird sample			-1.625***	-1.746**		
			(0.419)	(0.696)		
1(have inc data) * demeaned sel. prop.					1.138**	1.550
					(0.568)	(1.469)
ln(Income in 10,000s)* 1(Have Income)* demeaned sel. prop.					-0.245	-0.408
					(0.219)	(0.521)
Hunt birds? 1=yes, 0=no	-6.244	-9.505	-1.624	-2.001	-1.854	-2.453
	(7.528)	(10.69)	(1.238)	(1.593)	(1.833)	(2.934)
1(Mode - Nonmotorized)	-15.71	-27.16	-3.941	-4.696	-5.927	-7.336
	(19.02)	(27.03)	(2.885)	(4.701)	(4.581)	(9.667)
Constant	7.403**	9.568*	4.807***	4.868***	4.221***	4.298***
	(3.170)	(4.923)	(0.320)	(0.462)	(0.264)	(0.415)
Hunt birds? 1=yes, 0=no						
ln(Mean Census Tract Travel Time)	-0.0480	-0.0415	-0.0586*	-0.0487	-0.0567	-0.0443
	(0.0339)	(0.0355)	(0.0345)	(0.0364)	(0.0349)	(0.0375)
1(Female)	-0.118***	-0.125***	-0.114***	-0.122***	-0.109***	-0.118***
	(0.0207)	(0.0245)	(0.0209)	(0.0254)	(0.0210)	(0.0250)
Age less than 45 yrs	0.0344	0.0381	0.0285	0.0304	0.0243	0.0340
	(0.0312)	(0.0352)	(0.0304)	(0.0340)	(0.0355)	(0.0369)
Age more than 65 yrs	-0.0796***	-0.0749***	-0.0693***	-0.0665***	-0.0624***	-0.0568**
	(0.0252)	(0.0276)	(0.0231)	(0.0246)	(0.0234)	(0.0250)
1(Working)	0.00555	0.0120	0.00440	0.0112	0.00489	0.0184

Continued on next page

	(0.0232)	(0.0257)	(0.0232)	(0.0257)	(0.0239)	(0.0275)
1(Have self-reported income bracket)	0.0554	0.0640	0.0571	0.0671	0.0421	0.0417
	(0.0429)	(0.0445)	(0.0428)	(0.0436)	(0.0483)	(0.0387)
ln(Income in 10,000s)* 1(Have Income)	-0.00793	-0.00969	-0.00861	-0.0110	0.000859	0.00710
	(0.0175)	(0.0202)	(0.0175)	(0.0198)	(0.0223)	(0.0195)
1(Not Interested in Waterfowl)	-0.00178	-0.00112	-0.00122	-0.000630	0.00490	0.00120
	(0.0552)	(0.0454)	(0.0553)	(0.0455)	(0.0550)	(0.0448)
1(Not Interested in Wading Birds)	-0.0641	-0.0787*	-0.0623	-0.0776*	-0.0563	-0.0720*
	(0.0596)	(0.0405)	(0.0595)	(0.0404)	(0.0594)	(0.0395)
1(Not Interested in Birds of Prey)	0.00195	-0.0161	-0.00120	-0.0187	0.0108	-0.0114
	(0.0625)	(0.0474)	(0.0628)	(0.0475)	(0.0624)	(0.0464)
1(Not Interested in Perching Birds)	0.148***	0.159**	0.152***	0.163**	0.147***	0.158**
	(0.0556)	(0.0757)	(0.0556)	(0.0759)	(0.0559)	(0.0757)
1(Not Interested in Other Game Birds)	-0.0384	-0.0429	-0.0399	-0.0435	-0.0408	-0.0434
	(0.0374)	(0.0316)	(0.0373)	(0.0317)	(0.0374)	(0.0319)
IMR probit; for normal dep var in eBird sample	0.0345	0.0319				
	(0.0223)	(0.0218)				
IMR oprobit; for normal dep var in eBird sample			0.0727	0.0675		
			(0.0536)	(0.0488)		
1(have inc data) * demeaned sel. prop.					0.0321	0.0582

Continued on next page

					(0.0585)	(0.0628)
ln(Income in 10,000s)* 1(Have Income)* demeaned sel. prop.					-0.0248	-0.0428
					(0.0291)	(0.0336)
Constant	0.205*	0.187	0.286**	0.254**	0.303**	0.258**
	(0.120)	(0.124)	(0.116)	(0.125)	(0.118)	(0.131)
1(Mode - Nonmotorized)						
ln(Mean Census Tract Travel Time)	0.0112	0.00994	-0.0206	-0.0181	-0.0116	-0.00927
	(0.0186)	(0.0141)	(0.0356)	(0.0429)	(0.0320)	(0.0405)
1(Female)	0.0477**	0.0422***	0.0515**	0.0442**	0.0428**	0.0384*
	(0.0208)	(0.0108)	(0.0210)	(0.0221)	(0.0213)	(0.0223)
Age less than 45 yrs	-0.00541	-0.00912	0.00560	0.00198	0.0147	0.00779
	(0.0170)	(0.0140)	(0.0251)	(0.0279)	(0.0238)	(0.0256)
Age more than 65 yrs	0.0133	0.0144	-0.00156	-0.00550	-0.00121	-0.00256
	(0.0232)	(0.0148)	(0.0204)	(0.0151)	(0.0197)	(0.0194)
1(Working)	0.0502**	0.0532**	0.0474**	0.0509**	0.0483**	0.0503**
	(0.0231)	(0.0221)	(0.0228)	(0.0229)	(0.0235)	(0.0254)
1(Have self-reported income bracket)	0.117***	0.121**	0.106**	0.110**	0.0851*	0.0750
	(0.0424)	(0.0494)	(0.0428)	(0.0500)	(0.0482)	(0.0477)
ln(Income in 10,000s)* 1(Have Income)	-0.0569***	-0.0585***	-0.0530***	-0.0549**	-0.0445**	-0.0411*
	(0.0173)	(0.0214)	(0.0174)	(0.0217)	(0.0221)	(0.0214)
1(Not Interested in Waterfowl)	-0.0757	-0.0783	-0.0736	-0.0778	-0.0883	-0.0920
	(0.0552)	(0.0669)	(0.0552)	(0.0673)	(0.0550)	(0.0681)

Continued on next page

1(Not Interested in Wading Birds)	0.196***	0.222**	0.195***	0.221**	0.185***	0.211**
	(0.0596)	(0.0952)	(0.0595)	(0.0954)	(0.0595)	(0.0965)
1(Not Interested in Birds of Prey)	-0.0533	-0.0646	-0.0465	-0.0581	-0.0623	-0.0666
	(0.0626)	(0.0465)	(0.0628)	(0.0470)	(0.0625)	(0.0478)
1(Not Interested in Perching Birds)	0.0242	0.0103	0.0244	0.00911	0.0317	0.0141
	(0.0556)	(0.0440)	(0.0556)	(0.0452)	(0.0559)	(0.0446)
1(Not Interested in Other Game Birds)	-0.0413	-0.0489	-0.0411	-0.0482	-0.0410	-0.0494
	(0.0374)	(0.0323)	(0.0373)	(0.0328)	(0.0375)	(0.0320)
IMR probit; for normal dep var in eBird sample	-0.0527**	-0.0611***				
	(0.0213)	(0.0171)				
IMR oprobit; for normal dep var in eBird sample			-0.133**	-0.145***		
			(0.0526)	(0.0400)		
1(have inc data) * demeaned sel. prop.					0.0675	0.0971
					(0.0578)	(0.0783)
ln(Income in 10,000s)* 1(Have Income)* demeaned sel. prop.					-0.0143	-0.0243
					(0.0291)	(0.0391)
Constant	0.0968	0.120*	0.130	0.130	0.0506	0.0462
	(0.0731)	(0.0630)	(0.118)	(0.137)	(0.112)	(0.139)
/						
var(e.ln_l_maxradius)	16.33	48.09	1.454	1.850	2.524	3.640
	(39.28)	(98.55)	(1.259)	(2.521)	(3.138)	(8.379)

Continued on next page

var(e.birdhunt)	0.0610***	0.0618***	0.0610***	0.0618***	0.0611***	0.0618***
	(0.00328)	(0.00758)	(0.00328)	(0.00758)	(0.00329)	(0.00754)
var(e.mode_nonmotor)	0.0612***	0.0607***	0.0610***	0.0606***	0.0613***	0.0607***
	(0.00329)	(0.00773)	(0.00328)	(0.00773)	(0.00330)	(0.00772)
corr(e.birdhunt,e.ln_l_maxradius)	0.344**	0.325***	0.272	0.320**	0.224	0.274
	(0.152)	(0.0559)	(0.171)	(0.163)	(0.186)	(0.171)
corr(e.mode_nonmotor,e.ln_l_maxradius)	0.905***	0.935***	0.641**	0.708*	0.799***	0.846***
	(0.0873)	(0.0202)	(0.313)	(0.376)	(0.215)	(0.276)
corr(e.mode_nonmotor,e.birdhunt)	-0.0277	-0.00961	-0.0292	-0.0111	-0.0335	-0.0160
	(0.0381)	(0.0564)	(0.0380)	(0.0415)	(0.0380)	(0.0392)
Observations	691	691	691	691	691	691
Log Likelihood	-1526.99	-1528.08	-1524.13	-1526.67	-1533.77	-1531.70
AIC	3145.99	3148.16	3140.27	3145.34	3165.54	3161.39
BIC	3354.74	3356.91	3349.02	3354.09	3387.91	3383.76
Weighted?	No	Yes	No	Yes	No	Yes

Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

E Digression: Legitimate use of Inverse Mills Ratios to correct for sample selection

Over the last several decades, empirical researchers have become accustomed to the idea that estimating a sample-selection model via maximum likelihood methods, calculating the IMR, and including that estimated IMF into the desired "outcome" equation of interest will (somehow) purge the parameters of that outcome equation of any bias due to sample selection. However, it is crucial to remember that the IMR offers an appropriate correction for sample-selection bias only under some very specific conditions. Confidence that "including an IMR term" will "fix" selection bias hinges on the assumption that the selection equation and the outcome equation have error terms that are jointly normally distributed.

The joint normality assumption is critical because the IMR correction derives entirely from the formula for the expected value of a singly truncated bivariate normal distribution. If the conditional distribution latent variable in the selection equation is not normal and the conditional distribution of the dependent variable in the outcome equation is not normal (either observed or censored in some way, perhaps after some transformation), then the needed expected value of the singly truncated joint distribution of the errors in the selection equation and the outcome equation cannot automatically be assumed to be given by the usual IMR formulas.

Ideally, selection and outcome equations should be estimated jointly, in which case a wide variety of joint distributions for the two error terms can be assumed/employed, provided that the joint density can be derived and written down. In some cases, it is convenient to write the conditional joint distributions of the selection propensity and the outcome variable as the product of a conditional distribution and a marginal distribution.[1]

This insight is especially relevant for researchers who wish to estimate conditional logit "outcome" models based on people's choices across alternatives with different attributes. Nothing stops the analyst from estimating a binary probit sample selection model and calculating the usual IMR term from the fitted parameters. However, there is no rigorous statistical rationale for including this fitted IMR term like other respondent characteristics as a variable that might shift one or more slope characteristics or the coefficient on the status quo indicator variable, as is done in Yuan et al. (2015). Some types of joint models where IMR correction terms can make sense, statistically, include the following:

- The usual OLS outcome regression with a continuous dependent variable that is con-

[1] Stata now includes the "heckpoisson" estimator, following Terza (1998). Appropriately, this estimator is available only as a FIML estimator, not as a two-step estimator that relies on an IMR term. Jointly distributed variables that are not both normal have also bee used in a FIML model that combines a participation/experience variable (that is distributed either Poisson or zero-inflated-Poisson) with a censored-normal outcome variable is estimated jointly in Cameron and Englin (1997).

ditionally normal, perhaps after some transformation

- A Tobit outcome model (censored anywhere, at the bottom, the top, or both) with a partially censored normal propensity variable
- An interval-data outcome variable censored between known thresholds
- An ordered probit outcome model with a normally distributed latent propensity variable
- A censored normal outcome model with different censored points across observations

Simply appending an estimated IMR variable to a second-step outcome equation of interest cannot be assumed to be correct in any of the following cases:

- Count data models: Poisson, negative binomial
- Conditional logit models: fixed or random parameters
- Any other statistical model for the "outcome" equation, where the (perhaps latent) dependent is not normal (or at least normal after transformation)

F Additional complications to estimating IMR: Dealing with missing values for Z_j variables in the eBird sample

F.1 qBus sociodemographic variables have few missing values

Any empirical application of this methodology may have to confront the problem of what to do when there are missing values of some variables in one sample or the other. If the correction is based upon the standard sociodemographic variables available for qBus panel members, the data for those variables can be expected to be relatively complete. Any missing values in the qBus sample might be expected to be missing at random.[2]

F.2 eBird sociodemographics match Census, but have more missing values

Missing values in the citizen-science eBird sample, for the sociodemographic variables that conform to the set available the qBus sample, are likely to be more of a problem. For example,

[2] If other key variables intended to serve as regressors, Z_i, in our weighting strategy have been posted to qBus participants, however, it is entirely possible that there may be item non-response for some of those variables. We do not consider that issue in this analysis, since we rely exclusively on the standard sociodemographics available for all qBus panelists.

due to time constraints for our survey of eBird members, we elected not to ask about individuals' political ideologies. Had we anticipated being able to employ qBus questions to build sampling weights and estimated response propensities, it would have been prudent to be sure that the citizen science members were asked *every* standard sociodemographic question, verbatim, that is available with the qBus responses. For this first example of our procedure, we can assemble conformable measures for gender, race, ethnicity, income brackets, four regions of the U.S., employment status and educational attainment. Some aggregation of categories has been required in each sample to produce matching categories. In future applications of this method, it would be prudent to minimize this type of aggregation. In the eBird data, we used categories that matched the U.S. Census, which would facilitate more-conventional comparisons of marginal distributions in the eBird sample to marginal distributions in the general population. However, the U.S. Census does not provide any information about engagement in citizen science, so our special-purpose qBus sample is much superior in that way.

F.3 Using maximal available Z_j regressors for each eBird observation

Suppose there were no data in the eBird sample on any of the same sociodemographic regressors, Z_i provided by the qBus sample. There would still be valuable information in the qBus sample that could help construct either probability weights or propensity corrections. If one runs an ordered probit model to explain the engagement outcome in the eBird data, but use *no* explanatory variables, the result is a set of estimates for only the three cut-points between the four outcome levels in in that eBird data. If one then calculates the predicted probabilities for each of the four participation intensities, the means of these probabilities, across the sample, match the proportions of the sample observed at each level.[3]

F.4 If there are no Z_j regressors available for some eBird respondents

With no Z_j regressors available for some (small) subset of observations in the eBird sample, the best available option for weighting the observations at each level of participation intensity must be derived solely from (a) the predicted probabilities for each of the four relevant participation-intensity levels in the qBus sample (also estimated without regressors) relative to (b) the analogous predicted probabilities for the same four participation intensity levels in the qBus sample. The model being used to predict participation intensities, in this case, has no Z regressors, so there is no basis for observable systematic heterogeneity in these probabilities. The weights will differ across the four observed participation intensity levels,

[3]For binary probit and logit models, the means of the fitted probabilities will be either extremely close to the observed proportions, or exactly equal to those proportions, as can be proven by the algebra of the first-order conditions for the maximum likelihood estimation algorithm.

but will be the same for every person who has no available Z variables in the eBird sample.

F.5 If only some subset of Z_j regressors is available for some eBird respondents

The most-general approach to weighting by participation intensity level or correcting parameters for different-from-average participation intensity would exploit the maximum information available in both samples, on an observation-by-observation basis for the eBird sample. To simplify, assume that only three basic factors are available as explanatory variables. In practice, each factor may be captured by a set of indicators for the categories of that factor, but we will assume for now that there is one continuous variable per factor such that the universe of potential Z variables consists of Z_1, Z_2, and Z_3. All three variables (groups of variables) are available for each qBus observation, but different observations in the eBird sample have missing values for either one, two, or all three of these variables.

To fully exploit the available information, it is necessary to estimate an array of models for the qBus sample so that one of these models will be appropriate to transfer to every observation in the eBird sample. Suppose that we have indicators for the presence or absence of values for each of these three Z variables in the eBird sample. The number of necessary models using the qBus data could then be calculated using the sum of all the relevant combinations:

$$C_0^3 + C_1^3 + C_2^3 + C_3^3 = 1 + 3 + 3 + 1 = 8 \tag{1}$$

Of course, as the number of potential factors increases, the number of potentially relevant models to explain participation intensities in the qBus data can increase dramatically. With six different factors potentially missing for at least some observations in the eBird sample, the number of potentially relevant models would be 64.

References

Cameron, T. A. and J. Englin (1997). Respondent experience and contingent valuation of environmental goods. *Journal of Environmental Economics and Management 33*(3), 296–313.

Terza, J. V. (1998). Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics 84*(1), 129–154.

Yuan, Y., K. J. Boyle, and W. You (2015). Sample selection, individual heterogeneity, and regional heterogeneity in valuing farmland conservation easements. *Land Economics 91*(4), 627–649.