NOT FOR PUBLICATION

Appendix C: Supplementary materials for "Exclusionary policies in urban development: under-servicing migrant households in Brazilian cities," by Leo Feler and J. Vernon Henderson

This appendix provides supporting estimation results for all the robustness checks discussed but not presented in the main text. Here we provide tables and further discussion.

Robustness checks for Table 3

Table C1 presents robustness checks for Table 3, column 3 in the text, where we obtain using 2SLS a coefficient estimate (standard error) of 0.729 (0.419) for the effect that the share of small houses with water in 1991 has on the growth rate of urban households between 1991 and 2000. Because it is hard to reject weak instruments overall, we present in columns 1 and 2 of Table C1, results using Limited Information Maximum Likelihood (LIML) in place of 2SLS. LIML estimators are more robust to weak instruments but are also more sensitive to the length of the instrument list. Using the same instrument list as we used in Table 3, we show in Table C1, column 1, how the coefficient estimate increases to 1.251 (with standard error 0.660) when estimating with LIML. When using a shorter instrument list and estimating with LIML, in column 2, the coefficient estimate (standard error) is 0.853 (0.507), approximately the same as what we obtain with 2SLS in Table 3. With the short instrument list, the Kleibergen-Paap LM p-value for the underidentification test is 0.04, suggesting we can reject underidentification, although the precision of the coefficient estimate for the service variables is not improved. The shortened instrument list is adult illiteracy in 1970, share of the locality population that voted against the military in the 1982 legislative elections, the manufacturing-to-service employment ratio in the rest of the metro area in 1970, this ratio interacted with the log of the distance of the metro area to Sao Paulo, the log of the number of urban households in the rest of the metro area in 1970, the share of the locality's geology that is porous, this share porous interacted with mean insolation, and this share porous interacted with the standard deviation of insolation.

In Table C1, columns 3 and 4, we respectively drop 44 localities that are composed of multiple municipalities, but none are dominant with more than 85% of the locality population, and we drop all localities in the Rio de Janeiro and Sao Paulo metro areas. We obtain coefficient estimates (standard errors) of 0.654 (0.377) and 0.601 (0.400), respectively. These coefficients are more noisily estimated because of smaller sample sizes, but they do not differ greatly from what we obtain in Table 3.

In Table C1, columns 5 and 6, we experiment with a placebo exercise, looking at whether the share of small houses with a radio or TV set in 1991 has any effect on the subsequent growth rate of urban households. The ability of small households to have a radio or TV set, controlling for fixed effects for differences in metro area incomes, should not lead to higher growth in the number of households since radio and TV broadcasts are neither excludable (unlike public water connections) nor are they necessarily highly-valued by the poor. We find that although the OLS estimate is positive and significant as might be expected, at 0.250 (with standard error 0.093), the 2SLS estimate is zero and insignificant, at -0.172 (with standard error 0.317). Instruments for the share of small houses with a radio or TV set are weak, however, with a first-stage partial *F*-statistic of 7.10. Nevertheless, this placebo exercise is suggestive that the results we obtain for the water servicing variable are not just due to income differences.

In Table C1, columns 7 and 8, rather than using the share of small houses serviced, we use the log total count of small houses serviced as the policy variable, with its attendant indexing and scaling issues. We obtain a positive but insignificant OLS coefficient (standard error) of 0.044 (0.037), which rises to 0.149 (0.099) under 2SLS, suggesting that a 1% increase in the number of serviced small houses leads to a 0.15% increase in the growth of the locality. This indicates the same direction of effects but with less precision and weaker first-stage results as what we obtain in Table 3.

Finally, we experiment with the dependent variable, looking at rural and low-education in-migrants. This captures either the number of household heads coming from rural areas outside the locality or the number of household heads who are low-education (who have not completed primary school) coming from outside the locality in the previous five years. This captures gross and not net flows because it does not account for out-migration. In columns 9 and 10 of Table C1, we obtain an OLS estimate (standard error) of -0.451 (0.391), which becomes large and positive under 2SLS, at 2.358 (1.456), for the effect of the servicing variable on the number of rural in-migrants. Similarly, in columns 11 and 12, we obtain an OLS estimate (standard error) of -1.019 (0.345), which becomes 1.918 (1.454) under 2SLS. While servicing of small houses leads to higher growth in the number of households in a locality, it appears to lead to particularly strong in-migration of rural and low-education households (although these are noisily estimated).

Robustness checks for Table 4

In Table C2, we perform robustness checks for Table 4 in the text, which looks at the effect of servicing on the growth of low-education and high-education households. In Table C2, we estimate a ratio model looking at the growth of low-education relative to high-education households. The 2SLS estimate (standard error) is negative, at -0.389 (0.346), although not significantly different from zero, which hints that poor servicing of small houses has adverse effects on the growth of all education

households, suggesting a negative externality for high-education households. We note, however, that we have difficulty instrumenting for the ratio of low-education to high-education households in 1991, which is the reason the Hansen-Sargan test leads to a rejection of the null that the instruments are valid.

Robustness checks for Table 5

In Table C3, we performs two sets of robustness checks, separately dropping the 44 localities that are composed of multiple municipalities, but none are dominant with more than 85% of the locality population, in columns 1 and 2, and dropping all localities in the Rio de Janeiro and Sao Paulo metro areas, in columns 3 and 4. Our estimates for the effects of income, size, and the income-size interaction on the share of small houses serviced remains in line with what we obtained in Table 5, although we lose some significance because of smaller sample sizes. Our results in Table 5 for the determinants of servicing do not appear to be driven by issues relating to how those localities with multiple municipalities set servicing or how servicing is determined in Rio de Janeiro and Sao Paulo.

Robustness checks for Tables 6 and 7

Table 7 in the text looks at the counterfactual of servicing large households in 1991 and of servicing small households in 2000, after more than a decade of democracy and mandatory voting. As robustness checks, we replicate the quintile specification in Table 6 for these counterfactuals. Table C4 shows the quintile specification for the share of small houses serviced in 2000. Paralleling the result in Table 7, where the income-size interaction is insignificant, here we see that the income-size interactions for the two largest quintiles are either insignificant or positive, suggesting increased servicing of small houses in 2000 by the largest and wealthiest localities.

Table C5 presents the results from the quintile specification for the share of large houses serviced in 1991. Again, none of the income-size interactions are significant for the largest quintiles, paralleling the result in Table 7. The largest and wealthiest localities do not appear to be withholding service from large houses.

Robustness checks for Table 8

Our last set of robustness checks is for Table 8 in the text. In Table C6, columns 1 and 2, we show results for the incorrect, biased estimation of strategic interactions using OLS, following equation (13) in the text. In column 1, controlling for metro area characteristics but without metro area fixed effects, we obtain a positive coefficient estimate (standard error) of 0.269 (0.054) for the strategic interaction term, which becomes negative at -0.457 (0.159) when we estimate with OLS and metro area fixed effects. This is illustrative of the negative bias that results when including metro area fixed effects and estimating with OLS. In column 3, we return to the proper estimation with MLE and metro area

fixed effects, but look only at metro areas with 2 to 3 localities. This reduces our sample to 23 metro areas with 59 localities. We obtain a coefficient estimate (standard error) for the strategic interaction term of -0.753 (0.075), which is larger than what we obtain for the full sample of metro areas and localities (coefficient estimate of -0.590 and standard error of 0.081). We might expect this type of result: the strongest non-competitive behavior is likely to arise in cases when there are few localities in a metro area, making cooperation or collusion between localities more possible.

Table C1: Robustness checks for Table 3

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Growth rate of urban households, 1991-2000		Growth rate of urban households, 1991-2000		Growth rate of urban households, 1991-2000		Growth rate of urban households, 1991-2000		ln # rural migrant HHs, 2000		ln # low-educ migrant HHs, 2000	
	LIML	LIML, short inst. list	2SLS, drop 44 localities	2SLS, drop Rio & SP	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Share small houses with water 1991	1.251*	0.853*	0.654*	0.601					-0.451	2.358*	-1.019***	1.918
	(0.660)	(0.507)	(0.377)	(0.400)					(0.391)	(1.436)	(0.345)	(1.454)
Share small houses with radio or TV set 1991					0.250**	-0.172						
					(0.093)	(0.317)						
ln # small houses with water 1991							0.044	0.149				
							(0.037)	(0.099)				
ln # urban HHs 1991	-0.154***	-0.120***	-0.094***	-0.104***	-0.042***	-0.057***	-0.083**	-0.223**	0.848***	0.573***	0.829***	0.509***
	(0.050)	(0.039)	(0.027)	(0.027)	(0.008)	(0.015)	(0.041)	(0.108)	(0.051)	(0.091)	(0.033)	(0.112)
ln land area 1991	0.086**	0.065***	0.045**	0.052***	0.018*	0.023	0.017**	0.039***	-0.037	0.128**	-0.033	0.157*
	(0.034)	(0.024)	(0.020)	(0.013)	(0.009)	(0.015)	(0.007)	(0.013)	(0.047)	(0.055)	(0.034)	(0.081)
Metro area fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N [metro areas]	327 [54]	327 [54]	281 [51]	279 [52]	327 [54]	327 [54]	327 [54]	327 [54]	297 [54]	297 [54]	327 [54]	327 [54]
R-squared within metro areas					0.10		0.09		0.77		0.79	
Hansen-Sargan stat. p-value	0.99	0.97	0.84	0.97		0.54		0.38		0.84		0.66
Underidentification test p-value	0.17	0.04	0.18	0.08		0.15		0.27		0.23		0.17
Min. 1st stage partial F	13.82	14.19	22.41	25.01		7.10		30.29		10.55		13.82

Notes: * significant at 10%; ** significant at 5%; *** significant at 1% level.

Robust standard errors, clustered at the metro area level, reported in parentheses.

For columns 1, 3, 4, 6, 8, 10, and 12, excluded instruments are: locality adult illiteracy rate in 1970, adult illiteracy rate in the rest of the metro area in 1970, the share of locality votes for anti-military parties in the 1982 national legislative elections, the manufacturing-to-service employment ratio in the rest of the metro area in 1970, this manufacturing-to-service ratio interacted with the log of the distance of the locality to Sao Paulo, the adult illiteracy rate in the rest of the metro area in 1970 interacted with the log of the distance to Sao Paulo, the log number of households in the rest of the metro area in 1970, the share of households that are rural in the rest of the metro area in 1970, the share of a locality's land that is composed of porous geology, mean annual insolation in the locality, the standard deviation of average monthly insolation in the locality, the share of a locality's geology that is porous interacted with mean insolation, and the share porous geology interacted with the standard deviation of insolation.

For column 2, excluded instruments are: adult illiteracy in 1970, share of the locality population that voted against the military in the 1982 legislative elections, the manufacturing-to-service employment ratio in the rest of the metro area in 1970, this ratio interacted with the log of the distance of the metro area to Sao Paulo, the log of the number of urban households in the rest of the metro area in 1970, the share of the locality's geology that is porous, this share porous interacted with mean insolation, and this share porous interacted with the standard deviation of insolation.

For column 3, the sample excludes 44 localities composed of multiple municipalities, none of which has at least 85% of the locality population.

For column 4, the sample excludes all localities in the Rio de Janeiro and Sao Paulo metro areas.

Table C2: Robustness checks for Table 4

Dependent variable: Growth rate of low-education relative to high-education households, 1991-2000 [ln (# low educ HHs)(t) - ln (# low educ HHs)(t-1)]-[ln (# high educ HHs)(t) - ln (# high educ HHs)(t-1)]	(1) OLS	(2) 2SLS
Share small houses with water 1991	0.081	-0.389
	(0.075)	(0.346)
ln density 1991	-0.035***	0.004
	(0.013)	(0.041)
ln (# low educ. HHs/# high educ. HHs) 1991	-0.350***	-0.337***
	(0.042)	(0.120)
Metro area fixed effects	Yes	Yes
N [54 metro areas]	327	327
R-squared within metro areas	0.36	
Hansen-Sargan stat. p-value		0.03
Underidentification test p-value		0.09
Min. 1st stage partial F		13.99

Notes: * significant at 10%; ** significant at 5%; *** significant at 1% level.
Robust standard errors, clustered at the metro area level, reported in parentheses.
Excluded instruments are: locality adult illiteracy rate in 1970, adult illiteracy rate in the rest of the metro area in 1970, the share of locality votes for anti-military parties in the 1982 national legislative elections, the manufacturing-to-service employment ratio in the rest of the metro area in 1970, this manufacturing-to-service ratio interacted with the log of the distance of the locality to Sao Paulo, the adult illiteracy rate in the rest of the metro area in 1970 interacted with the log of the distance to Sao Paulo, the log number of households in the rest of the metro area in 1970, the share of households that are rural in the rest of the metro area in 1970, the share of a locality's land that is composed of porous geology, mean annual insolation in the locality, the standard deviation of average monthly insolation in the locality, the share of a locality's geology that is porous interacted with mean insolation, and the share porous geology interacted with the standard deviation of insolation.

Table C3: Robustness checks for Table 5

Dependent variable: Share small houses with water connection in own locality, 1991	(1)	(2)	(3)	(4)
	Drop 44 localities		Drop Rio & SP	
ln median HH income 1980	0.439***	0.327***	0.456***	0.422***
	(0.133)	(0.113)	(0.163)	(0.143)
ln # urban HHs 1980	0.281**	0.222*	0.250	0.237*
	(0.135)	(0.114)	(0.155)	(0.131)
ln median HH income * ln # urban HHs 1980	-0.027**	-0.024**	-0.026	-0.025*
	(0.013)	(0.011)	(0.016)	(0.013)
Share anti-military vote 1982 national elections		0.134		0.118
		(0.088)		(0.072)
ln density 1980		0.042***		0.007
		(0.015)		(0.009)
Share all urban HHs with water 1970		0.059		0.072*
		(0.042)		(0.036)
Metro area fixed effects	Yes	Yes	Yes	Yes
N [metro areas]	246 [50]	246 [50]	230 [48]	230 [48]
R-squared within metro areas	0.24	0.33	0.23	0.26

Notes: * significant at 10%; ** significant at 5%; *** significant at 1% level.
Robust standard errors, clustered at the metro area level, reported in parentheses.

Table C4: Robustness checks for Tables 6 and 7, quintile specification for small houses serviced in 2000

Dependent variable: Share small houses with water connection in own locality in 2000				
	Quintile 2	Quintile 3	Quintile 4	Quintile 5
ln median HH income 1980	0.071	0.073	-0.016	0.037
	(0.043)	(0.051)	(0.078)	(0.066)
ln # urban HHs 1980	0.037	0.047	0.043	-0.091
	(0.061)	(0.044)	(0.074)	(0.056)
		Relevant Interactions		
4th income * 4th size quintile		-0.014		
		(0.073)		
4th income * 5th size quintile		0.157**		
		(0.078)		
5th income * 4th size quintile		-0.094		
		(0.072)		
5th income * 5th size quintile		0.052		
		(0.054)		
Metro area fixed effects		Yes		
Locality controls from Table 5, column 2		Yes		
N [50 metro areas]		276		
R-squared within metro areas		0.30		

Notes: * significant at 10%; ** significant at 5%; *** significant at 1% level.
Robust standard errors, clustered at the metro area level, reported in parentheses.
Locality controls are: share of votes for the anti-military party in the 1982 national elections, natural log of population density in 1980, and share of all urban households with connection to water in 1970.

Table C5: Robustness checks for Tables 6 and 7, quintile specification for large houses serviced in 1991

Dependent variable: Share large houses with water connection in own locality in 1991				
	Quintile 2	Quintile 3	Quintile 4	Quintile 5
ln median HH income 1980	0.016	0.009	0.007	0.026
	(0.037)	(0.054)	(0.055)	(0.040)
ln # urban HHs 1980	-0.029	0.034	-0.010	0.002
	(0.046)	(0.041)	(0.051)	(0.055)
		Relevant Interactions		
4^{th} income * 4^{th} size quintile		0.058		
		(0.057)		
4^{th} income * 5^{th} size quintile		0.056		
		(0.064)		
5^{th} income * 4^{th} size quintile		-0.006		
		(0.047)		
5^{th} income * 5^{th} size quintile		0.024		
		(0.053)		
Metro area fixed effects		Yes		
Locality controls from Table 5, column 2		Yes		
N [50 metro areas]		276		
R-squared within metro areas		0.32		

Notes: * significant at 10%; ** significant at 5%; *** significant at 1% level.
Robust standard errors, clustered at the metro area level, reported in parentheses.
Locality controls are: share of votes for the anti-military party in the 1982 national elections, natural log of population density in 1980, and share of all urban households with connection to water in 1970.

Table C6: Robustness checks for Table 8

Dependent variable: Share small houses with water connection in own locality, 1991	(1) OLS Inverse Density Weights	(2) OLS Inverse Density Weights	(3) MLE Inverse Density Weights, 2-3 localities only
Weighted avg. other localities share of servicing of small houses	0.269***	-0.457***	-0.753***
	(0.054)	(0.159)	(0.075)
ln median HH income 1980	0.396***	0.299***	0.414***
	(0.145)	(0.104)	(0.141)
ln # urban HHs 1980	0.302**	0.190*	0.352**
	(0.147)	(0.099)	(0.151)
ln median HH income * ln # urban HHs 1980	-0.032**	-0.020**	-0.038**
	(0.015)	(0.010)	(0.016)
Spatial correlation of error terms			-0.795***
			(0.077)
Locality controls	Yes	Yes	Yes
Metro area controls	Yes	N/A	N/A
Metro area fixed effects	No	Yes	Yes
N [metro areas]	276 [50]	276 [50]	59 [23]
R-squared [within metro areas]	0.53	[0.36]	0.99

Notes: * significant at 10%; ** significant at 5%; *** significant at 1% level.
Standard errors reported in parentheses. For columns 1 and 2, errors are clustered at the metro area level. For column 3, errors are corrected for spatial correlation within metro areas.
Weights refer to the weighting of other localities' share of servicing of small houses in 1991, for localities within the same metro area. Locality controls are: share of votes for the anti-military party in the 1982 national elections, natural log of population density in 1980, and share of all urban households with connection to water in 1970. Metro area controls are: natural log of the distance from the metro area to Sao Paulo, natural log of the metro area land area, the illiteracy rate in the metro area, and the manufacturing-to-service employment ratio in the metro area, all for 1980.