Method Article

How to apply dynamic panel bootstrap-corrected fixed-effects (xtbcfe) and heterogeneous dynamics (panelhetero)☆

Samuel Asumadu Sarkodie*, Phebe Asantewaa Owusu

Nord University Business School (HHN), Post Box 1490, 8049 Bodø, Norway

A B S T R A C T

The characteristics of panel data namely, inter alia, missing values, cross-sectional dependence, serial correlation, small time period bias, omitted variable bias, country-specific fixed-effects, time effects, heterogeneous effects and convergence often lead to misspecification, and spurious regression, thus, affecting the consistency and robustness of the model. In this regard, a more sophisticated panel estimation technique that accounts for the attributes and challenges is worthwhile. The novel panel bootstrap-corrected fixed-effects estimator (*xtbcfe*) and heterogeneous dynamics (*panelhetero*) recommended in this study meets almost all the requirements for robust and consistent panel estimation with an interface for user modifications. We further demonstrate how to use empirical CDF, moments and kernel density estimation to investigate heterogeneous effects. Due to the complexities in the application of *xtbcfe* and *panelhetero* algorithm, we provide a step-by-step procedure and guidelines for the estimation approach. We apply the *xtbcfe* and *panelhetero* algorithm for global estimation of mortality, disability-adjusted life years and welfare cost from exposure to ambient air pollution. Importantly, the *xtbcfe* algorithm can be applied to any panel data-based studies in social science, environmental science, environmental economics, health economics, energy economics, and among others.

- Procedures useful for data imputation and transforming negative variables for time series, cross-sectional and panel data are presented.
- Contrary to traditional models, we show how a novel approach can be modified and used to examine the degree of heterogeneous effects across cross-sectional units of panel data.

- We demonstrate how the dynamic panel bootstrap-corrected fixed-effects estimator is useful in estimating higher-order panel data models and accounting for challenges such as omitted-variable bias, convergence, cross-section dependence and heterogeneous effects.
- We apply the imputation technique, *panelhetero*, and *xtbcfe* algorithms to examine the nexus between ambient air pollution and health outcomes.

Specifications table

Subject Area:	Environmental Science, Health Economics, Econometrics
More specific subject area:	*Energy, Environmental and Health econometrics*
Method name:	*Dynamic panel bootstrap-corrected fixed-effects and Heterogeneous dynamics*
Name and reference of original method:	Everaert, Gerdie, and Lorenzo Pozzi. "Bootstrap-based bias correction for dynamic panels." *Journal of Economic Dynamics and Control* 31.4 (2007): 1160-1184.
	Okui, R., Yanagi, T., 2019. Panel data analysis with heterogeneous dynamics. *Journal of Econometrics.* 212, 451-475.
Resource availability:	Data used in this study were extracted from the environmental risk and health database of the Organization for Economic Co-operation Development [1].

Method details

Data pre-processing

Pre-processing techniques of unevenly spaced data and negative value are critical in time series and panel data estimation techniques. For example, first and second generational panel unit root tests namely Im-Pesaran-Shin (IPS) [2], Levin-Lin-Chu (LLC), Harris-Trazvalis (HT), Breitung [3], cross-sectionally augmented IPS (CIPS) and cross-section augmented Dickey-Fuller (CADF) require either a strongly balanced panel or no missing values (no gaps). While there are several techniques available, we recommend a unique data imputation algorithm in Orange software that is advantageous in retaining the pattern of original data with little or no changes in centring, distribution, dispersion, minimum and maximum values of the raw data. The imputation method is valid in all missing mechanisms ranging from missing not at random (MNAR), missing at random (MAR), and missing completely at random (MCAR) [4]. The Orange data mining software is an open-source project which is freely available to download[1] [5]. It utilizes a drag and drop user interface, which is easy for input-output type of estimation, hence, beneficial to researchers with limited coding skills.

Once the software has been downloaded and installed, follow the steps depicted in Fig. 1. Drag and drop the file to upload the dataset. Select the type of data: numeric, categorical, text, datetime; and roles of the data: feature, target, meta and skip. Drag and drop the feature statistics and connect the link back to the file module. Feature statistics helps to examine the characteristics of the uploaded data. It produces descriptive statistics namely distribution, centre, dispersion, minimum, maximum, and total number of missing inputs. At the impute module, select the desired impute function based on data characteristics. The impute function includes don't impute (skip imputation of a particular variable), average/most frequent, as a distinct value, model-based imputer (using simple tree), random

---

[1] Freely available to download here: https://buff.ly/2BQucdJ.

Feature Statistics –
Before Imputation

Feature Statistics –
After Imputation

Data Table to observe
the imputed variables

Data

Data

Select the desired
imputation function

Data

File

Data

Impute

Data Table

Upload your data and specify
the type of data and role

Save your new dataset
in available formats
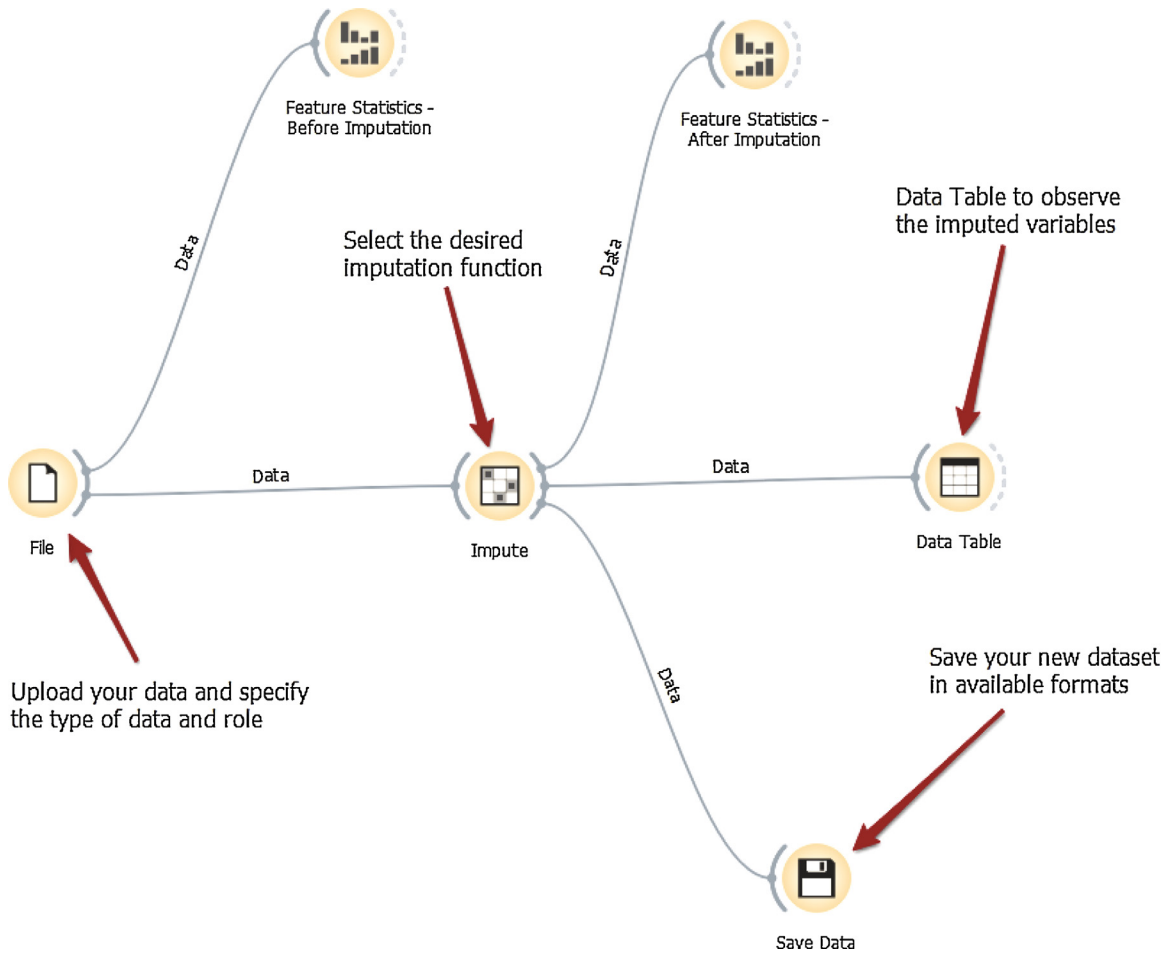
Data

Save Data

Fig. 1. Procedure for missing data imputation in Orange software.

Table 1
Characteristics of data during pre-imputation and post-imputation.

Data	Name	Center	Dispersion	Min.	Max.	Missing
Raw Data	PM 2.5	29.93	0.61	5.76	103.16	2608 (57%)
Imputed Data	PM 2.5	29.93	0.61	5.76	103.16	0(0%)
Raw Data	TOT_DALY	9.26	0.51	1.72	39.16	0(0%)
Imputed Data	TOT_DALY	9.26	0.51	1.72	39.16	0(0%)
Raw Data	TOT_MOR	309.21	0.63	40.02	1257.4	0(0%)
Imputed Data	TOT_MOR	309.21	0.63	40.02	1257.4	0(0%)
Raw Data	TOT_MOR_V	13959	5.17	4.83	1029848	0(0%)
Imputed Data	TOT_MOR_V	13959	5.17	4.83	1029848	0(0%)
Raw Data	TOT_SC_V	17281.76	4.2	1.47	1580136	268(4%)
Imputed Data	TOT_SC_V	17836.43	4.35	1.47	1580136	0(0%)

Notes: Raw Data is the original dataset with missing values whereas Imputed Data denotes the output of the imputation algorithm. Legend: TOT_MOR is the total mortality from exposure to outdoor $PM_{2.5}$ and ozone, TOT_MOR_V denotes premature deaths from exposure to outdoor $PM_{2.5}$ and ozone, TOT_DALY is the total Disability-Adjusted Life Year from exposure to outdoor $PM_{2.5}$ and ozone, TOT_SC_V is the total welfare cost of premature deaths from exposure to outdoor $PM_{2.5}$ and ozone and PM 2.5 is the exposure to ambient particulate matter. The Table presented is reproduced from Owusu and Sarkodie [6].

values, remove instances with unknown values and custom value. Once the impute function has been selected and applied, then the feature statistics can be observed and presented in the data Table. The desired imputed dataset can be saved using the save data module. We test the procedure on ambient air pollution and health outcomes presented in Table 1. It can be observed that the characteristics of both raw and imputed data are relatively equal.

Researchers often encounter negative values, a challenge prior to the application of data transformation. While many scholars tend to use absolute values of negative inputs, such procedure is fundamentally wrong and distorts the pattern of the original data. To maintain the pattern and characteristics of negative dataset, the normalization technique presented in Sarkodie, Adams [7] can be applied manually with Microsoft Excel or using OriginPro software. In OriginPro, the technique can be applied by selecting the column of the negative data. The user goes to the Analysis module → Mathematics → Normalize columns → Open dialog and chooses the desired normalization method.

Robust and consistent panel estimation

Most often, data are log-transformed to offer variables a constant variance and control for potential heteroskedasticity. To ensure robust and consistent panel estimates, we begin by testing for panel unit root tests.

Step 1: Without any assumption, we apply two of the first generational unit root tests namely Breitung and IPS. The tests are important to deal with highly persistent series that may influence the model estimation and produce spurious and biased statistical inferences. Thus, we investigate the stationarity properties of the panel data under the null hypothesis of unit root. We use STATA version 16 software, declare the data as panel and run the test in both level and first difference. For example, we test the panel unit root of mortality as: *xtunitroot breitung lnTOT_MOR* and *xtunitroot breitung d.lnTOT_MOR*. Other options like trend, lags, demean, kernel, and noconstant can be accounted for. Sample results of the first generational panel unit root tests are presented in Table 2.

Step 2: Panel data estimation are often affected by unobserved and global common shocks, of which failure to account for it renders the model spurious and misspecified. To account for this challenge, we adopt the test of cross-section dependence (CD) expounded in Pesaran [8], Pesaran [9]. Importantly, the CD test can be applied to unevenly spaced and balanced panel to assess the average correlation between sampled countries assuming a standard normal distribution. The CD test technique is based on the null hypothesis of either weak cross-section dependence or strict cross-section independence [8,9]. For example, we run the pre-estimation CD test algorithm in STATA using: *xtcd lnTOT_MOR*. We present the results of the estimated CD test in Table 3. It can be observed in

Table 2

First generational panel unit root tests.

Variable	Breitung		IPS	
	Level	First Difference	Level	First Difference
PM 2.5	-39.132***	-37.890***	-36.528***	-50.915***
TOT_DALY	21.974	-25.140***	4.966	-30.589***
TOT_MOR	21.255	-27.366***	8.697	-31.327***
TOT_MOR_V	32.947	-28.873***	22.250	-32.250***
TOT_SC_V	30.867	-25.738***	27.921	-31.840***
TOT_SC_V_TOT_DALY	22.479	-27.153***	18.659	-33.700***

Notes: *** denotes rejection of the null hypothesis of unit root at 1% significance level. The Table presented is reproduced from Owusu and Sarkodie [6].

Table 3

Test for cross-sectional dependence.

Variable	CD-test	p-value	average joint T	mean ρ	mean abs(ρ)
TOT_MOR	5.751	0.000	28	0.010	0.650
TOT_MOR_V	231.128	0.000	28	0.32	0.74
PM 2.5	2.425	0.015	28	0.000	0.220
TOT_DALY	32.001	0.000	28	0.040	0.650
TOT_SC_V	405.366	0.000	28	0.560	0.670
TOT_SC_V_TOT_DALY	184.500	0.000	28	0.250	0.590

Notes: Under the null hypothesis of cross-section independence, CD ~ N(0,1) P-values close to zero indicate data are correlated across panel groups. The Table presented is reproduced from Owusu and Sarkodie [6].

Table 3 that the null hypothesis is rejected — providing strong evidence of cross-section dependence across countries.

Step 3: We investigate the prospect of heterogeneity across 195 countries due to varying income groups and differences in economic structure. We utilize four estimation techniques namely modified Wald (MWALD) statistics in a fixed-effect regression, empirical CDF, moments, and kernel density estimation to examine heterogeneous effects. We apply the MWALD test in two stages: first, run *xtreg lnTOT_MOR lnPM25, fe* and second, run *xttest3*. The MWALD test assumes normality of errors under the null hypothesis of homoskedasticity [10].

We validate MWALD test using modern panel techniques for heterogeneous dynamics (*panelhetero*) expounded in Okui and Yanagi [11]. It is important to note that the *panelhetero* algorithm requires a strongly balanced panel. The panel heterogeneous dynamics (*panelhetero*) is a model-free estimation technique that controls for misspecification bias with three methods namely naïve estimation *("naive")* — without bias correction and two bias correction procedures based on split-panel jackknife — namely half-panel *("hpj")* and third-order jackknife *("toj")* [11]. Thus, the panel heterogeneous dynamics approach estimates statistics such as mean, autocovariance and autocorrelation for each country and calculates empirical cumulative distribution function (CDF), moments and kernel density. We apply the model-free approach using *panelhetero* algorithm via these procedures: first, we estimate moments using *phmoment lnTOT_MOR, method("hpj") boot(200) acov_order(0) acor_order(1)*; second, we estimate empirical CDF using *phecdf lnTOT_MOR, method("hpj") acov_order(0) acor_order(1)*; and third, we estimate kernel density using *phkd lnPM25, method("hpj") acov_order(0) acor_order(1)*. The options in the *panelhetero* algorithm can be modified to suit the user's estimation needs. The moments compute 9 statistical attributes when heterogeneity is observed across countries. The sample estimation for mortality rates using moments is presented in Table 4.

The sample estimation for mortality rates and ambient air pollution using empirical CDF technique are depicted in Figs. 2–4.

Similarly, the kernel density estimation for mortality rates and ambient air pollution are presented in Figs. 5–7. We observe from all the three panel heterogeneous dynamics techniques that the estimates are with the 95% confidence interval, hence, confirm the presence of heterogeneous effects across countries.

Table 4
Moments estimation for Mortality rates (*lnTOT_MOR*).

Parameters	Estimate	S.E	Low*	High**
Mean of Mean	5.539	0.042	5.470	5.621
Mean of Autocovariance	0.040	0.004	0.031	0.046
Mean of Autocorrelation	1.036	0.014	1.010	1.062
Variance of Mean	0.373	0.031	0.319	0.435
Variance of Autocovariance	0.002	0.000	0.000	0.002
Variance of Autocorrelation	-0.011	0.005	-0.020	-0.001
Correlation between Mean and Autocovariance	-0.278	0.095	-0.469	-0.088
Correlation between Mean and Autocorrelation	-0.361	0.131	-0.599	-0.123
Correlation between Autocovariance and Autocorrelation	0.447	0.062	0.336	0.573

Notes: *,** denotes 95 % Confidence Intervals for Moments based on bootstrapping across cross-sectional units; S.E represents Standard Errors of the estimates based on bootstrapping across cross-sectional units.
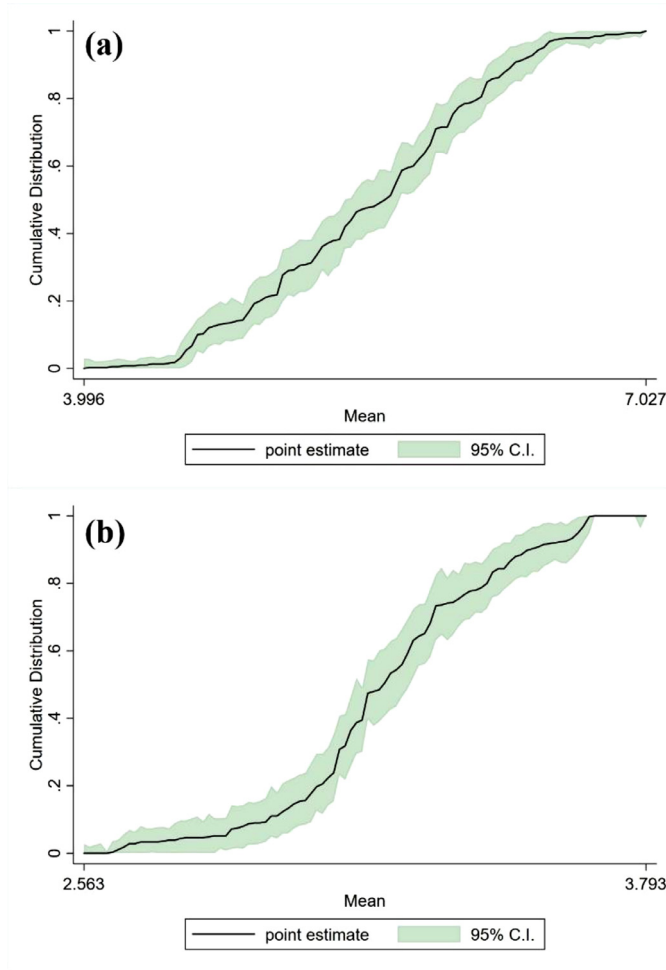

Fig. 2. Empirical CDF Estimation for Mean (a) Mortality (b) Ambient Air Pollution.
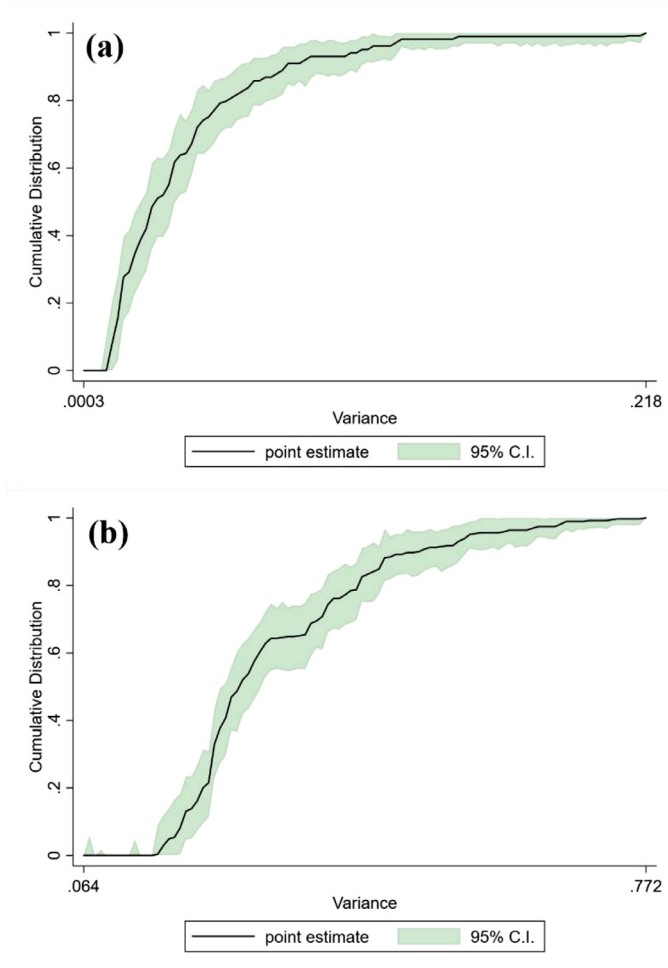
Fig. 3. Empirical CDF Estimation for Variance (a) Mortality (b) Ambient Air Pollution.

Step 4: With the presence of strong evidence of cross-section dependence, we re-examine the panel unit root using second generational panel unit root tests useful in heterogeneous panel with strong correlation across countries. We utilize CIPS and CADF panel unit root tests under the null hypothesis of homogeneous non-stationary for CIPS [12] and non-stationary series for CADF [2]. To test for CIPS, we run *xtcips lnTOT_MOR, maxlags(2) bglags(1)* and *pescadf lnTOT_MOR, lags(1)* at level for CADF test. The empirical results of the second generational panel unit root tests are presented in Table 5, which validate the results of first generational panel unit root tests.

Step 5: We test for panel cointegration using Westerlund estimation technique as *xtcointtest westerlund lnTOT_MOR lnPM25*.

Step 6: Contrary to the numerous panel estimation techniques available and used in the extant literature [13–16], we utilize the novel dynamic panel bootstrap-corrected fixed-effects estimator to construct mortality-DALYs-PM$_{2.5}$ models with lagged dependent explanatory variables. In contrast to traditional panel techniques that require only large time dimension (T) for model estimations to be asymptomatically binding, the bootstrap-corrected fixed-effects — least squares dummy variable estimator corrects the small T bias in panel dynamic models [17,18] using a simplified algorithm
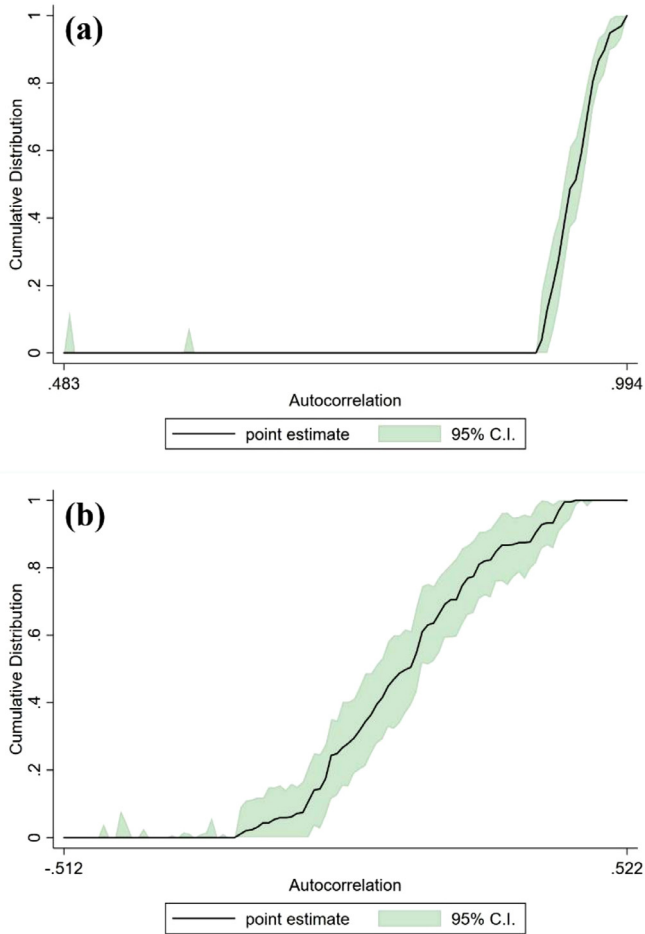
Fig. 4. Empirical CDF Estimation for Autocorrelation of order 1 (a) Mortality (b) Ambient Air Pollution.

Table 5
Second generational panel unit root tests.

Variable	CIPS[a]		PESCADF[b]	
	Level	First Difference	Level	First Difference
PM 2.5	-4.533***	-6.071***	-15.390***	-34.090***
TOT_DALY	-1.551	-3.366***	1.872	-11.245***
TOT_MOR	-1.387	-3.420***	4.752	-13.206***
TOT_MOR_V	-1.508	3.243***	2.228	-2.741***
TOT_SC_V	-1.112	-3.915***	27.921	-4.504***
TOT_SC_V_TOT_DALY	-1.861	-4.236***	1.336	-7.357***

Notes: [a]H_0 (homogeneous non-stationary): bi = 0 for all I whereas [b] the null hypothesis assumes all series are non-stationary in a heterogeneous panel with cross-sectional dependence. The Table presented is reproduced from Owusu and Sarkodie [6].
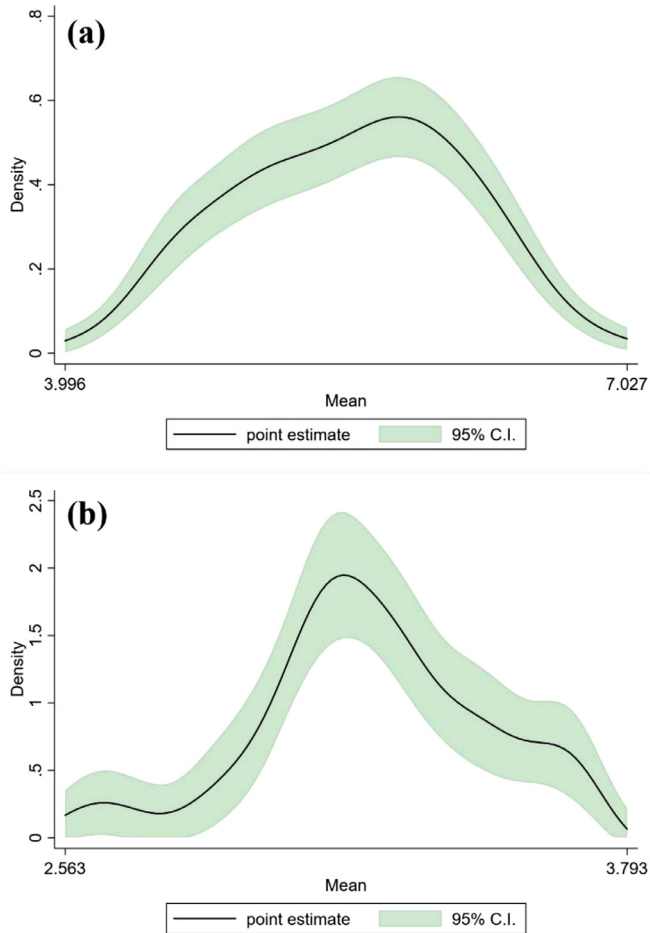
Fig. 5. Kernel Density Estimation for Mean (a) Mortality (b) Ambient Air Pollution.

introduced in Everaert and Pozzi [19]. Thus, the bootstrap-corrected fixed-effects estimator is useful in estimating higher-order panel data models that contradict the standard error structure, a situation encountered in this study. Using the suitable resampling option in the dynamic panel estimator, challenges such as, *inter alia*, omitted-variable bias, convergence, cross-section dependence and heteroskedasticity that undermine the analytical error correction procedures are controlled.

For briefness, the generic representation of dynamic panel estimation models based on bootstrap-corrected fixed-effects is expressed as [6,20]:

$$y_{i,t} = \gamma * y_{i,t-1} + \beta * x_{i,t} + \mu_i + \varepsilon_{i,t}, \tag{1}$$

Where $y_{i,t}$ represents dependent variables, $x_{i,t}$ denotes strongly exogenous independent variables, γ denotes the autoregressive coefficient of lagged dependent variable, β represents estimated vector coefficients of regressors, μ_i denotes unobserved heterogeneity or uncorrelated and exogenous country-specific fixed-effects with zero mean and greater than zero variance, and $\varepsilon_{i,t}$ denotes unobserved and uncorrelated white noise across countries $i = 1, \ldots, N$ and time dimension $t = 2, \ldots, T$. γ is assumed to be less than 1 ($\gamma < 1$) to achieve a dynamic stable relationship between $y_{i,t}$ and $x_{i,t}$.
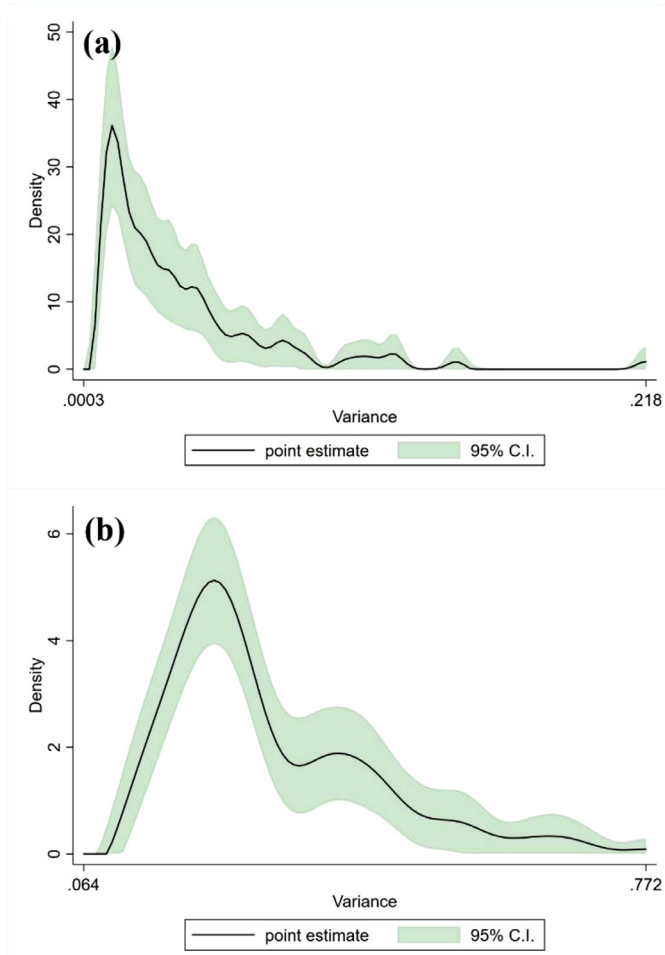
Fig. 6. Kernel Density Estimation for Variance (a) Mortality (b) Ambient Air Pollution.

To control for heterogeneous effects, we modify the model specification to include resampling of error terms using randomized temporal heteroskedasticity scheme with analytical heterogeneous initialization. This implies that the algorithm resamples the entire time period, with subsequent resampling of error terms within the specified time periods ($t=1, ..., 28$). Our sampling method used has characteristics of multivariate normal distribution and initial conditions consisting of cross-sectional specific means and variance-covariance matrices [20]. We run the bootstrap-corrected fixed-effects estimator (*xtbcfe*) using *xtbcfe lnTOT_MOR lnPM25, bciters(2500) res(thet_r) ini(ahe) lags(1) infer(inf_ci) infit(1000) te*. The estimation procedure, resampling, inferences, scheme and initialization and post-estimation can be modified to suit the needs of the user. The sample parameter estimates of ambient air pollution and health outcomes using dynamic panel bootstrap-corrected fixed-effects are presented in Table 6.

Model validation

To make unbiased statistical inferences while preserving the dynamic panel structure of the estimated models, we utilized the nonparametric bootstrap option of the simulation to resample the original data series and subsequently apply the bootstrapping bias-correction to the estimated
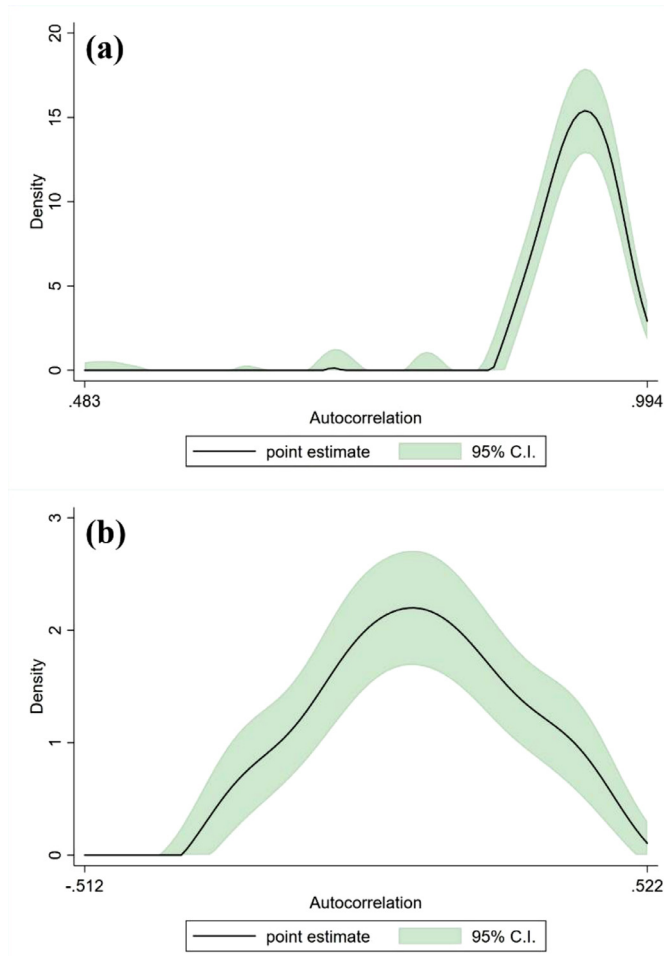
Fig. 7. Kernel Density Estimation for Autocorrelation of order 1 (a) Mortality (b) Ambient Air Pollution.

Table 6
Parameter estimates of ambient air pollution and health outcomes using dynamic panel bootstrap-corrected fixed-effects.

Estimation	Mortality	Premature	DALYs	Welfare Cost	Mortality	Mortality
$|\gamma| < 1$	0.310***	0.274***	0.315***	0.211**	0.841***	0.831***
	[0.036]	[0.034]	[0.030]	[0.085]	[0.028]	[0.028]
DALYs	—	—	—	—	0.147***	0.108***
					[0.026]	[0.024]
Welfare Cost	—	—	—	—	0.004***	-0.012***
					[0.001]	[0.004]
Welfare Cost $\times$ DALYs	—	—	—	—	—	0.008***
						[0.002]
PM$_{2.5}$	0.001***	0.001***	0.001***	0.048***	0.005***	0.005***
	[0.000]	[0.000]	[0.000]	[0.014]	[0.001]	[0.001]
1992	—	—	—	0.073	0.004**	0.004**
				[0.061]	[0.002]	[0.002]

(*continued on next page*)

Table 6 (continued)

Estimation	Mortality	Premature	DALYs	Welfare Cost	Mortality	Mortality
1993	0.000	0.000	0.000	-0.012	0.007***	0.007***
	[0.002]	[0.002]	[0.002]	[0.065]	[0.002]	[0.002]
1994	-0.008***	-0.009***	-0.007***	0.003	0.000	0.001
	[0.002]	[0.002]	[0.002]	[0.075]	[0.002]	[0.002]
1995	-0.010***	-0.011***	-0.010***	0.078	-0.002	-0.001
	[0.002]	[0.002]	[0.002]	[0.062]	[0.002]	[0.002]
1996	-0.001	-0.002	-0.000	0.169**	0.005**	0.006***
	[0.002]	[0.002]	[0.002]	[0.079]	[0.002]	[0.002]
1997	-0.006***	-0.006***	-0.003	0.066	0.005*	0.006**
	[0.002]	[0.002]	[0.002]	[0.067]	[0.003]	[0.003]
1998	-0.008***	-0.008***	-0.006**	0.171**	0.003	0.004*
	[0.002]	[0.002]	[0.003]	[0.073]	[0.002]	[0.002]
1999	-0.012***	-0.013***	-0.010***	0.154*	0.000	0.001
	[0.002]	[0.002]	[0.002]	[0.090]	[0.003]	[0.003]
2000	-0.017***	-0.018***	-0.015***	0.148**	-0.005*	-0.004
	[0.002]	[0.002]	[0.003]	[0.066]	[0.003]	[0.003]
2001	-0.005**	-0.006***	-0.003	0.279***	0.004	0.005*
	[0.002]	[0.002]	[0.002]	[0.083]	[0.003]	[0.003]
2002	-0.005**	-0.006***	-0.005**	0.225**	0.008***	0.009***
	[0.002]	[0.002]	[0.002]	[0.106]	[0.003]	[0.003]
2003	-0.008***	-0.009***	-0.007***	0.211***	0.007**	0.008***
	[0.002]	[0.002]	[0.002]	[0.070]	[0.003]	[0.003]
2004	-0.016***	-0.016***	-0.013***	0.303***	0.001	0.002
	[0.002]	[0.002]	[0.002]	[0.105]	[0.003]	[0.003]
2005	-0.011***	-0.012***	-0.009***	0.469***	0.002	0.004
	[0.002]	[0.002]	[0.002]	[0.099]	[0.003]	[0.003]
2006	-0.004*	-0.005**	-0.000	0.445***	0.009***	0.011***
	[0.002]	[0.002]	[0.002]	[0.105]	[0.003]	[0.003]
2007	-0.006***	-0.007***	-0.003	0.448***	0.010***	0.012***
	[0.002]	[0.002]	[0.002]	[0.099]	[0.003]	[0.003]
2008	-0.009***	-0.010***	-0.007***	0.510***	0.008**	0.009***
	[0.002]	[0.002]	[0.002]	[0.128]	[0.004]	[0.003]
2009	-0.010***	-0.011***	-0.008***	0.422***	0.007**	0.009**
	[0.002]	[0.002]	[0.004]	[0.124]	[0.004]	[0.004]
2010	-0.012***	-0.013***	-0.011***	0.534***	0.006*	0.008**
	[0.002]	[0.003]	[0.002]	[0.097]	[0.004]	[0.004]
2011	0.005*	0.003	-0.006**	0.533***	0.021***	0.022***
	[0.003]	[0.003]	[0.003]	[0.111]	[0.004]	[0.004]
2012	-0.026***	-0.027***	-0.025***	0.521***	0.000	0.002
	[0.003]	[0.003]	[0.004]	[0.138]	[0.005]	[0.005]
2013	-0.015***	-0.017***	-0.014***	0.568***	0.003	0.005
	[0.003]	[0.003]	[0.003]	[0.104]	[0.004]	[0.004]
2014	-0.018***	-0.021***	-0.017***	0.650***	0.002	0.004***
	[0.003]	[0.003]	[0.003]	[0.140]	[0.004]	[0.004]
2015	0.031***	0.028***	0.031***	0.619***	0.044***	0.046
	[0.006]	[0.006]	[0.006]	[0.107]	[0.006]	[0.006]
2016	-0.041***	-0.042***	-0.041***	0.747***	-0.003	0.000***
	[0.004]	[0.004]	[0.004]	[0.125]	[0.004]	[0.004]
2017	0.012***	0.008***	0.012***	0.652***	0.028***	0.030
	[0.003]	[0.002]	[0.002]	[0.111]	[0.004]	[0.004]
Convergence	YES	YES	YES	YES	YES	YES
MWALD test (Prob>chi²)‡	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***

Notes: [.] denotes Bootstrapped standard errors, Bootstrap 95% (percentile-based) confidence intervals and Inference performed with non-parametric bootstrap; γ represents the lagged dependent variable; ***,**,* denote statistical significance at 1, 5, and 10% levels.
[a,b,c] denote Mortality ~ f(Ambient air pollution), Mortality ~ f(DALYs, welfare cost and ambient air pollution) and Mortality ~ f(DALYs, welfare cost, ambient air pollution and interaction between DALYs and welfare cost). ‡denotes the modified Wald test used as a post-estimation technique to examine groupwise heteroskedasticity under the null hypothesis, H_0: $\sigma_{(i)}^2 = \sigma^2$ for all i. **Legend:** DALY is the average total Disability-Adjusted Life Years from exposure to $PM_{2.5}$ and ozone, MWALD means the modified Wald statistics, and $Prob>chi^2$ is the probability of Chi-squared test. The Table presented is reproduced from Owusu and Sarkodie [6].
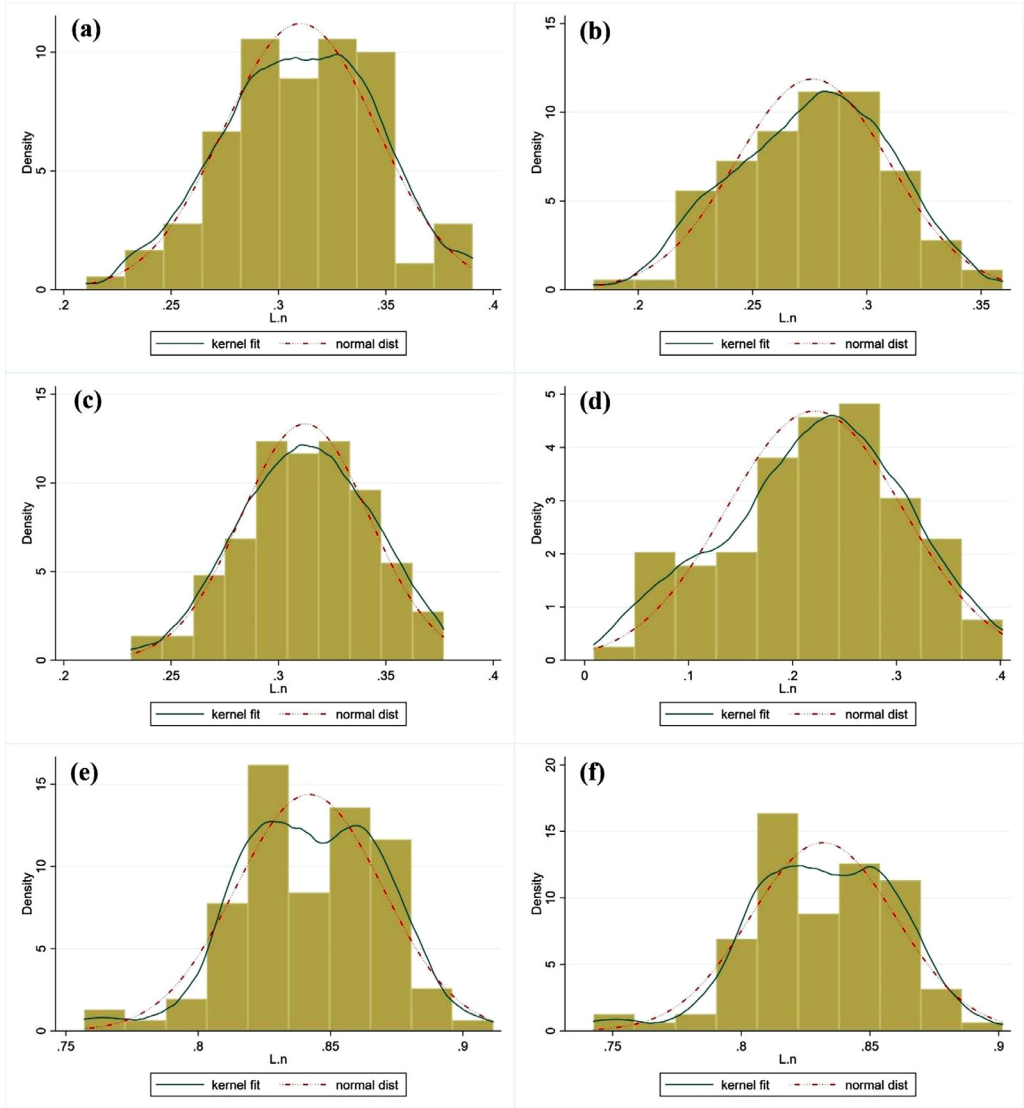
Fig. 8. Post-estimation bootstrap-stimulated distribution of autoregressive (AR) coefficients and their sum for the: (a) relationship between mortality and $PM_{2.5}$ (b) relationship between premature deaths and $PM_{2.5}$ (c) relationship between DALYs and $PM_{2.5}$ (d) relationship between welfare cost of premature deaths from exposure to $PM_{2.5}$ and ozone and $PM_{2.5}$ (e) relationship between mortality versus $PM_{2.5}$, DALYs, and the welfare cost of premature deaths from exposure to $PM_{2.5}$ and ozone (f) relationship between mortality versus $PM_{2.5}$, DALYs, the welfare cost of premature deaths from exposure to $PM_{2.5}$ and ozone, and the interactive effective of DALYs and welfare cost of premature deaths from exposure to $PM_{2.5}$ and ozone. Figure presented is reproduced from Owusu and Sarkodie [6].

fixed-effects of each constructed samples [20]. We validate the estimated parameters using the post-estimation bootstrap-stimulated distribution of autoregressive (AR) coefficients and their sum. This can be estimated by modifying the model specification as: *xtbcfe lnTOT_MOR lnPM25, bciters(2500) res(thet_r) ini(ahe) lags(1) infer(inf_ci) infit(1000) te dist(all)*. Sample post-estimation bootstrap-stimulated distribution of autoregressive (AR) coefficients is presented in Fig. 8.

Conclusion

With regards to the growing complexities of panel data modelling, we provide a step-by-step process and guideline for estimating dynamic panel models using the novel dynamic panel bootstrap-corrected fixed-effects estimator (*xtbcfe*). It is important to note that *xtbcfe* algorithm cannot compute statistical inferences without achieving convergence, hence, affect the robustness and consistency of estimated parameters. Three key components namely number of iterations (*infit*), convergence criterion (*crit*) and number of samples for bootstrapping (*bciters*) affect the computation time of *xtbcfe* algorithm. Additionally, we provided guidelines on data imputation of unevenly spaced dataset and steps to apply the normalization technique to negative inputs before data transformation. We demonstrated how to estimate a model-free panel heterogeneous dynamics (*panelhetero*) that control for misspecification using empirical CDF, moments, and kernel density estimation. Finally, we applied the guidelines to examine the empirical data on ambient air pollution and health outcomes. The estimated parameters confirmed ambient air pollution attributed mortality, Disability-Adjusted Life Years, and welfare cost. We note that the recommended pre-processing and panel estimation techniques can be applied in several disciplines such as, *inter alia*, social sciences, energy, health, environmental and resource economics.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] OECD, Environmental Risks and Health, 2018 [December 15th, 2019] Available from: https://stats.oecd.org/.
[2] M.H. Pesaran, K.S. Im, Y. Shin, Testing for unit roots in heterogeneous panels, J. Econom. 115 (1) (2003) 53–74.
[3] J. Breitung, The Local Power of Some Unit Root Tests for Panel Data, 1999 Discussion Papers, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes.
[4] X. Liu, Chapter 14 - Methods for handling missing data, in: X. Liu (Ed.), Methods and Applications of Longitudinal Data Analysis, Academic Press, Oxford, 2016, pp. 441–473.
[5] J. Demšar, et al., Orange: data mining toolbox in Python, J. Mach. Learn. Res. 14 (1) (2013) 2349–2353.
[6] P.A. Owusu, S.A. Sarkodie, Global estimation of mortality, disability-adjusted life years and welfare cost from exposure to ambient air pollution, Sci. Total Environ. 742 (2020) 140636.
[7] S.A. Sarkodie, S. Adams, T. Leirvik, Foreign direct investment and renewable energy in climate change mitigation: Does governance matter? J. Clean. Prod. (2020) 121262.
[8] M.H. Pesaran, General Diagnostic Tests for Cross section Dependence in Panels, 2004.
[9] M.H. Pesaran, Testing weak cross-sectional dependence in large panels, Econom. Rev. 34 (6-10) (2015) 1089–1117.
[10] W.H. Greene, Econometric Analysis, Prentice-Hall, New York, 2000.
[11] R. Okui, T. Yanagi, Panel data analysis with heterogeneous dynamics, J. Econom. 212 (2) (2019) 451–475.
[12] M.H. Pesaran, A simple panel unit root test in the presence of cross-section dependence, J. Appl. Econom. 22 (2) (2007) 265–312.
[13] A.A. Alola, F.V. Bekun, S.A. Sarkodie, Dynamic impact of trade policy, economic growth, fertility rate, renewable and non-renewable energy consumption on ecological footprint in Europe, Sci.Total Environ. 685 (2019) 702–709.
[14] F.V. Bekun, A.A. Alola, S.A. Sarkodie, Toward a sustainable environment: Nexus between CO2 emissions, resource rent, renewable and nonrenewable energy in 16-EU countries, Sci. Total Environ. 657 (2019) 1023–1029.
[15] G.N. Ike, O. Usman, S.A. Sarkodie, Testing the role of oil production in the environmental Kuznets curve of oil producing countries: New insights from Method of Moments Quantile Regression, Sci. Total Environ. 711 (2020) 135208.
[16] S.A. Sarkodie, P.A. Owusu, T. Leirvik, Global effect of urban sprawl, industrialization, trade and economic development on carbon dioxide emissions, Environ. Res. Lett. 15 (3) (2020) 034049.
[17] S. Nickell, Biases in dynamic models with fixed effects, Econom. J. Econom. Soc. (1981) 1417–1426.
[18] J.F. Kiviet, On bias, inconsistency, and efficiency of various estimators in dynamic panel data models, J. Econom. 68 (1) (1995) 53–78.
[19] G. Everaert, L. Pozzi, Bootstrap-based bias correction for dynamic panels, J. Econ. Dyn. Control 31 (4) (2007) 1160–1184.
[20] I. De Vos, G. Everaert, I. Ruyssen, Bootstrap-based bias correction and inference for dynamic panels with fixed effects, Stata J. 15 (4) (2015) 986–1018.