10_SBC_final

April 17, 2020

1 Final tables

1.1 Instruction Notebook

The notebook allows the user to construct to different level of aggregation:

- `industry`
- `geocode4_corr` -> city

By default, the `aggregation_param` parameter is set to `industry`. To switch to city, change for `geocode4_corr`. Then launch the notebook for new results

In the paper, we define a threshold to distinguish cities or industries. Choose among this set of threshold:

- 5
- 6
- 7
- 8
- mean

By default, the `threshold_full` parameter is set to `6`. To switch to another threshold, change the threshold `threshold_full`. Then launch the notebook for new results

```
Service account storage and Bigquery are now connected.
Service account storage is stored as <google.cloud.storage.client.Client object
at 0xa242a2890> and accessible with "Storage_account"
Service account Bigquery is stored as <google.cloud.bigquery.client.Client
object at 0xa26923950> and accessible with "bigquery_account"
```

2 Load Data

2.1 Paper dataset

```
/Users/thomas/anaconda3/lib/python3.7/site-packages/pyarrow/feather.py:83:
FutureWarning: The SparseDataFrame class is removed from pandas. Accessing it
from the top-level namespace will also be removed in the next version
  if isinstance(df, _pandas_api.pd.SparseDataFrame):
```

2.2 Compute Herfhindal: proxy Size

$$H = \sum_{i=1}^{N} s_i^2$$

where s_i is the market share of industry[city] i in a city [industry], and N is the number of firms.

We proceed as follow: - Step 1: Compute the share [output, capital, employment] by city-industry: `market_share_cit` - Step 2: compute the sum of squared market share by industry[city]: `Herfindahl_agg_t` - Step 3: Compute the average across time: `Herfindahl_agg` - Step 4: Compute the deciles of step 3: `decile_herfhindal_agg` - Low decile implies a low concentration within sectors - High decile implies a high concentration within sectors

2.3 Compute Ownership: proxy Foreign/SOE

$$\sum output_{agg,o} / \sum output_{agg}$$

- with agg stands for industry[city]
- o stands for ownership (Foreign vs Domestic or SOE vs private)

2.3.1 Foreign vs domestic

We proceed as follow: - Step 1: Compute the share [output, capital, employment] by industry[city], ownership (Foreign/Domestic): `Share_X_agg_o` - ~Step 2: Compute dummy when share Foreign above share domestic by industry[city]~ - Step 3: Compute decile by industry[city]-ownership - Note, high decile in Foreign means the industry[city] has relatively high share of foreign output, but not in absolule value as in step 2. A decile 9 in foreign can be a decile 2 or 3 in Domestic

2.3.2 SOE vs Private

We proceed as follow: - Step 1: Compute the share [output, capital, employment] by industry[city], ownership (SOE/PRIVATE): `Share_X_agg_o` - ~Step 2: Compute dummy when share SOE above share domestic by industry[city]~ - Step 3: Compute decile by industry[city]-ownership - Note, high decile in SOE means the industry[city] has relatively high share of SOE output, but not in absolule value as in step 2. A decile 9 in SOE can be a decile 2 or 3 in Domestic

2.4 Load TCZ_list_china from Google Spreadsheet

Feel free to add description about the dataset or any usefull information.

Profiling will be available soon for this dataset

2.5 Load chinese_city_characteristics from Google Spreadsheet

Feel free to add description about the dataset or any usefull information.

Profiling will be available soon for this dataset

2.6 Create R tables

3 Table 3

1. Full sample

$$LogSO2emission_{ikt} = \alpha\,(\text{ Period } \times \text{ Target }_i \times \text{ Polluting sectors }_k) + \nu_i + \lambda_t + \phi_k + \epsilon_{ikt}$$

$$LogSO2emission_{ikt} = \alpha\,(\text{ Period } \times \text{ Target }_i \times \text{ Polluting sectors }_k) + \nu_{ct} + \lambda_{kt} + \phi_{ck} + \epsilon_{ikt}$$

2. SOE dominated
3. Foreign dominated

1. TRUE 2. TRUE 3. TRUE

Table 1: Table Baseline Panel 1 - SOE

	Dependent variable SO2 emission $_{ikt}$							
	Full sample		Output		Capital		Employment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
output$_{cit}$	0.144	−0.066	0.875*	0.216	1.092**	0.308	0.995**	0.087
	(0.152)	(0.089)	(0.457)	(0.257)	(0.464)	(0.397)	(0.421)	(0.280)
capital$_{cit}$	1.481**	0.971***	−3.614***	−0.611	−4.106***	−0.485	−4.094***	−0.393
	(0.730)	(0.366)	(1.088)	(0.617)	(1.146)	(0.773)	(1.104)	(0.708)
labour$_{cit}$	3.089***	1.538**	11.671***	4.601***	12.228***	5.216***	12.327***	4.783***
	(0.832)	(0.767)	(2.326)	(1.475)	(2.374)	(1.353)	(2.359)	(1.548)
target$_c$ × Period	−0.003		0.087		0.012		0.128	
	(0.104)		(0.381)		(0.379)		(0.379)	
target$_c$ × Polluted$_i$	0.438***		0.838**		0.748*		0.839**	
	(0.140)		(0.383)		(0.390)		(0.379)	
target$_c$ × Period × Polluted$_i$	−0.291**	−0.430***	−0.430	0.009	−0.344	−0.159	−0.468	−0.215
	(0.146)	(0.132)	(0.465)	(0.465)	(0.443)	(0.435)	(0.432)	(0.461)
City fixed effects	Yes	No	Yes	No	Yes	No	Yes	No
Industry fixed effects	Yes	No	Yes	No	Yes	No	Yes	No
Year fixed effects	Yes	No	Yes	No	Yes	No	Yes	No
City-year fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Industry-year fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
City-industry fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	30,676	30,676	9,165	9,165	9,149	9,149	9,011	9,011
R^2	0.346	0.851	0.377	0.872	0.376	0.868	0.373	0.869

Due to limited space, only the coefficients of interest are presented for the regressions with city,industry, year fixed effect (i.e. columns 1-3). * Significance at the 10%, ** Significance at the 5%, *** Significance at the 1% heteroscedasticity-robust standard errors in parentheses are clustered by city

Table 1: Table Baseline Panel 2 - No SOE

	Dependent variable SO2 emission $_{ikt}$							
	Full sample		Output		Capital		Employment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
output$_{cit}$	0.144	−0.066	0.116	−0.077	0.107	−0.077	0.113	−0.073
	(0.152)	(0.089)	(0.134)	(0.094)	(0.130)	(0.090)	(0.130)	(0.095)
capital$_{cit}$	1.481**	0.971***	1.885***	0.942**	1.930***	0.905**	1.744***	0.911**
	(0.730)	(0.366)	(0.575)	(0.445)	(0.557)	(0.401)	(0.551)	(0.441)
labour$_{cit}$	3.089***	1.538**	2.592***	1.423*	2.567***	1.329*	2.611***	1.376
	(0.832)	(0.767)	(0.718)	(0.798)	(0.707)	(0.778)	(0.731)	(0.847)
target$_c$ × Period	−0.003		−0.050		−0.031		−0.049	
	(0.104)		(0.105)		(0.105)		(0.105)	
target$_c$ × Polluted$_i$	0.438***		0.388***		0.394***		0.369**	
	(0.140)		(0.148)		(0.145)		(0.150)	
target$_c$ × Period × Polluted$_i$	−0.291**	−0.430***	−0.250*	−0.429***	−0.284*	−0.430***	−0.256*	−0.434***
	(0.146)	(0.132)	(0.150)	(0.141)	(0.151)	(0.140)	(0.151)	(0.144)
City fixed effects	Yes	No	Yes	No	Yes	No	Yes	No
Industry fixed effects	Yes	No	Yes	No	Yes	No	Yes	No
Year fixed effects	Yes	No	Yes	No	Yes	No	Yes	No
City-year fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Industry-year fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
City-industry fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	30,676	30,676	21,511	21,511	21,527	21,527	21,665	21,665
R^2	0.346	0.851	0.355	0.855	0.359	0.856	0.358	0.856

Due to limited space, only the coefficients of interest are presented for the regressions with city,industry, year fixed effect (i.e. columns 1-3). * Significance at the 10%, ** Significance at the 5%, *** Significance at the 1% heteroscedasticity-robust standard errors in parentheses are clustered by city

4 Table 4 parallel trend

$$\log \text{SO2 emission }_{ikt} = \sum_{t=2002}^{2007} \alpha \left(Target_i \times \text{ Polluting sectors }_k \times year_t \right) + \theta X_{ikt} + \nu_{ik} + \lambda_{it} + \phi_{kt} + \epsilon_{ikt}$$

1. Full sample
2. SOE dominated vs No SOE dominated
   - output
   - capital
   - employment

1. TRUE 2. TRUE 3. TRUE 4. TRUE 5. TRUE 6. TRUE

Table 1: Parallel trend threshold used decile 6

	Dependent variable SO2 emission $_{ikt}$						
		Output		Capital		employment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Full sample	No SOE	SOE	No SOE	SOE	No SOE	SOE
2003	−0.230	−0.232	−0.581	−0.214	−0.326	−0.250	−0.308
	(0.222)	(0.248)	(0.838)	(0.243)	(0.833)	(0.242)	(0.846)
2004	−0.156	−0.174	0.148	−0.197	0.498	−0.229	0.498
	(0.222)	(0.240)	(0.808)	(0.237)	(0.772)	(0.239)	(0.807)
2005	−0.347	−0.364	−0.290	−0.350	−0.137	−0.307	−0.336
	(0.258)	(0.277)	(0.886)	(0.273)	(0.884)	(0.271)	(0.905)
2006	−0.598**	−0.645**	−0.082	−0.625**	−0.019	−0.630**	−0.246
	(0.263)	(0.279)	(0.950)	(0.276)	(0.956)	(0.276)	(0.978)
2007	−0.715***	−0.686**	−0.308	−0.703**	−0.247	−0.712**	−0.273
	(0.265)	(0.280)	(0.926)	(0.276)	(0.902)	(0.282)	(0.947)
City-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City-industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30,676	21,511	9,165	21,527	9,149	21,665	9,011
R^2	0.852	0.855	0.872	0.856	0.868	0.856	0.869

Due to limited space, only the coefficients of interest are presented for the regressions with city,industry, year fixed effect (i.e. columns 1-3). * Significance at the 10%, ** Significance at the 5%, *** Significance at the 1% heteroscedasticity-robust standard errors in parentheses are clustered by city

5 Diffusion chanels

- TCZ VS non TCZ
- Concentrated VS non concentrated
- Kuznet
  - TCZ
  - Concentration
  - SOE
- TFP

5.1 TCZ VS non TCZ

1. TRUE 2. TRUE 3. TRUE

Table 1: Diffusion Chanel TCZ VS No TCZ

	Dependent variable SO2 emission $_{ikt}$			
	TCZ		No TCZ	
	(1)	(2)	(3)	(4)
target$_c$ × Period	−0.012		0.899	
	(0.111)		(0.874)	
target$_c$ × Polluted$_i$	0.388***		2.570**	
	(0.146)		(1.182)	
target$_c$ × Period × Polluted$_i$	−0.254	−0.413***	−1.335	−1.415
	(0.155)	(0.138)	(1.128)	(1.059)
City fixed effects	Yes	No	Yes	No
Industry fixed effects	Yes	No	Yes	No
Year fixed effects	Yes	No	Yes	No
City-year fixed effects	No	Yes	No	Yes
Industry-year fixed effects	No	Yes	No	Yes
City-industry fixed effects	No	Yes	No	Yes
Observations	23,333	23,333	7,343	7,343
R^2	0.346	0.847	0.416	0.892

Due to limited space, only the coefficients of interest are presented for the regressions with city,industry, year fixed effect (i.e. columns 1-3). * Significance at the 10%, ** Significance at the 5%, *** Significance at the 1% heteroscedasticity-robust standard errors in parentheses are clustered by city

5.2 Concentrated VS non concentrated

1. TRUE 2. TRUE 3. TRUE

Table 1: Diffusion Chanel TCZ VS No TCZ

	Dependent variable SO2 emission $_{ikt}$			
	Concentrated		No Concentrated	
	(1)	(2)	(3)	(4)
target$_c$ × Period	−0.012		0.899	
	(0.111)		(0.874)	
target$_c$ × Polluted$_i$	0.388***		2.570**	
	(0.146)		(1.182)	
target$_c$ × Period × Polluted$_i$	−0.254	−0.413***	−1.335	−1.415
	(0.155)	(0.138)	(1.128)	(1.059)
City fixed effects	Yes	No	Yes	No
Industry fixed effects	Yes	No	Yes	No
Year fixed effects	Yes	No	Yes	No
City-year fixed effects	No	Yes	No	Yes
Industry-year fixed effects	No	Yes	No	Yes
City-industry fixed effects	No	Yes	No	Yes
Observations	23,333	23,333	7,343	7,343
R^2	0.346	0.847	0.416	0.892

Due to limited space, only the coefficients of interest are presented for the regressions with city,industry, year fixed effect (i.e. columns 1-3). * Significance at the 10%, ** Significance at the 5%, *** Significance at the 1% heteroscedasticity-robust standard errors in parentheses are clustered by city

5.3 Kuznet

Estimate the following model

$$LogSO2emission_{ikt} = \alpha(\text{ Period } \times \text{ Target } i \times \text{ Polluting sectors } k) + Kuznet + \nu i + \lambda_t + \phi_k + \epsilon_{ikt}$$

$$LogSO2emission_{ikt} = \alpha(\text{ Period } \times \text{ Target } i \times \text{ Polluting sectors } k) + Kuznet + \nu_{ct} + \lambda_{kt} + \phi_{ck} + \epsilon_{ikt}$$

1. TCZ vs No TCZ
2. Concentrated vs No Concentrated
3. SOE vs No SOE
   - output
   - capital
   - employment

1. TRUE 2. TRUE 3. TRUE

Table 1: Diffusion Chanel Kuznet Decile 6

	Dependent variable: SO2 emission $_{ikt}$									
	City		Concentration		Output		Capital		Employment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	TCZ	No TCZ	Concentrated	No Concentrated	SOE dominated	SOE No dominated	SOE dominated	SOE No dominated	SOE dominated	SOE No dominated
(ln gdp per cap)$_{ct}$	2.708***	0.269	−0.626	3.391***	−0.096	2.611***	1.437	2.559***	0.434	2.984***
	(0.927)	(1.428)	(1.192)	(0.971)	(1.784)	(0.837)	(1.882)	(0.847)	(1.555)	(0.854)
(ln gdp per cap) squared$_{ct}$	−0.132***	0.002	0.039	−0.158***	0.017	−0.127***	−0.062	−0.126***	−0.015	−0.143***
	(0.045)	(0.071)	(0.061)	(0.047)	(0.090)	(0.041)	(0.096)	(0.041)	(0.080)	(0.042)
(ln population)$_{ct}$	0.062	0.011	0.328*	0.026	0.043	0.040	−0.093	0.055	0.054	0.036
	(0.103)	(0.168)	(0.191)	(0.102)	(0.211)	(0.096)	(0.221)	(0.094)	(0.218)	(0.092)
turning point RMB	28795	-	-	45396	-	30264	-	24867	-	35190
turning point Dollar	3568	-	-	5625	-	3750	-	3081	-	4361
City fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,865	7,330	9,153	21,042	9,160	21,035	9,137	21,058	8,993	21,202
R^2	0.344	0.415	0.405	0.340	0.376	0.353	0.375	0.357	0.372	0.357

Due to limited space, only the coefficients of interest are presented for the regressions with city,industry, year fixed effect (i.e. columns 1-3). * Significance at the 10%, ** Significance at the 5%, *** Significance at the 1% heteroscedasticity-robust standard errors in parentheses are clustered by city

5.4 TFP

$$TFP_{ikt} = \alpha\,(\text{ Period } \times \text{ Target } i \times \text{ Polluting sectors } k) + \nu i + \lambda_t + \phi_k + \epsilon_{ikt}$$

$$TFP_{ikt} = \alpha\,(\text{ Period } \times \text{ Target } i \times \text{ Polluting sectors } k) + \nu_{ct} + \lambda_{kt} + \phi_{ck} + \epsilon_{ikt}$$

1. SOE/No SOE
2. TCZ/No TCZ
3. Coastal/no Coastal
4. Turning point

6 Create Report