Promoting Green Technical Efficiency: A Regional and Urban Agglomeration Cluster Analysis Supplementary

Kai Du

School of Business, University of Wollongong, NSW 2522, Australia

Appendix S1 Methodology

The data on inputs (x_i^t), output (y_i^t) and emissions (b_i^t) as undesirable output can be used to estimate the true technology Ψ^t, where i represents each province, t is the time period. The sample size is $N = \sum_{t=1}^{2} n_t$ in this study and there are two time periods where n_t is the number of provinces in each period.

The DEA model estimates the true technology Ψ^t by enveloping all provinces in the sample as shown below:

$$\begin{aligned}\hat{E}_i^t = \max\{\delta :\ & \delta y_i^t \leq \textstyle\sum_{j=1}^{n_t} \lambda_j^t y_j^t, \\ & x_i^t \geq \textstyle\sum_{j=1}^{n_t} \lambda_j^t x_j^t, \\ & b_i^t = \textstyle\sum_{j=1}^{n_t} \lambda_j^t b_j^t, \\ & \delta \geq 0,\ \lambda_j^t \geq 0,\ j = 1, \dots, n_t\}\end{aligned}$$

where $\lambda_1^t, \dots, \lambda_{n_t}^t$ are intensity variables, which are also the weight of each observation in the reference set, and the summation of λs is given by $\sum_{j=1}^{n_t} \lambda_j^t y_j^t$ for desirable output and j refers to each province within each constraint.

The TE score for a province i is then estimated when the efficiency measure δ is maximised with the following three constraints related to desirable output ($\delta y_i^t \leq \sum_{j=1}^{n_t} \lambda_j^t y_j^t$), inputs ($x_i^t \geq \sum_{j=1}^{n_t} \lambda_j^t x_j^t$) and emissions/undesirable output ($b_i^t = \sum_{j=1}^{n_t} \lambda_j^t b_j^t$).

The first constraint $\delta y_i^t \leq \sum_{j=1}^{n_t} \lambda_j^t y_j^t$ shows that, given the maximised δ, the desirable output y_i^t of the province i in the period t should be smaller than or equal to the weighted average of the desirable output of all provinces in period t. The second constraint $x_i^t \geq \sum_{j=1}^{n_t} \lambda_j^t x_j^t$ shows that the inputs x_i^t have to be equal to (or at least be as large as) the weighted average of the inputs of all provinces in period t. The third constraint $b_i^t = \sum_{j=1}^{n_t} \lambda_j^t b_j^t$ shows that the emissions/undesirable output b_i^t has to be equal to the weighted average of the emissions/undesirable output of all provinces in period t.

While extending the SW approach to include undesirable outputs is relatively straightforward, the bootstrapping procedure warrants a detailed explanation:

Step 1. For each period t ($t = 1$ and 2), we use the original data, denoted as $S_{n_t}^t := \{(x_i^t, y_i^t): i = 1, \dots, n_t\}$, to compute $\hat{E}_i^t$ *separately* for each period t.

Step 2. Organise the DEA efficiency estimates and their regressors into the panel data set $S_N := \left\{\left\{\left(\hat{E}_i^1, Z_i^1, D^1\right)\right\}_{i=1}^{n_t}, \left\{\left(\hat{E}_i^2, Z_i^2, D^2\right)\right\}_{i=1}^{n_t}, \dots, \left\{\left(\hat{E}_i^T, Z_i^T, D^T\right)\right\}_{i=1}^{n_t}\right\}$, with sample size $N = \sum_{t=1}^{T} n_t$. Exclude the observations on the boundary (i.e., remove the "spuriously efficient" observations and use only $m_t < n_t$ observations for which at this stage) and use the method of maximum likelihood estimation to obtain the estimated and in the truncated regression of on and , denoting them as and .$\hat{E}_i^t > 1 \beta, \gamma \sigma_\epsilon \hat{E}_i^t Z_i^t D^t \hat{\beta}, \hat{\gamma} \hat{\sigma}_\epsilon^2$

Step 3. Loop over the next four steps L_1 times to obtain a set of bias-corrected estimates $\mathcal{B}_i^t = \{E_{i,b}^{t^*}\}_{b=1}^{L_1}$ as follows:

Step 3.1. For each $i = 1, \dots, n_t$ and $t = 1$ and 2, draw $\hat{\epsilon}_{i,b}^t$ from $\mathcal{N}(0, \hat{\sigma}_\epsilon^2)$ distribution with truncation on the left at $(1 - Z_i^t \hat{\beta} - D^t \hat{\gamma})$.

Step 3.2. For each $i = 1, \dots, n_t$ and $t = 1$ and 2, compute the bootstrap analogues of the efficiency scores as $E_{i,b}^{t^*} = Z_i^t \hat{\beta} + D^t \hat{\gamma} + \hat{\epsilon}_{i,b}^t$.

Step 3.3. Define $x_{i,b}^{t^*} = x_i^t$, $y_{i,b}^{t^*} = (\hat{E}_i^t / E_{i,b}^{t^*}) y_i^t$, $Z_{i,b}^{t^*} = Z_i^t$ for all $i = 1, \dots, n_t$ and $t = 1$ and 2.

Step 3.4. *Separately* for each period ($t = 1$ and 2), compute $\hat{E}_{i,b}^{t^*}$ using formulation (2), but after replacing y_j^t and x_j^t with their bootstrapping analogues $y_{j,b}^{t^*}$ and $x_{j,b}^{t^*}$ for all $j = 1, \dots, n_t$.

Step 4. For each $i = 1, \dots, n_t$ and $t = 1$ and 2, compute the bias-corrected estimates $\hat{\hat{E}}_i^t$ defined by $\hat{\hat{E}}_i^t = \hat{E}_i^t - B(\hat{E}_i^t)$, where $B(\hat{E}_i^t)$ is the bootstrap-based estimate of the bias of $\hat{E}_i^t$ using the bootstrapping estimates in $\mathcal{B}_i^t$ obtained in Step 3.

Step 5. Organise the bias-corrected efficiency estimates and their regressors into the panel data again $S_N := \left\{\left\{\left(\hat{\hat{E}}_i^1, Z_i^1, D^1\right)\right\}_{i=1}^{n_t}, \left\{\left(\hat{\hat{E}}_i^2, Z_i^2, D^2\right)\right\}_{i=1}^{n_t}, \dots, \left\{\left(\hat{\hat{E}}_i^T, Z_i^T, D^T\right)\right\}_{i=1}^{n_t}\right\}$, and use the method of maximum likelihood on the full sample (of size $N = \sum_{t=1}^{2} n_t$) to estimate the truncated regression of $\hat{\hat{E}}_i^t$ on Z_i^t and D^t, yielding new (refined) estimates of the regression, denoting them as ($\hat{\hat{\beta}}$, $\hat{\hat{\gamma}}$ and $\hat{\hat{\sigma}}_\epsilon$).

Step 6. Loop over the next three steps L_2 times to obtain a set of bootstrap analogues of parameters of the regression $\{(\hat{\beta}^*, \hat{\gamma}^*, \hat{\sigma}_\epsilon^*)_b\}_{b=1}^{L_2}$ as follows:

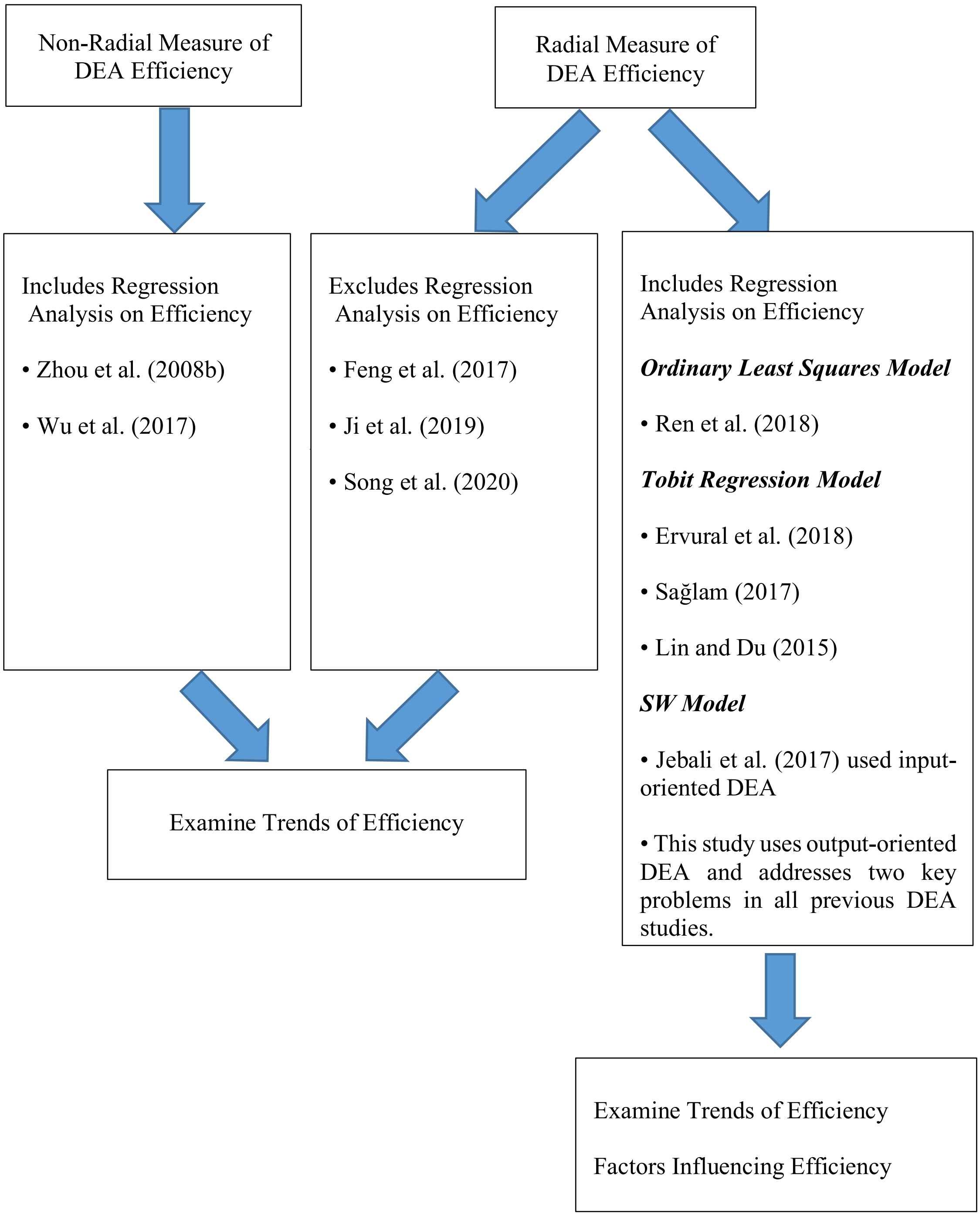
Step 6.1. For each observation ($i = 1, \dots, n_t$ and $t = 1$ and 2), draw $\hat{\hat{\epsilon}}_{i,b}^t$ from $\mathcal{N}(0, \hat{\hat{\sigma}}_\epsilon)$ with left-truncation at $(1 - Z_i^t \hat{\hat{\beta}} - D^t \hat{\hat{\gamma}})$.

Step 6.2. Obtain the double-bootstrap analogues of efficiency scores as follows: $E_{i,b}^{t^{**}} = Z_i^t \hat{\hat{\beta}} + D^t \hat{\hat{\gamma}} + \hat{\hat{\epsilon}}_{i,b}^t$, for each $i = 1, \dots, n_t$ and $t = 1$ and 2.

Step 6.3. Use the maximum likelihood method to estimate the truncated regression of $E_{i,b}^{t^{**}}$ on Z_i^t and D^t, yielding estimates $\hat{\hat{\beta}}^*$, $\hat{\hat{\gamma}}^*$ and $\hat{\hat{\sigma}}_\epsilon{}^*$.

Step 7. Use the bootstrapping values in $\{(\hat{\beta}^*, \hat{\gamma}^*, \hat{\sigma}_\epsilon^*)_b\}_{b=1}^{L_2}$and the refined estimates $\hat{\hat{\beta}}$, $\hat{\hat{\gamma}}$ and $\hat{\hat{\sigma}}_\epsilon$ to construct bootstrap-based confidence intervals (CIs) for each element of β, γ and σ_ϵ as reported in Appendix 6.

Appendix S2 Literature Review on DEA Efficiency Related to Pollutants



Appendix S3 Correlation and Variance Inflation Factor of Regressors

	Correlation Matrix					Variance Inflation Factor (VIF)
	Urbanisation	Services	Energy	Education	Openness	
Urbanisation	1.000	0.719	0.061	0.810	0.811	5.084
Services	0.719	1.000	-0.108	0.850	0.678	4.088
Energy	0.061	-0.108	1.000	-0.123	0.111	1.118
Education	0.810	0.850	-0.123	1.000	0.651	5.552
Openness	0.811	0.678	0.111	0.651	1.000	3.323

Notes: The high correlations among Urbanisation, Services, Education, and Openness reflect a common development pathway — as provinces urbanise, they simultaneously expand services, increase educational attainment, and open to trade. Energy consumption is largely orthogonal to these variables (all correlations below 0.13 in absolute value), explaining its stable coefficient across all model specifications in Table 5. VIFs are computed from the full five-regressor model. While all VIFs are below the conventional threshold of 10, Education (5.55) and Urbanisation (5.08) approach levels where individual coefficient precision is materially affected. See Models 2–6 in Table 5 for robustness checks that address this concern.

Appendix S4 Truncated Regression Results for Efficiency (Energy Intensity as DEA Input)

Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Panel A: Economic Growth with Carbon Emissions					
Urbanisation	-0.634***	-1.267***	-0.428***	-0.933***	-0.877***	
Service Sector	-0.368***	-0.622***	-0.299***	-0.522***		-1.441***
Education	0.306**	0.546***				
Openness	-0.301***		-0.388***			
Time Dummy	-0.257**	-0.572***	-0.423***	-0.695***	-0.445**	-1.034***
Intercept	0.843***	0.441***	0.909***	0.477**	0.763***	-0.152*
$\hat{\sigma}$	0.231	0.387	0.241	0.425	0.328	0.793
Log-Likelihood	260.979	258.614	256.703	247.551	242.323	225.141
	Panel B: Economic Growth without Carbon Emissions					
Urbanisation	-0.647***	-0.843***	-0.452***	-0.650***	-0.677***	
Service Sector	-0.139*	-0.165*	-0.095	-0.108		-0.632**
Education	0.282***	0.375***				
Openness	-0.237***		-0.308***			
Time Dummy	-0.259**	-0.311***	-0.384***	-0.511***	-0.456***	-0.691**
Intercept	1.200***	1.238***	1.239***	1.303***	1.304***	1.178***
$\hat{\sigma}$	0.204	0.215	0.218	0.241	0.237	0.414
Log-Likelihood	222.272	214.991	216.996	205.592	204.169	172.475

Notes: A negative coefficient denotes an improvement in TE. DEA inputs: employment, capital stock, and energy intensity. Algorithm II of SW (2007) was used with 2000 iterations for the bootstrapping to obtain confidence intervals for valid inference. As this table serves as a robustness check, only the summary results are reported here; full confidence intervals are available from the authors upon request. If the confidence interval for the estimate does not cover zero, the variable is statistically significant at the 10% (*), 5% (**), and 1% (***) levels of significance.

Appendix S5 Truncated Regression Results for Efficiency (Outlier Excluded)

Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Panel A: Economic Growth with Carbon Emissions					
Urbanisation	-0.335***	-0.311***	-0.264***	-0.361***	-0.302***	
Service Sector	-0.103***	-0.105**	-0.082**	-0.075*		-0.257***
Energy Use	-0.056***	-0.058***	-0.054**			
Education	0.108**	0.103**		0.105**		
Openness	0.027		0.009	0.027		
Time Dummy	-0.083	-0.075	-0.125***	-0.064	-0.066	-0.119*
Intercept	1.309***	1.301***	1.327***	1.296***	1.303***	1.297***
$\hat{\sigma}$	0.054***	0.054***	0.056***	0.056***	0.058***	0.081***
Log-Likelihood	280.776	280.381	277.195	277.081	271.709	245.204
	Panel B: Economic Growth without Carbon Emissions					
Urbanisation	-0.372***	-0.318***	-0.322***	-0.413***	-0.288***	
Service Sector	0.029	0.031	0.051	0.070**		-0.147***
Energy Use	-0.069***	-0.070***	-0.068***			
Education	0.079**	0.064		0.076*		
Openness	0.060**		0.045	0.060*		
Time Dummy	-0.078*	-0.060	-0.111**	-0.056	-0.112***	-0.099*
Intercept	1.480***	1.468***	1.499***	1.465***	1.481***	1.446***
$\hat{\sigma}$	0.052***	0.053***	0.053***	0.056***	0.063***	0.094***
Log-Likelihood	248.395	246.343	246.206	240.981	234.996	196.740

Notes: A negative coefficient denotes an improvement in TE. We exclude 26 observations (10.8%) with efficiency scores > 2.0 from the two-stage analysis: Hainan (2010–2017), Zhejiang (2010–2017), Jiangxi (2012–2017), and Hebei (2014–2017). Algorithm II of SW (2007) was used with 2000 iterations for the bootstrapping to obtain confidence intervals for valid inference. As this table serves as a robustness check, only the summary results are reported here; full confidence intervals are available from the authors upon request. If the confidence interval for the estimate does not cover zero, the variable is statistically significant at the 10% (*), 5% (**), and 1% (***) levels of significance.

Appendix S6 Detailed Benchmark Results with Confidence Intervals

Table S1 Confidence Intervals for Table 5 Estimates — Model 1, GDP/CO_2 as DEA Output

Variables	Coeffi.	90% CI	90% CI	95% CI	95% CI	99% CI	99% CI
Urbanisation	-0.438***	-0.603	-0.274	-0.634	-0.246	-0.680	-0.179
Service Sector	-0.323***	-0.451	-0.187	-0.472	-0.161	-0.515	-0.099
Education	0.096	-0.042	0.229	-0.067	0.250	-0.108	0.289
Energy Use	0.135***	0.061	0.213	0.046	0.223	0.018	0.245
Openness	-0.092	-0.203	0.013	-0.225	0.037	-0.259	0.075
Time Dummy	-0.363***	-0.543	-0.201	-0.574	-0.158	-0.624	-0.084
Intercept	1.292***	1.167	1.414	1.146	1.449	1.110	1.509
$\hat{\sigma}$	0.167	0.129	0.214	0.118	0.220	0.097	0.231
Log-Likelihood	235.161						

Table S2 Confidence Intervals for Table 5 Estimates — Model 1, GDP as DEA Output

Variables	Coeffi.	90% CI	90% CI	95% CI	95% CI	99% CI	99% CI
Urbanisation	-0.544***	-0.715	-0.378	-0.748	-0.350	-0.791	-0.272
Service Sector	-0.114	-0.225	0.011	-0.250	0.032	-0.291	0.087
Education	0.101	-0.042	0.241	-0.071	0.270	-0.148	0.323
Energy Use	0.138***	0.062	0.218	0.045	0.235	0.013	0.256
Openness	0.005	-0.104	0.115	-0.124	0.137	-0.172	0.179
Time Dummy	-0.288**	-0.461	-0.106	-0.490	-0.075	-0.566	0.005
Intercept	1.450***	1.334	1.565	1.312	1.597	1.272	1.640
$\hat{\sigma}$	0.195	0.150	0.246	0.137	0.251	0.114	0.262
Log-Likelihood	185.424						

Table S3 Confidence Intervals for Table 5 Estimates — Model 2, GDP/CO_2 as DEA Output

Variables	Coeffi.	90% CI	90% CI	95% CI	95% CI	99% CI	99% CI
Urbanisation	-0.522***	-0.668	-0.373	-0.688	-0.336	-0.737	-0.246
Service Sector	-0.334***	-0.469	-0.189	-0.492	-0.147	-0.532	-0.081
Education	0.109	-0.037	0.251	-0.068	0.278	-0.143	0.331
Energy Use	0.162***	0.094	0.237	0.081	0.254	0.056	0.275
Time Dummy	-0.397***	-0.575	-0.213	-0.607	-0.170	-0.659	-0.074
Intercept	1.296***	1.169	1.437	1.151	1.480	1.122	1.548
$\hat{\sigma}$	0.172	0.125	0.218	0.111	0.225	0.085	0.233
Log-Likelihood	234.767						

Table S4 Confidence Intervals for Table 5 Estimates — Model 2, GDP as DEA Output

Variables	Coeffi.	90% CI	90% CI	95% CI	95% CI	99% CI	99% CI
Urbanisation	-0.540***	-0.677	-0.400	-0.706	-0.376	-0.740	-0.303
Service Sector	-0.117	-0.235	0.005	-0.257	0.029	-0.295	0.087
Education	0.101	-0.044	0.250	-0.084	0.278	-0.154	0.340
Energy Use	0.138***	0.066	0.210	0.052	0.224	0.024	0.255
Time Dummy	-0.288***	-0.451	-0.117	-0.477	-0.083	-0.551	-0.020
Intercept	1.446***	1.335	1.566	1.321	1.596	1.291	1.653
$\hat{\sigma}$	0.196	0.148	0.248	0.138	0.254	0.109	0.266
Log-Likelihood	185.529						

Table S5 Confidence Intervals for Table 5 Estimates — Model 3, GDP/CO_2 as DEA Output

Variables	Coeffi.	90% CI	90% CI	95% CI	95% CI	99% CI	99% CI
Urbanisation	-0.416***	-0.589	-0.230	-0.624	-0.190	-0.703	-0.111
Service Sector	-0.362***	-0.501	-0.204	-0.528	-0.175	-0.579	-0.111
Education	0.204**	0.068	0.347	0.031	0.373	-0.038	0.420
Openness	-0.178***	-0.287	-0.069	-0.313	-0.047	-0.357	-0.003
Time Dummy	-0.320***	-0.495	-0.128	-0.525	-0.088	-0.594	-0.003
Intercept	1.220***	1.079	1.386	1.056	1.429	1.016	1.536
$\hat{\sigma}$	0.186	0.132	0.238	0.118	0.244	0.080	0.259
Log-Likelihood	231.554						

Table S6 Confidence Intervals for Table 5 Estimates — Model 3, GDP as DEA Output

Variables	Coeffi.	90% CI	90% CI	95% CI	95% CI	99% CI	99% CI
Urbanisation	-0.495***	-0.663	-0.324	-0.695	-0.284	-0.756	-0.218
Service Sector	-0.141*	-0.262	-0.019	-0.282	0.010	-0.319	0.063
Education	0.202**	0.060	0.342	0.031	0.364	-0.021	0.413
Openness	-0.086	-0.191	0.018	-0.215	0.035	-0.257	0.073
Time Dummy	-0.236**	-0.404	-0.076	-0.435	-0.038	-0.491	0.038
Intercept	1.417***	1.299	1.545	1.278	1.578	1.235	1.662
$\hat{\sigma}$	0.204	0.154	0.257	0.143	0.265	0.105	0.277
Log-Likelihood	181.672						

Table S7 Confidence Intervals for Table 5 Estimates — Model 4, GDP/CO_2 as DEA Output

Variables	Coeffi.	90% CI	90% CI	95% CI	95% CI	99% CI	99% CI
Urbanisation	-0.391***	-0.541	-0.255	-0.566	-0.226	-0.614	-0.162
Service Sector	-0.307***	-0.436	-0.174	-0.458	-0.154	-0.496	-0.093
Energy Use	0.158***	0.088	0.231	0.071	0.243	0.046	0.263
Openness	-0.095	-0.202	0.015	-0.228	0.036	-0.271	0.091
Time Dummy	-0.412***	-0.578	-0.250	-0.607	-0.212	-0.669	-0.167
Intercept	1.316***	1.199	1.443	1.182	1.472	1.157	1.537
$\hat{\sigma}$	0.167	0.126	0.212	0.111	0.218	0.084	0.224
Log-Likelihood	235.064						

Table S8 Confidence Intervals for Table 5 Estimates — Model 4, GDP as DEA Output

Variables	Coeffi.	90% CI	90% CI	95% CI	95% CI	99% CI	99% CI
Urbanisation	-0.495***	-0.655	-0.343	-0.681	-0.308	-0.725	-0.251
Service Sector	-0.096	-0.205	0.022	-0.231	0.052	-0.267	0.092
Energy Use	0.161***	0.091	0.239	0.075	0.249	0.041	0.280
Openness	0.002	-0.110	0.112	-0.130	0.134	-0.169	0.168
Time Dummy	-0.338***	-0.497	-0.185	-0.527	-0.155	-0.580	-0.063
Intercept	1.469***	1.362	1.590	1.343	1.617	1.319	1.668
$\hat{\sigma}$	0.198	0.150	0.247	0.139	0.254	0.109	0.268
Log-Likelihood	184.837						

Table S9 Confidence Intervals for Table 5 Estimates — Model 5, GDP/CO_2 as DEA Output

Variables	Coeffi.	90% CI	90% CI	95% CI	95% CI	99% CI	99% CI
Urbanisation	-0.585***	-0.740	-0.383	-0.765	-0.337	-0.816	-0.259
Time Dummy	-0.337***	-0.516	-0.150	-0.549	-0.106	-0.611	-0.029
Intercept	1.174***	1.002	1.406	0.980	1.459	0.945	1.611
$\hat{\sigma}$	0.243	0.151	0.314	0.126	0.323	0.083	0.341
Log-Likelihood	212.682						

Table S10 Confidence Intervals for Table 5 Estimates — Model 5, GDP as DEA Output

Variables	Coeffi.	90% CI	90% CI	95% CI	95% CI	99% CI	99% CI
Urbanisation	-0.485***	-0.596	-0.363	-0.610	-0.338	-0.645	-0.271
Time Dummy	-0.317***	-0.463	-0.172	-0.485	-0.136	-0.531	-0.057
Intercept	1.459***	1.346	1.590	1.333	1.624	1.305	1.691
$\hat{\sigma}$	0.217	0.155	0.272	0.140	0.281	0.104	0.294
Log-Likelihood	174.741						

Table S11 Confidence Intervals for Table 5 Estimates — Model 6, GDP/CO_2 as DEA Output

Variables	Coeffi.	90% CI	90% CI	95% CI	95% CI	99% CI	99% CI
Service Sector	-0.722***	-0.933	-0.442	-0.964	-0.371	-1.016	-0.177
Time Dummy	-0.621***	-0.861	-0.326	-0.897	-0.260	-0.961	-0.091
Intercept	1.209***	1.024	1.472	0.993	1.565	0.954	1.719
$\hat{\sigma}$	0.287	0.157	0.380	0.121	0.391	0.019	0.408
Log-Likelihood	207.760						

Table S12 Confidence Intervals for Table 5 Estimates — Model 6, GDP as DEA Output

Variables	Coeffi.	90% CI	90% CI	95% CI	95% CI	99% CI	99% CI
Service Sector	-0.455***	-0.604	-0.269	-0.628	-0.220	-0.669	-0.125
Time Dummy	-0.472***	-0.679	-0.223	-0.712	-0.166	-0.770	-0.073
Intercept	1.432***	1.293	1.613	1.272	1.650	1.226	1.814
$\hat{\sigma}$	0.304	0.197	0.391	0.168	0.405	0.097	0.428
Log-Likelihood	155.066						