Online Appendix for Urbanization in the developing world: Too early or too slow?

J. Vernon Henderson and Matthew A. Turner

J. Vernon Henderson is School Professor of Economic Geography, London School of Economics, London, United Kingdom. Henderson is also a Research Fellow of the Centre for Economic Policy Research. Matthew A. Turner is Professor of Economics, Brown University, Providence, Rhode Island. Turner is also Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts, and Senior Research Fellow, Property and Environment Research Center, Bozeman, Montana. Both authors are Research Affiliates, International Growth Center, London, United Kingdom. Their email addresses are J.V.Henderson@lse.ac.uk and matthew_turner@brown.edu.

Tables A1a and A1b present our regression results for the outcomes from the LSMS, DHS and Afrobarometer data, as discussed in the paper. For each outcome listed in the left column we report from left to right: coefficient of log density in a regression with no controls, the R-square of this regression, the corresponding coefficient and R-square from a regression that also includes the demographic controls listed at the bottom of each panel, the mean of the outcome and log population density in the no-controls regression sample, and finally, the count of survey respondents, survey clusters and countries upon which the no-controls regression sample is based.

Table A2 lists the countries covered by each of the three surveys, LSMS, DHS and Afrobarometer. Table A1's count of the countries on which each regression is based will sometimes be lower than that from Table A2. This primarily reflects the fact that some of the DHS survey units are conducted in only a subset of DHS countries. Throughout our analysis, we consistently use the largest set of survey respondents that is available for each particular question. As a consequence, some of the density gradients we report are based on quite different samples of countries. Given this, some caution is required in comparing regression results across outcomes. Refinements of these estimates are an obvious area for further work.

Table A1a: Density gradients for Afrobarometer, LSMS and DHS outcomes.

	No controls		Controls						
Outcome	β	R^2	β	R^2	$\overline{y}$	$\overline{x}$	N	Clusters	Countries
	s.e.		s.e.		s.e.	s.e.			
Data: *LSMS*									
ln(Income)	.3126^a	0.067	.3170^a	0.856	4.097	5.77	35,231	2,118	6
	(.0161)		(.0141)		(2.014)	(1.67)			
ln(Wage)	.1177^a	0.019	.0488^a	0.553	1.191	6.38	18,806	1,704	6
	(.0152)		(.0094)		(1.435)	(1.69)			
Controls: $\mathbb{1}$(*Kindergarten*), $\mathbb{1}$(*Some prim. sch.*), $\mathbb{1}$(*Some high sch.*), age O(2), $\mathbb{1}$(*fem.*).									
Data: *DHS household*									
Electricity	.0797^a	0.084	.0444^a	0.827	.691	5.96	987,081	28,088	39
	(.0012)		(.0010)		(.462)	(1.68)			
Safe Water	.0853^a	0.083	.0576^a	0.655	.510	5.95	1,005,468	28,604	40
	(.0013)		(.001)		(.500)	(1.69)			
Imp. Sanitation	.0825^a	0.079	.0630^a	0.662	.572	5.95	1,005,283	28,604	40
	(.0010)		(.0010)		(.495)	(1.69)			
Controls: *H.H. size O(2)*, $\mathbb{1}$(*fem. HoH*),*age HoH O(2)*, $\mathbb{1}$(*Some prim. sch. HoH*), $\mathbb{1}$(*Some sec. sch. HoH*), $\mathbb{1}$(*> sec. sch. HoH*).									
Data: *DHS school*									
School≥8yr	.0497^a	0.029	.0158^a	0.719	.611	5.94	95,687	25,529	40
	(.0014)		(.0011)		(.488)	(1.67)			
Controls: $\mathbb{1}$(*fem.*), $\mathbb{1}$(*fem. HoH*), age HoH O(2), $\mathbb{1}$(*Some prim. sch. HoH*), $\mathbb{1}$(*Some sec. sch. HoH*), $\mathbb{1}$(*> sec. sch. HoH*).									
Data: *DHS female*									
Contraception	.0297^a	0.011	.0122^a	0.595	.496	5.9	183,273	19,294	37
	(.0016)		(.0009)		(.500)	(1.76)			
Justified Beating	-.0361^a	0.017	-.0120^a	0.499	.384	5.87	575,495	20,129	40
	(.0016)		(.0009)		(.486)	(1.76)			
Victim	.0001	0.000	.0074^a	0.320	.277	5.8	194,157	17,951	32
	(.0010)		(.0009)		(.448)	(1.77)			
Tot. # births	-.0278^a	0.008	-.0109^a	0.370	.298	6.01	1,110,331	28,604	40
	(.0007)		(.0004)		(.531)	(1.68)			
Controls: *age O(2)*, $\mathbb{1}$(*Some prim. sch.*), $\mathbb{1}$(*Some sec. sch.*), $\mathbb{1}$(*> sec. sch.*), $\mathbb{1}$(*fem. HoH*),*age HoH O(2)*, $\mathbb{1}$(*Some prim. sch. HoH*), $\mathbb{1}$(*Some sec. sch. HoH*), $\mathbb{1}$(*> sec. sch. HoH*).									
Data: *DHS birth*									
Infant Death	-.0006^a	0.000	.0008^a	0.038	.035	5.75	294,385	28,205	40
	(.0002)		(.0002)		(.184)	(1.71)			
Controls: $\mathbb{1}$(*fem.*), *age (mother) O(2)*, $\mathbb{1}$(*Some prim. sch.(mother)*), $\mathbb{1}$(*Some sec. sch.(mother)*), $\mathbb{1}$(*> sec. sch.(mother)*), $\mathbb{1}$(*fem. HoH*), *age HoH O(2)*, $\mathbb{1}$(*Some prim. sch. HoH*), $\mathbb{1}$(*Some sec. sch. HoH*), $\mathbb{1}$(*> sec. sch. HoH*).									

Note: *Regressions of respondent level 'outcome' on log population density in a 5km disk. Standard errors are clustered by 'survey cluster'. Each row reports results from two regressions, one without demographic controls and one with; $^a = 1\%$, $^b = 5\%$, $^c = 10\%$, all two-tailed tests. Relevant demographic controls are listed at the bottom of each panel. $\overline{y}$ and $\overline{x}$ are mean of outcome and* ln(*pop. density*) *in the 'no-controls' sample. Except for the LSMS panel, we lose only a tiny number of observations when we add controls.*

Table A1b: Density gradients for Afrobarometer, LSMS and DHS outcomes.

	No controls		Controls					Clusters	Countries
Outcome	β	R^2	β	R^2	$\overline{y}$	$\overline{x}$	N		
	s.e.		s.e.		s.e.	s.e.			

Data: *DHS children*

Diarrhea	-.0035^a	0.000	.0030^a	0.160	.125	5.76	512,855	28,507	40
	(.0005)		(.0004)		(.331)	(1.71)			
DPT3	.0209^a	0.007	.0123^a	0.798	.763	5.76	95,334	24,914	40
	(.0013)		(.0011)		(.425)	(1.71)			
Cough	-.0001	0.000	.0038^a	0.255	.188	5.76	513,082	28,507	40
	(.0008)		(.0006)		(.391)	(1.71)			

Controls: *age O(2), 𝟙(Some prim. sch.(mother)), 𝟙(Some sec. sch.(mother)), 𝟙(> sec. sch.(mother)),age HoH O(2), 𝟙(Some prim. sch. HoH), 𝟙(Some sec. sch. HoH), 𝟙(> sec. sch. HoH).*

Data: *DHS lifestyle*

High B.P.	.0076^a	0.001	.0108^a	0.260	.244	6.17	475,157	15,838	2
	(.0008)		(.0008)		(.430)	(1.57)			
Asthma	0.00002	0.000	.00012	0.019	.015	6.18	712,978	15,546	1
	(.00012)		(.00012)		(.122)	(1.57)			
Diabetes	.0019^a	0.001	.0015^a	0.028	.014	6.19	677,232	15,545	1
	(.0001)		(.0001)		(.117)	(1.57)			
Obese	.0128^a	0.006	.0100^a	0.154	.077	6.07	851,767	28,330	39
	(.0005)		(.0003)		(.267)	(1.67)			

Controls: *age O(2), 𝟙(Some prim. sch.), 𝟙(Some sec. sch.), 𝟙(> sec. sch.), 𝟙(fem. HoH),age HoH O(2), 𝟙(Some prim. sch. HoH), 𝟙(Some sec. sch. HoH), 𝟙(> sec. sch. HoH).*

Data: *Afrobarometer*

Fear Walking	.0157^a	0.003	.0155^a	0.430	.381	5.65	26,437	2,210	24
	(.0037)		(.0034)		(.486)	(1.76)			
Fear at Home	.0094^a	0.001	.0102^a	0.386	.334	5.65	26,437	2,210	24
	(.0037)		(.0036)		(.472)	(1.76)			
Theft at Home	.0042	0.000	.0059^b	0.320	.288	5.65	26,476	2,210	24
	(.0028)		(.0026)		(.453)	(1.76)			
Attacked	.0026	0.000	.0024	0.147	.103	5.65	26,468	2,210	23
	(.0019)		(.0019)		(.303)	(1.76)			

Controls: *𝟙(< Primary sch.), 𝟙(Some sec. sch.), 𝟙(> high sch.), age O(2), 𝟙(fem.), H.H. size*

Note: *Regressions of respondent level 'outcome' on log population density in a 5km disk. Standard errors are clustered by 'survey cluster'. Each row reports results from two regressions, one without demographic controls and one with; $^a = 1\%$, $^b = 5\%$, $^c = 10\%$, all two-tailed tests. Relevant demographic controls are listed at the bottom of each panel. $\overline{y}$ and $\overline{x}$ are mean of outcome and $\ln$(pop. density) in the 'no-controls' sample. Except for the LSMS panel, we lose only a tiny number of observations when we add controls.*

Table A2: Country lists for Afrobarometer, LSMS and DHS outcomes.

Data: *LSMS*
Ethiopia, Ghana, Malawi, Nigeria, Tanzania, Uganda.
Data: *Afrobarometer*
Algeria, Angola, Benin, Eswatini, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritius, Morocco, Mozambique, Namibia, Niger, Nigeria, Sao Tome and Principe, Senegal, Sierra Leone, South Africa, Sudan, Togo, Tunisia, Uganda, United Republic of Tanzania, Zambia, Zimbabwe.
Data: *DHS*
Angola 2015-16, Bangladesh 2014, Benin 2011-12, Burkina Faso 2010, Burundi 2010, Cambodia 2014, Cameroon 2011, Chad 2014-15, Colombia 2010, Comoros 2012, Congo Democratic Republic 2013-14, Cote d'Ivoire 2011-12, Dominican Republic 2013, Ethiopia 2016, Gabon 2012, Ghana 2014, Guatemala 2014-15, Guinea 2012, Haiti 2012, Honduras 2011-12, India 2015-16, Kenya 2014, Lesotho 2014, Liberia 2013, Malawi 2015-16, Mali 2012-13, Mozambique 2011, Myanmar 2015-16, Namibia 2013, Nepal 2016, Nigeria 2013, Rwanda 2014-15, Senegal 2010-11, Sierra Leone 2013, Tanzania 2015-16, Timor-Leste 2016, Togo 2013-14, Uganda 2016, Zambia 2013-14, Zimbabwe 2015.