MODS 207 : Gas price and urbanization

Final presentation - 21st May

Nghia Nguyen Danh

Théo Nguyen Ngoc

Content

1. Research question
2. Literature
3. Our data
4. Empirical strategy
5. Results and robustness checks

1. Research question

What is the correlation between the price of gas and the level of urbanisation in a country?

- To understand the impact oil has on the development of countries
- To consider the possible development in an era without petrol or fossil fuel
- We will also consider the speed at which the urbanisation happens !

2. Literature review

- Several articles somehow related to the topic were found
- Instinctively, there must be a correlation !
- Our previous MODS206 project pointed out a correlation, based on 4 rich countries

3. Our data

Criterias	Old data	New data
# observations	112	808
# countries	4	76
# countries of continent	Europe (2), North America (2)	Asia (28), Europe (31), Africa (4), SA (7), Oceania (2), NA (4)
# observations of continent	Europe (56). North America (56)	Asia (281), Europe (338), Africa (47), SA (82), Oceania (21), NA (39)
# years	28	13
year frame	1992-2019	1991-2016
control variables	gdp, num_car, oil consumption, services, industry	gdp, oil consumption, oil production, services, industry, region

4. Empirical strategy

Step 1: From the last project, decide what we should expand on this project

Step 2: Collect, process and clean data

Step 3: Calculate the some variables (urban rate, oil_import, produce_oil)

Step 4: Make the simple regressions with both urban and urban_rate

Step 5: Add fixed effects (entity fixed effects at both region and country level) → make interpretation

Step 6: Add nonlinear part into models

Step 7: Add interaction effects to answer the research questions

Step 8: Make causal interpretation and conclusion

5. Results and robustness checks

5.1 Simple regression - urban vs urban_rate

Explained variables

OLS Regression Results

Dep. Variable:	urban	R-squared:	0.481
Model:	OLS	Adj. R-squared:	0.477
Method:	Least Squares	F-statistic:	18.78
Date:	Sun, 20 Jun 2021	Prob (F-statistic):	3.27e-13
Time:	07:47:40	Log-Likelihood:	-3238.8
No. Observations:	808	AIC:	6492.
Df Residuals:	801	BIC:	6524.
Df Model:	6		
Covariance Type:	cluster		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	-6.8069	18.271	-0.373	0.709	-42.618	29.004
gas_price	0.5981	1.833	0.326	0.744	-2.994	4.190
gdp	0.0004	9.56e-05	3.981	0.000	0.000	0.001
oil_cons	-0.0023	0.001	-3.230	0.001	-0.004	-0.001
oil_prod	0.0032	0.001	2.457	0.014	0.001	0.006
ind	0.3914	0.222	1.763	0.078	-0.044	0.827
ser	0.9573	0.239	4.004	0.000	0.489	1.426

Omnibus:	22.547	Durbin-Watson:	0.219
Prob(Omnibus):	0.000	Jarque-Bera (JB):	24.255
Skew:	-0.378	Prob(JB):	5.41e-06
Kurtosis:	3.385	Cond. No.	4.35e+05

Warnings:
[1] Standard Errors are robust to cluster correlation (cluster)

vs

OLS Regression Results

Dep. Variable:	urban_rate	R-squared:	0.143
Model:	OLS	Adj. R-squared:	0.136
Method:	Least Squares	F-statistic:	1.920
Date:	Sun, 20 Jun 2021	Prob (F-statistic):	0.0885
Time:	07:47:40	Log-Likelihood:	-145.68
No. Observations:	808	AIC:	305.4
Df Residuals:	801	BIC:	338.2
Df Model:	6		
Covariance Type:	cluster		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.1774	0.311	0.570	0.568	-0.432	0.787
gas_price	0.0765	0.051	1.489	0.136	-0.024	0.177
gdp	-3.259e-06	1.76e-06	-1.852	0.064	-6.71e-06	1.9e-07
oil_cons	6.803e-05	3.7e-05	1.841	0.066	-4.41e-06	0.000
oil_prod	-2.832e-05	2.63e-05	-1.075	0.282	-7.99e-05	2.33e-05
ind	0.0078	0.005	1.486	0.137	-0.002	0.018
ser	-0.0032	0.004	-0.820	0.412	-0.011	0.005

Omnibus:	106.767	Durbin-Watson:	0.278
Prob(Omnibus):	0.000	Jarque-Bera (JB):	162.505
Skew:	0.899	Prob(JB):	5.16e-36
Kurtosis:	4.262	Cond. No.	4.35e+05

Warnings:
[1] Standard Errors are robust to cluster correlation (cluster)

Clustered SEs

5. Results and robustness checks

5.2 Fixed effects - region level

Dep. Variable:	urban	R-squared:	0.606
Model:	OLS	Adj. R-squared:	0.600
Method:	Least Squares	F-statistic:	15.78
Date:	Sun, 20 Jun 2021	Prob (F-statistic):	1.64e-16
Time:	09:14:43	Log-Likelihood:	-3127.2
No. Observations:	808	AIC:	6282.
Df Residuals:	794	BIC:	6348.
Df Model:	13		
Covariance Type:	cluster		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	8.0176	20.144	0.398	0.691	-31.463	47.498
gas_price	3.8243	2.632	1.453	0.146	-1.334	8.983
gdp	0.0004	0.000	4.320	0.000	0.000	0.001
oil_cons	-0.0015	0.001	-1.923	0.054	-0.003	2.81e-05
oil_prod	0.0025	0.001	2.178	0.029	0.000	0.005
ind	0.4577	0.227	2.014	0.044	0.012	0.903
ser	0.8683	0.258	3.359	0.001	0.362	1.375
asia	-22.6989	4.174	-5.438	0.000	-30.879	-14.518
europe	-19.8834	3.525	-5.640	0.000	-26.793	-12.974
africa	-20.2008	4.655	-4.340	0.000	-29.324	-11.077
NA	-21.3419	5.834	-3.658	0.000	-32.777	-9.907
oceania	-10.5839	4.530	-2.336	0.019	-19.463	-1.705
le_2000	4.1162	2.342	1.758	0.079	-0.473	8.706
le_2010	2.6449	1.315	2.012	0.044	0.068	5.222

Why are they positive?

Dep. Variable:	urban_rate	R-squared:	0.192
Model:	OLS	Adj. R-squared:	0.179
Method:	Least Squares	F-statistic:	4.577
Date:	Sun, 20 Jun 2021	Prob (F-statistic):	1.04e-05
Time:	07:47:41	Log-Likelihood:	-121.73
No. Observations:	808	AIC:	271.5
Df Residuals:	794	BIC:	337.2
Df Model:	13		
Covariance Type:	cluster		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.0705	0.308	0.229	0.819	-0.534	0.675
gas_price	0.0977	0.062	1.576	0.115	-0.024	0.219
gdp	-2.48e-06	1.91e-06	-1.296	0.195	-6.23e-06	1.27e-06
oil_cons	7.181e-05	3.48e-05	2.066	0.039	3.7e-06	0.000
oil_prod	-1.208e-05	3.31e-05	-0.365	0.715	-7.7e-05	5.28e-05
ind	0.0064	0.005	1.382	0.167	-0.003	0.016
ser	-0.0019	0.004	-0.491	0.623	-0.010	0.006
asia	0.0761	0.075	1.013	0.311	-0.071	0.223
europe	-0.0458	0.071	-0.644	0.519	-0.185	0.094
africa	0.1282	0.141	0.908	0.364	-0.148	0.405
NA	-0.2148	0.147	-1.465	0.143	-0.502	0.073
oceania	-0.0914	0.075	-1.226	0.220	-0.237	0.055
le_2000	0.0305	0.050	0.605	0.545	-0.068	0.129
le_2010	0.0412	0.037	1.114	0.265	-0.031	0.114

le_2000 > le_2010?

5. Results and robustness checks

5.3 Fixed effects - country level - urban

OLS Regression Results

Dep. Variable:	urban	R-squared:	0.989
Model:	OLS	Adj. R-squared:	0.988
Method:	Least Squares	F-statistic:	329.4
Date:	Sun, 20 Jun 2021	Prob (F-statistic):	3.57e-55
Time:	07:47:44	Log-Likelihood:	-1688.6
No. Observations:	808	AIC:	3545.
Df Residuals:	724	BIC:	3939.
Df Model:	83		
Covariance Type:	cluster		

	coef	std err	z	P>|z|	[0.025	0.975]
gas_price	0.5673	0.592	0.958	0.338	-0.594	1.728
gdp	-2.979e-05	4.66e-05	-0.639	0.523	-0.000	6.16e-05
oil_cons	0.0036	0.001	4.057	0.000	0.002	0.005
oil_prod	0.0001	0.001	0.189	0.850	-0.001	0.002
ser	-0.0023	0.063	-0.036	0.971	-0.127	0.122
ind	-0.0083	0.061	-0.136	0.891	-0.128	0.112
intercept	29.2955	4.353	6.730	0.000	20.763	37.828
le_2000	-3.9470	0.751	-5.256	0.000	-5.419	-2.475
le_2010	-2.0811	0.380	-5.481	0.000	-2.825	-1.337

The coefficient changed much from the model at region level (3.82 vs 0.56) although p-value is still very large.
→ we need the nonlinear model.

The coefficients now become reasonable and they are significant.

Continued with 75 binary country variables (76 countries)

5. Results and robustness checks

5.3 Fixed effects - country level - urban_rate

Dep. Variable:	urban_rate	R-squared:	0.793
Model:	OLS	Adj. R-squared:	0.770
Method:	Least Squares	F-statistic:	27.86
Date:	Sun, 20 Jun 2021	Prob (F-statistic):	1.53e-19
Time:	07:47:44	Log-Likelihood:	429.08
No. Observations:	808	AIC:	-690.2
Df Residuals:	724	BIC:	-295.8
Df Model:	83		
Covariance Type:	cluster		

	coef	std err	z	P>|z|	[0.025	0.975]
gas_price	-0.0078	0.043	-0.183	0.855	-0.092	0.076
gdp	1.564e-06	3.36e-06	0.465	0.642	-5.03e-06	8.16e-06
oil_cons	7.21e-06	3.68e-05	0.196	0.845	-6.49e-05	7.93e-05
oil_prod	-3.099e-05	4.49e-05	-0.690	0.490	-0.000	5.71e-05
ser	0.0067	0.004	1.628	0.103	-0.001	0.015
ind	0.0062	0.005	1.290	0.197	-0.003	0.016
intercept	0.1077	0.307	0.351	0.726	-0.494	0.709
le_2000	0.0189	0.044	0.432	0.666	-0.067	0.104
le_2010	0.0302	0.032	0.931	0.352	-0.033	0.094

The coefficient changed from positive to negative although p-value is still very large

The coefficients are still positive and le_2000 is still smaller than le_2010.
→ This is a characteristic of our data

Continued with 75 binary country variables (76 countries)

5. Results and robustness checks

5.4 Nonlinear models - urban (country level)

OLS Regression Results

Dep. Variable:	urban	R-squared:	0.989
Model:	OLS	Adj. R-squared:	0.988
Method:	Least Squares	F-statistic:	386.0
Date:	Mon, 14 Jun 2021	Prob (F-statistic):	3.15e-60
Time:	16:42:29	Log-Likelihood:	-1665.9
No. Observations:	808	AIC:	3504.
Df Residuals:	722	BIC:	3908.
Df Model:	85		
Covariance Type:	cluster		

	coef	std err	z	P>|z|	[0.025	0.975]
gas_price	13.5457	3.820	3.546	0.000	6.058	21.034
gdp	-1.515e-05	4.28e-05	-0.354	0.723	-9.91e-05	6.87e-05
oil_cons	0.0033	0.001	3.933	0.000	0.002	0.005
oil_prod	-5.421e-05	0.001	-0.077	0.939	-0.001	0.001
ser	-0.0401	0.063	-0.637	0.524	-0.163	0.083
ind	-0.0620	0.061	-1.021	0.307	-0.181	0.057
intercept	28.3586	4.143	6.844	0.000	20.238	36.479
le_2000	-3.6511	0.671	-5.445	0.000	-4.965	-2.337
le_2010	-2.0127	0.359	-5.605	0.000	-2.716	-1.309
is_Algeria	39.9308	1.141	34.999	0.000	37.695	42.167
is_Argentina	61.2658	0.694	88.308	0.000	59.906	62.626
is_Australia	55.9257	1.812	30.858	0.000	52.374	59.478
...						
is_Uzbekistan	20.2351	0.891	22.702	0.000	18.488	21.982
is_Venezuela	64.1633	2.059	31.163	0.000	60.128	68.199
gas_price^2	-10.5948	3.059	-3.464	0.001	-16.590	-4.600
gas_price^3	2.5808	0.743	3.471	0.001	1.124	4.038

After adding non linear parts, the p-values have become very small → the correlation is nonlinear rather than linear.

P-value: 1.68e-9 (F-test, nonlinear = 0)

Causal interpretation:

- When gas price increases from 1 to 2$ per litter, urbanization will **decrease 0.1731%**
- When gas price increases from 2 to 3$ per litter, urbanization will **increase 9.6087%**

5. Results and robustness checks

5.4 Nonlinear models - urban rate (country level)

OLS Regression Results

Dep. Variable:	urban_rate	R-squared:	0.794
Model:	OLS	Adj. R-squared:	0.770
Method:	Least Squares	F-statistic:	43.24
Date:	Mon, 14 Jun 2021	Prob (F-statistic):	4.18e-27
Time:	16:53:54	Log-Likelihood:	430.10
No. Observations:	808	AIC:	-688.2
Df Residuals:	722	BIC:	-284.5
Df Model:	85		
Covariance Type:	cluster		

	coef	std err	z	P>|z|	[0.025	0.975]
gas_price	-0.2087	0.268	-0.780	0.435	-0.733	0.316
gdp	1.33e-06	3.29e-06	0.404	0.686	-5.12e-06	7.78e-06
oil_cons	1.226e-05	3.69e-05	0.332	0.740	-6e-05	8.45e-05
oil_prod	-2.735e-05	4.61e-05	-0.594	0.553	-0.000	6.29e-05
ser	0.0073	0.004	1.627	0.104	-0.001	0.016
ind	0.0071	0.005	1.387	0.165	-0.003	0.017
intercept	0.1206	0.306	0.395	0.693	-0.479	0.720
le_2000	0.0140	0.040	0.345	0.730	-0.065	0.093
le_2010	0.0292	0.032	0.912	0.362	-0.034	0.092
is_Algeria	-0.0131	0.109	-0.121	0.904	-0.226	0.200
is_Argentina	-0.4991	0.047	-10.599	0.000	-0.591	-0.407
...						
is_Venezuela	-0.6482	0.160	-4.051	0.000	-0.962	-0.335
gas_price^2	0.1604	0.214	0.751	0.453	-0.258	0.579
gas_price^3	-0.0382	0.052	-0.727	0.467	-0.141	0.065

P-value: 0.40 (F-test, nonlinear = 0)

It seems less useful to use the polynomial regression when we look at the urbanisation rate.

Can not make the causal interpretation between gas_price and urbanization rate (the coefficients are not significant, both linear and nonlinear case)

5. Results and robustness checks

5.5 Interaction effects - oil_import - urban

OLS Regression Results

Dep. Variable:	urban	R-squared:	0.990
Model:	OLS	Adj. R-squared:	0.988
Method:	Least Squares	F-statistic:	324.8
Date:	Mon, 14 Jun 2021	Prob (F-statistic):	2.65e-60
Time:	17:21:56	Log-Likelihood:	-1653.9
No. Observations:	808	AIC:	3486.
Df Residuals:	719	BIC:	3904.
Df Model:	88		
Covariance Type:	cluster		

	coef	std err	z	P>|z|	[0.025	0.975]
gas_price	14.5661	3.780	3.853	0.000	7.157	21.975
gdp	-6.667e-06	5.04e-05	-0.132	0.895	-0.000	9.2e-05
oil_cons	0.0027	0.001	3.053	0.002	0.001	0.004
ser	-0.0443	0.062	-0.710	0.478	-0.167	0.078
ind	-0.0572	0.061	-0.933	0.351	-0.177	0.063
intercept	28.1738	4.141	6.804	0.000	20.058	36.290
le_2000	-3.6226	0.678	-5.345	0.000	-4.951	-2.294
le_2010	-2.0104	0.365	-5.506	0.000	-2.726	-1.295
is_Algeria	39.9506	1.109	36.027	0.000	37.777	42.124
is_Argentina	61.4143	0.753	81.513	0.000	59.938	62.891
...						
gas_price^2	-11.6529	3.039	-3.835	0.000	-17.609	-5.697
gas_price^3	2.7934	0.729	3.830	0.000	1.364	4.223
oil_import	0.0009	0.001	0.980	0.327	-0.001	0.003
gas_price X oil_import	-0.0032	0.003	-1.143	0.253	-0.009	0.002
gas_price^2 X oil_import	0.0037	0.003	1.372	0.170	-0.002	0.009
gas_price^3 X oil_import	-0.0008	0.001	-1.123	0.261	-0.002	0.001

P-value: 8.38e-7 (F-test, all cross variables = 0) → The interaction effect is significant.

Causal interpretation:

- Oil_import = 100TWh
  - When gas price increases from 1 to 2$ per litter, urbanization will **decrease 0.6088%**
  - When gas price increases from 2 to 3$ per litter, urbanization will **increase 9.3862%**
- Oil_import = -100TWh
  - When gas price increases from 1 to 2$ per litter, urbanization will **decrease 1.0688%**
  - When gas price increases from 2 to 3$ per litter, urbanization will **increase 9.3662%**

5. Results and robustness checks

5.7 Compare regions to Asia (bonus)

OLS Regression Results

Dep. Variable:	urban	R-squared:	0.606
Model:	OLS	Adj. R-squared:	0.600
Method:	Least Squares	F-statistic:	15.78
Date:	Sun, 20 Jun 2021	Prob (F-statistic):	1.64e-16
Time:	15:44:13	Log-Likelihood:	-3127.2
No. Observations:	808	AIC:	6282.
Df Residuals:	794	BIC:	6348.
Df Model:	13		
Covariance Type:	cluster		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	-14.6813	19.967	-0.735	0.462	-53.817	24.454
gas_price	3.8243	2.632	1.453	0.146	-1.334	8.983
gdp	0.0004	0.000	4.320	0.000	0.000	0.001
oil_cons	0.0011	0.001	1.028	0.304	-0.001	0.003
oil_import	-0.0025	0.001	-2.178	0.029	-0.005	-0.000
ind	0.4577	0.227	2.014	0.044	0.012	0.903
ser	0.8683	0.258	3.359	0.001	0.362	1.375
SA	22.6989	4.174	5.438	0.000	14.518	30.879
europe	2.8155	4.165	0.676	0.499	-5.348	10.979
africa	2.4981	4.983	0.501	0.616	-7.269	12.265
NA	1.3570	5.375	0.252	0.801	-9.178	11.892
oceania	12.1150	4.798	2.525	0.012	2.710	21.520
le_2000	4.1162	2.342	1.758	0.079	-0.473	8.706
le_2010	2.6449	1.315	2.012	0.044	0.068	5.222

We use the model at region level, with entity fixed effects at region level (taking Asia as a reference).

Only the coefficients of SA and Oceania are significant (p-values < 0.05). We can conclude:

- An observation in SA with the same profile as one in Asia will have the higher urbanization (22.69%)
- An observation in Oceania with the same profile as one in Asia will have the higher urbanization (12.1150%)

Conclusion

- The gas price does have a positive impact on urbanization (at country level), but this correlation is nonlinear (cubic correlation)
- However, the gas_price seems to have no impact on urbanization rate at country level.
- The oil import has a huge effect on the correlation between gas price and urbanization (there was a big difference between an oil import country and an oil export country)

Thank you for your attention!