Appendix to *Relocating or Redefined: A New Perspective on Urbanization in China*

Description of Tracking Algorithm

In each year, there are nearly 700,000 community-level administrative units in China. For every two consecutive years, around 5% to 10% of these communities would have their community ID codes change for reasons discussed in Section 2.3. Also in Section 2.3, we briefly introduce the algorithm we use to track communities. In this appendix, we would in details talk about each step and the corresponding results about tracking communities for every two consecutive years from 2009 to 2017.

As shown in Figure 2, for each community, we know its community ID and Chinese name. From the code, we can extract information regarding all upper-level administrative units (e.g., province, prefecture, county and town-level governments) it belongs to, beyond which we have no further information. Since, we cannot obtain geographic locations (i.e., longitude and latitude) for each community, it would be impossible for us to track communities due to merges and splits using our algorithm. For example, Communities A and B merged as a new community. If the merged community in the latter year shares the same community ID and name with one of these two communities, (e.g., A), then our tracking methods would track Community A in both years and leave B in the former year unmatched. If the merged community in the latter year has a new code and name, then they would be left unmatched in both years.[1] Apart from that, our tracking algorithm works pretty well in one-to-one community tracking in any two consecutive years. Put differently, almost all communities that are left unmatched are due to merges and splits.

---

[1]Since we could not distinguish both cases, we are not aware of the share of each case. Nevertheless, we can take the county-level administrative units merge as a reference. If the new county shares the same name with one of the previous counties, then it will keep the corresponding ID (For example, in 2015, *Jing'an* District and *Zhabei* District in *Shanghai* merged as a new district named *Jing'an*, then the new district kept the original *Jing'an* District ID code). Otherwise, it would have a new code (for example, in 2016, *Qinghe* District and *Qingpu* District in *Huai'an, Jiangsu* merged as a new district named *Qingjiangpu*, then the new district was assigned a new county-level ID code).

Another fact that is worth mentioning is the variable we choose as the identifier. First, community name itself cannot work since there are multiple villages sharing the same Chinese name. Though each name is unique within a town, there might be several villages with the same name across towns in the same county. Second, community ID itself is not ideal as a single identifier, at least in the first few steps. There is a possibility that Community *A* with an ID in the former year is reassigned a new ID in the latter year while this new ID belonged to Community *B* in the same town in the former year. For example, in 2009, under *Pingzhuangcheng* Street, *Yuanbaoshan* District in *Chifeng* City, *Inner Mongolia*, the ID *150403002005* was assigned to *Qianjin Road No.2 Community*. In 2010, this ID was assigned to *Fangzhi Road No.1 Community* while that community used to have another ID *150403002014* in 2009. Thus, in this case, if we use only the community ID as the key variable, we could mistakenly track two different communities. As a result, using only the twelve-digit community ID code as the key identifier would cause a mismatch problem, even within the same town-level units, though this case is very rare in our dataset. In consequence, in order to prevent such mismatch, we use both community ID and its name (in Chinese) as key variables in our tracking algorithm.

Specifically, steps of our tracking algorithm are as follows and summarized in Table A1. After each step, tracked communities are dropped and we would deal only with those left unmatched in the next step, which would avoid the double counting problem.

Step 0: In this step, we track communities without code or name change using both variables as the identifier, which account for the majority of all tracked communities.

Starting from Step 1, we begin dealing with communities whose upper-level administrative units have been adjusted. As has been mentioned, since digits in the code represent a corresponding level of administrative units, we will proceed from the highest level (province) to the lowest level (town). For administrative changes at the provincial, city and county-level governments, we search for formal documentation on adjustment details from the website

of the Ministry of Civil Affairs.[2] First, we manually check changes at the province level. It turns out to be that in any two consecutive years between 2009 and 2017, there are no administrative unit adjustments across provinces. Put differently, no administrative units are reassigned to a different province. Thus, we start from the prefecture level.

Step 1: In this step, we track communities with prefecture-level administrative units change, which corresponds to the third and fourth digits of the community ID. There are three reasons for prefecture-level units adjustment: establish of new prefectures, abolishment of current prefectures and changes of type of units.[3] For example, in 2011, *Chaohu* City in *Anhui* Province was revoked and its counties were assigned to three cities in the province. Examples of establishment of new prefectures include *Sansha* and *Danzhou* City in *Hainan* province. Though, the majority of administrative units change at the prefecture level takes the form of readjustment, with many prefecture-level regions in Southwestern China areas or autonomous regions reclassified as cities during this period. For all these three cases, we check communities under those changed prefectures and manually do the one-to-one matching.

Step 2: In this step, we track communities with county-level administrative units change. Such changes are shown from the fifth and sixth digits of the code. Similarly, most adjustments come from changes of type of county-level administrative units.[4] In particular, many counties or county-level cities are reclassified as districts. In addition to such reclassification, county-level units merges and splits are also a key component for adjustments. We also check communities under those changed counties and manually do the one-to-one matching.

The next two steps are associated with town-level ID code change. Since there are no detailed formal documents on adjustments at the town level and lower, we mainly rely on adjustments due to type changes of town-level administrative units.

---

[2]The website is http://xzqh.mca.gov.cn/description?dcpid=1

[3]Types of prefecture-level administrative units include prefecture-level cities (*Shi* in Chinese), prefecture-level regions (*Diqu*), autonomous prefectures (*Zizhizhou*) and leagues (*Meng*).

[4]Types of county-level administrative units include districts (*Qu* in Chinese), counties (*Xian*), county-level cities (*Xianjishi*) and autonomous counties (*Zizhixian*), banners (*Qi*) and autonomous banners (*Zizhiqi*). The latter two are particular for minority-grouped areas in *Inner Mongolia*.

Step 3: We first separate the characters indicating the type of the town-level units[5] from its full name and define what is left as its "short name".[6] For town-level units that cannot be matched using their full names within a county in two consecutive years, We match them using their short names. Those matched samples may represent type adjustments.[7] Then we are able to track communities under those towns using both the new code by changing the corresponding digits (seventh to ninth) in the community ID and community name as key variables.

Step 4: We follow similar procedures in Step 3, only allowing matching town-level units with short name as identifier across county within the same prefecture in two consecutive years.

Beginning from Step 5, we track communities with code changes at the community level.

Step 5: Since the community name is uniquely identifiable within a town-level unit. In this step, we track communities by using town-level ID code and community name as key variables, which accounts for re-orderings for community lists within a town-level administrative unit.

Step 6: In this step, we use similar methods as we did for type adjustments at the town level. For all unmatched communities, we construct their short names by dropping the characters indicating their types.[8] We then match communities within town using their short names as key variables.

Step 7: This step follows Step 5, but using county-level ID code and community name as key variables, allowing communities reassignment to another town-level unit within the same county.

---

[5]Types of town-level administrative units include towns (*Zhen* in Chinese), townships (*Xiang*), Streets (*Jiedao*) and autonomous townships (*Zizhixiang*) and types indicating a particular use (e.g., prisons, farms, free trade zones, industrial zones).

[6]For example, if the full name of a town-level administrative unit is *Guanghua Township*, we just keep *Guanghua* as the short name.

[7]For example, if the *Guanghua Township* is reclassified as *Guanghua Street* in the next year, though we cannot match these two units with their full name, we may achieve that result using the short-name matching.

[8]Types of community-level administrative units include communities (*Shequ* in Chinese), villages (*Cun*) and types indicating a particular use.

Step 8: Similar to Step 7, this step follows Step 6, using county-level ID code and community short name as key variables.

Step 9: In this step, we keep the first two characters of a community's full name and use this as the key identifier to track communities within town in two consecutive years. Though not as sharp as Step 0-8, in most cases they are accurate.

Step 10: The final step uses community ID code only as the identifier for those unmatched communities.[9] The reason that we would adopt this final step is because we observe that after previous steps, most unmatched communities result from character changes in the name of the community. Especially for communities in the minority-grouped regions, there exist changes in translating community names from their own language to Mandarin Chinese.

---

[9]Before the last step, we would match communities with the same name within a town across year, so the aforementioned mismatch problem would be alleviated.

Tables to Appendix

Table A1: Description of Tracking Steps

Description of Algorithm	Step	Key Variables Used in Matching
Tracking with unchanged code and name	0	Community Code, Name
Tracking with changed code at the prefecture level	1	Manually
Tracking with changed code at the county level	2	Manually
Tracking with changed code at the town level		
with: administrative type adjusted within county	3	Community Code[a], Name
with: reassignment across counties within prefecture	4	Community Code[a], Name
Tracking with changed code at the community level		
with: no name change within town	5	Town Code, Name
with: no short name change within town	6	Town Code, Short Name[b]
with: no name change across town within county	7	County Code, Name
with: no short name change across town within county	8	County Code, Short Name
with: first 2 characters of community name within town	9	Town Code, First 2 Characters
with: community code only	10	Community Code

[a] This is revised community code with substituting the 7th to 9th digits representing the town-level administrative unit with the corresponding digits of the matched town-level units after matching with short names as identifiers.
[b] "Short Name" refers to community name dropping the characters indicating the type of the community. For example, short name for *Guanghua Village* is *Guanghua*.

Table A2: Tracking Results for Communities between 2009 and 2010

Step	Communities	Reason for Adjustment	Related Code (2009)	Related Code (2010)
0	666,778			
1	-			
2	1,051			
	90	merge	*110103*	*110101*
	107	merge	*110104*	*110102*
	336	merge & establishment	*120107, 120108, 120109*	*120116*
	52	abolishment	*320304*	*320302, 320311, 320312*
	326	reclassification	*320323*	*320312*
	39	establishment	*360425, 360426, 360427*	*360482*
	101	reclassification	*532522*	*532503*
3	9,275			
4	453			
5	971			
6	4,820			
7	4,666			
8	556			
9	1,900			
10	2,455			
Total	692,925			

Table A3: Tracking Results for Communities between 2010 and 2011

Step	Communities	Reason for Adjustment	Related Code (2010)	Related Code (2011)
0	648,805			
1	6,463			
	1,018	abolishment	*3414*	*3401, 3402, 3405*
	1,718	reclassification	*5222*	*5206*
	3,727	reclassification	*5224*	*5205*
2	1,929			
	72	merge	*310103*	*310101*
	65	reassignment	*321003*	*321002*
	113	merge	*321011*	*321003*
	333	reclassification	*321088*	*321012*
	156	reclassification	*430122*	*430112*
	439	merge & reclassification	*500110, 500222*	*500110*
	312	merge & reclassification	*500111, 500225*	*500111*
	237	reclassification	*511522*	*511503*
	65	reclassification	*530121*	*530114*
	137	reclassification	*532621*	*532601*
3	8,986			
4	580			
5	1,318			
6	3,291			
7	12,209			
8	644			
9	1,646			
10	2,868			
Total	688,739			

Table A4: Tracking Results for Communities between 2011 and 2012

Step	Communities	Reason for Adjustment	Related Code (2011)	Related Code (2012)
0	675,115			
1	3			
	3	establishment	*4690*	*4603*
2	981			
	131	reclassification	*130207, 130230*	*130209*
	269	merge & establishment	*320502, 320503, 320504*	*320508*
	310	reclassification	*320584*	*320509*
	26	merge	*340502*	*340503*
	43	establishment	*340521*	*340506*
	202	reclassification	*511821*	*511803*
3	4,327			
4	199			
5	1,029			
6	2,026			
7	4,640			
8	861			
9	929			
10	1,412			
Total	691,522			

Table A5: Tracking Results for Communities between 2012 and 2013

Step	Communities	Reason for Adjustment	Related Code (2012)	Related Code (2013)
0	661,722			
1	1,648			
	1,648	reclassification	*6321*	*6302*
2	5,619			
	21	establishment	*150781*	*150703*
	407	reclassification	*220724*	*220781*
	58	merge	*320103*	*320104*
	53	merge	*320107*	*320106*
	120	reclassification	*320124*	*320117*
	148	reclassification	*320125*	*320118*
	319	reclassification	*321284*	*321204*
	93	merge	*370205*	*370203*
	1,023	merge	*370284*	*370211*
	206	reclassification	*441827*	*441803*
	526	reclassification	*445121*	*445102, 445103*
	265	establishment	*445202, 445221*	*445202, 445203*
	168	reclassification	*450322*	*450312*
	57	merge	*450404*	*450403*
	12	reassignment	*450403*	*450421*
	79	establishment	*450421*	*450406*
	116	establishment	*450902*	*450903*
	282	establishment	*511602*	*511603*
	849	reclassification	*511721*	*511703*
	435	establishment	*511902*	*511903*
	59	merge	*520114, 520181*	*520111*
	75	establishment	*520112*	*520115*
	140	reclassification	*532526*	*532504*
	31	establishment	*542430*	*542431*
	75	reclassification	*632721*	*632701*
	2	establishment	*652701*	*652702*
3	6,257			
4	246			
5	733			
6	3,621			
7	6,161			
8	275			
9	1,864			
10	2,795			
Total	690,941			

Table A6: Tracking Results for Communities between 2013 and 2014

Step	Communities	Reason for Adjustment	Related Code (2013)	Related Code (2014)
0	639,153			
1	2,514			
	492	reassignment	*1306*	*1390*
	354	reassignment	*1301*	*1390*
	1,668	reclassification	*5423*	*5402*
2	8,261			
	69	abolishment	*130103*	*130102, 130104*
	173	reclassification	*130124*	*130111*
	224	reclassification	*130182*	*130109*
	215	reclassification	*130185*	*130110*
	277	reclassification	*220181*	*220113*
	196	merge	*320705, 320706*	*320706*
	468	reclassification	*320721*	*320707*
	401	reclassification	*330621*	*330602, 330603*
	439	reclassification	*330682*	*330604*
	330	reclassification	*360782*	*360702, 360703*
	281	merge	*370802*	*370811*
	437	reclassification	*370882*	*370812*
	733	reclassification	*371081*	*371002, 371003*
	990	reclassification	*371421*	*371403*
	443	reclassification	*371624*	*371603*
	375	reclassification	*420321*	*420304*
	337	reclassification	*440183*	*440118*
	265	reclassification	*440184*	*440117*
	426	merge & reclassification	*440903, 440923*	*440904*
	388	reclassification	*441421*	*441403*
	142	reclassification	*445302, 445323*	*445302, 445303*
	140	establishment	*460201*	*460202 - 460205*
	326	reclassification	*500224*	*500151*
	170	reclassification	*500227*	*500120*
	16	establishment	*632802*	*632857, 632858, 632859*
3	6,964			
4	487			
5	525			
6	1,875			
7	4,486			
8	343			
9	839			
10	1,153			
Total	666,600			

Table A7: Tracking Results for Communities between 2014 and 2015

Step	Communities	Reason for Adjustment	Related Code (2014)	Related Code (2015)
0	638,107			
1	1,860			
	1,142	reclassification	*5421*	*5403*
	496	reclassification	*5426*	*5404*
	222	reclassification	*6521*	*6504*
2	7,425			
	311	reclassification	*120221*	*120117*
	418	reclassification	*120223*	*120118*
	573	reclassification	*130323*	*130302, 130304, 130306*
	221	merge & establishment	*130603, 130604*	*130606*
	193	reclassification	*130621*	*130607*
	272	reclassification	*130622*	*130608*
	304	reclassification	*130625*	*130609*
	330	reclassification	*230182*	*230113*
	80	merge	*320405, 320412*	*320402, 320404, 320411, 320412*
	109	reclassification	*320482*	*320413*
	245	reclassification	*320982*	*320904*
	305	reclassification	*330183*	*330111*
	103	reclassification	*330303, 330322*	*330305*
	217	reclassification	*350784*	*350703*
	279	reclassification	*350822*	*350803*
	338	reclassification	*360122*	*360112*
	223	reclassification	*361122*	*361103*
	345	reclassification	*410224*	*410212*
	60	merge	*440116*	*440112*
	352	reclassification	*441283*	*441204*
	171	reclassification	*441723*	*441704*
	218	reclassification	*450122*	*450110*
	291	reclassification	*451025*	*451081*
	303	reclassification	*500223*	*500152*
	168	reclassification	*500226*	*500153*
	108	reclassification	*511422*	*511403*
	244	reclassification	*513321*	*513301*
	146	reclassification	*520421*	*520403*
	220	reclassification	*530522*	*530581*
	64	reclassification	*533421*	*533401*
	96	reclassification	*610126*	*610117*
	118	reclassification	*630221*	*630203*
3	7,467			
4	70			
5	327			
6	2,140			
7	5,082			
8	256			
9	395			
10	1,123			
Total	664,252			

Table A8: Tracking Results for Communities between 2015 and 2016

Step	Communities	Reason for Adjustment	Related Code (2015)	Related Code (2016)
0	604,947			
1	1,083			
	285	establishment	*4690*	*4603*
	554	reclassification	*5422*	*5405*
	244	reclassification	*6522*	*6505*
2	9,901			
	426	reclassification	*110228*	*110118*
	422	reclassification	*110229*	*110119*
	979	reclassification	*120225*	*120119*
	300	merge	*130721*	*130702, 130703, 130705*
	182	reclassification	*130729*	*130708*
	214	reclassification	*130733*	*130709*
	382	reclassification	*131181*	*131103*
	232	reclassification	*210122*	*210115*
	254	reclassification	*210282*	*210214*
	167	reclassification	*211121*	*211104*
	55	reclassification	*230833*	*230883*
	122	reclassification	*231024*	*231086*
	205	merge	*310108*	*310106*
	340	reclassification	*310230*	*310151*
	154	merge & establishment	*320202, 320203, 320204*	*320213*
	118	merge & establishment	*320802, 320811*	*320812*
	122	reclassification	*320829*	*320813*
	40	merge & establishment	*340702, 340703*	*340705*
	115	reclassification	*340721*	*340706*
	257	reassignment	*340823*	*340722*
	278	reassignment	*341521*	*340422*
	82	reclassification	*360402, 360427*	*360483*
	330	reclassification	*370521*	*370505*
	367	reclassification	*371727*	*371703*
	272	reclassification	*411222*	*411203*
	147	reclassification	*450221*	*450206*
	522	reclassification	*500234*	*500154*
	282	reclassification	*510122*	*510116*
	261	reclassification	*510724*	*510705*
	853	reassignment	*512081*	*510185*
	108	reclassification	*513229*	*513201*
	267	reclassification	*520303, 520321*	*520302, 520303, 520304*
	130	reclassification	*530328*	*530303*
	72	reclassification	*530421*	*530403*
	76	reclassification	*533321*	*533301*
	34	reclassification	*540125*	*540103*
	150	reclassification	*610521*	*610503*
	217	reclassification	*610624*	*610603*
	367	reclassification	*610823*	*610803*
3	6,066			
4	32			
5	447			
6	8,682			
7	4,306			
8	2,876			
9	2,162			
10	2,226			
Total	642,728			

Table A9: Tracking Results for Communities between 2016 and 2017

Step	Communities	Reason for Adjustment	Related Code (2016)	Related Code (2017)
0	637,228			
1	890			
	890	reclassification	*1390*	*1301, 1306*
2	9,020			
	265	reclassification	*130428*	*130407*
	410	reclassification	*130429*	*130408*
	158	abolishment	*130421*	*130402, 130471*
	109	reassignment	*130427*	*130473*
	250	reclassification	*130823*	*130881*
	316	reclassification	*330185*	*330112*
	75	abolishment	*330204*	*330212*
	398	reclassification	*330283*	*330213*
	194	reassignment	*330212*	*330203*
	319	reclassification	*331021*	*331083*
	148	reclassification	*360421*	*360404*
	294	reclassification	*360721*	*360704*
	181	reclassification	*361029*	*361003*
	910	reclassification	*370181*	*370114*
	1,097	reclassification	*370282*	*370215*
	70	abolishment	*410211*	*410202*
	438	reclassification	*411023*	*411003*
	271	reclassification	*430124*	*430182*
	100	establishment	*440306*	*440309*
	24	establishment	*440307*	*440310*
	210	reclassification	*451281*	*451203*
	343	reclassification	*500228*	*500155*
	210	reclassification	*500232*	*500156*
	207	reclassification	*510124*	*510117*
	126	reclassification	*510626*	*510604*
	413	reclassification	*511028*	*511083*
	504	reclassification	*520222*	*520281*
	136	reclassification	*530122*	*530115*
	190	reclassification	*610125*	*610118*
	313	reclassification	*610721*	*610703*
	341	reclassification	*610821*	*610882*
3	6,722			
4	302			
5	477			
6	2,436			
7	2,722			
8	227			
9	921			
10	2,056			
Total	663,001			