The University Of Sheffield.
School Of Economics

Re-evaluating China's New Energy Demonstration Cities with two-way fixed effects: A comment

You Wu and Andrew Burlinson

Sheffield Economic Research Paper Series

SERPS no. 2026002

ISSN 1749-8368

05 March 2026

Re-evaluating China's New Energy Demonstration Cities with two-way fixed effects: A comment

You Wu and Andrew Burlinson*

School of Economics, University of Sheffield, Sheffield S10 2TU, United Kingdom.

Email: ywu306@sheffield.ac.uk

This version: 05 March 2026

Please do not cite or distribute without the author's permission.

Abstract

China's New Energy Demonstration City (NEDC) policy aims to address energy and environmental challenges. While two-way fixed effects (TWFE) regression is often used to evaluate such policies within difference-in-differences frameworks, it can lead to incorrect conclusions if unit and time effects are mis-specified. Re-evaluating the impact of the NEDC policy on energy poverty (EP), as studied in Ma et al. (2025), by correctly specifying unit-specific *and* time-specific effects, we find that: i) the parallel trends assumption fails, and ii) the previously reported impact on EP becomes economically and statistically insignificant. Whether NEDCs reduce EP therefore remains an open question.

Keywords: Policy evaluation; Two-way fixed effects regression; Difference-in-differences; Energy Poverty

JEL codes: C12; C23; Q41; Q48

---

* We thank Subhasish M. Chowdhury, Karl Taylor, and Peter Wright for their thoughtful comments. Any remaining errors are our own. You Wu acknowledges the support of CSC-Sheffield Joint PhD Scholarship programme.

1. Introduction

Concerns related to energy consumption, including but not limited to climate change, poverty and inequality, have strengthened the need for credible policy solutions (see, e.g., Wu et al., 2024; Nepal et al., 2025). China's New Energy Demonstration City (NEDC) policy has been touted as one such solution, in particular in improving access and affordability of modern energy for basic needs – often referred to as Energy Poverty (EP). However, its effectiveness remains uncertain. This is not uncommon, as uncertainty regarding energy policy efficacy in tackling both environmental and equity concerns is rife, largely due to the dearth of rigorous (causal) policy evaluation (Zhang et al., 2024). The tide is turning with the combined rise in the use of difference-in-differences (DiD) and two-way fixed effects (TWFE) regression methods in energy policy evaluation (see, e.g., Clay et al., 2023; Hoffmann et al., 2025).[1] We contribute to the literature by showing the potential consequences for policy evaluation when *two-way* fixed effects (FEs) are not correctly specified.

A recent study by Ma, Wan, and Zhou (2025; MWZ hereafter) serves as our example. Despite using a household survey of individuals in China, the authors specify their FEs at the regional (*provincial*) and time (*year*) levels, and find that the NEDC policy has a statistically significant effect on mitigating EP. However, the authors do not justify why *unit* (individual) effects were overlooked in favour of regional effects. The authors further argue that the parallel trends assumption holds, i.e., before the treatment is introduced, the average change of the outcome variable would have remained the same in the treatment and control groups. In the traditional 2x2 (homogenous treatment) setting, if parallel trends hold, the treatment coefficient in a TWFE regression is equal to the DiD estimator obtained directly from sample means, i.e., TWFE estimators are unbiased for the average treatment effect (ATE) (Callaway and Sant'Anna, 2021). Indeed, the TWFE estimator of an ATE using within transformation and time effects is identical to a pooled ordinary least square (POLS) that controls for both unit-specific effects (time-invariant) and time-specific FEs. Overlooking the former set of unit-

[1] While TWFE is a popular approach of policy evaluation beyond energy policy, see for e.g., Ren and West (2023), Kang et al. (2024), and Brehm et al. (2025), it has also been criticised when applied to *heterogeneous* treatment effects (de Chaisemartin and D'Haultfœuille, 2023). For the purpose of this study, we assume treatments are homogenous.

specific effects however simply reverts to a POLS model, e.g., solely controlling for regional and time effects in MWZ, which could lead to incorrect policy conclusions. [2]

We demonstrate this issue by revisiting MWZ and correctly specifying the TWFE regression. This allows us to assess the authors' baseline results, including parallel trends, in order to compare them with estimates that account for *both* individual unit-specific and time-specific FEs. This leads to two crucial findings: i) based on both MWZ's main EP index and a central sub-component of the index, the parallel trends assumption appears to fail; and ii) regardless of the potential violation of parallel trends, upon controlling for unit-specific FEs, the NEDC policy's impact on EP turns out to be economically and statistically insignificant.

Our findings contribute to the literature by highlighting a case in which correctly defining *two-way* (unit-specific and time-specific) FEs is necessary, in the presence of omitted variable bias, when evaluating potential outcomes of energy and environmental policy. This further allows us to show the potential uncertainty in the NEDCs ability to mitigate EP, as well as the need for further evidence and avenues for research, for policymakers who may be considering expanding (replicating) the NEDC within (outside) of China.

2. Empirical analysis

2.1. Data

MWZ measured EP using an index constructed by five dimensions: cooking fuels, electricity, affordability, appliances, and energy expenditure; assigning a weight of 0.6 to cooking and weight of 0.1 to all other dimensions.[3]

We use this measure and other data on covariates exactly as it was published by MWZ. The unbalanced panel data covers 126 cities with 28 treated cities (NEDCs), from 2011 to 2021 (six waves), and is sourced from the China Family Panel Studies (CFPS). The total observations are 67906, with 10765 in the treatment group and 57141 in the control.

---

[2] This could potentially occur when using "reghdfe" Stata package, as used by MWZ, while overlooking the fact that unit-effects must be specified alongside time and other higher dimensional effects if needed. Unlike the standard "xtreg, fe" package, which automatically accounts for unit-specific FEs.

[3] MWZ use different weights as a robustness check, while their choice of applying a greater weight to cooking in the main findings may have driven some of the statistical significance overall, this is not the focus of our findings. Our findings herein are the same either way.

2.2. Econometric methodology

While MWZ report using TWFE, they adopt an atypical approach by defining the effects at the regional (*provincial*) and time (year) levels, instead of the standard *unit* (individual) and time levels as discussed above. Indeed, their model can be represented as a POLS regression as follows:

$$EP_{ict} = \beta_{POLS}DiD + X_{ict}\gamma + \alpha_p + \lambda_t + \varepsilon_{ict} \quad (1)$$

where EP_{ict} denotes the EP index for individual i residing in city c in year t. DiD is the interaction term for the policy exposure ($treatment \times post_{2014}$), which equals 1 if the household lives in an NEDC in 2014 onwards, and 0 otherwise. X_{ict} represents a vector of covariates, specifically: household size, education attainment, income level, age, workplace type, health status, per capita GDP, urban-rural location, and government interventions (i.e., the share of fiscal expenditure in GDP).[4] α_p is the *provincial*-level effects, λ_t is the time effects. And ε_{ict} is a composite error term which subsumes the unobserved unit-specific time-invariant heterogeneity (a_i) and the idiosyncratic factors (u_{ict}) affecting EP_{ict}. β_{POLS} is the coefficient of interest, and γ are estimated coefficients for the controls.

According to Imai and Kim (2021), among others, in order to rigorously evaluate policy, TWFE should be defined at the unit-specific (individual) level as follows:

$$EP_{ict} = \beta_{TWFE}DiD + X_{ict}\gamma + \alpha_p + \lambda_t + a_i + u_{ict} \quad (2)$$

where, a_i captures individual time-invariant heterogeneity taken out from the composite error term in Eq.1 and eliminated by time-demeaning in the FE model. Again, β_{TWFE} is the coefficient of interest. All other parameters are defined as above.

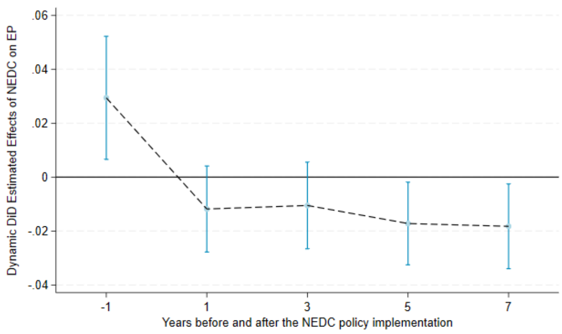
In the unlikely situation in which $Cov(DiD, a_i) = 0$, the estimators of our coefficient of interest will be equivalent in terms of unbiasedness (consistency) and the model should be chosen on grounds of efficiency. In this case, the Random Effects (RE) models would outperform both POLS and TWFE models. By contrast if $Cov(DiD, a_i) \neq 0$, as we show later, TWFE will be less biased (more consistent) than POLS. This ultimately determines the preferred specification for the analysis of the NEDC policy in the next section.

[4] See MWZ for definitions.

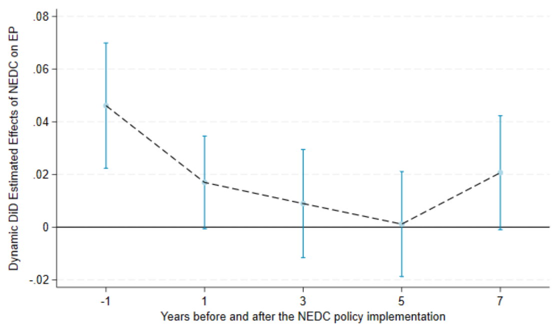
2.3. Results

Before proceeding to the regression analysis, we check for parallel trends either (a) without *or* (b) with controlling for individual effects (both control for provincial effects, time effects, and covariates). As discussed by MWZ, there are two pre-treatment periods, with 2013 set as the baseline (coded as 0). Using the EP index as the dependent variable, Figures 1(a) and 1(b), show that the trends are not parallel with or without controlling for individual effects.

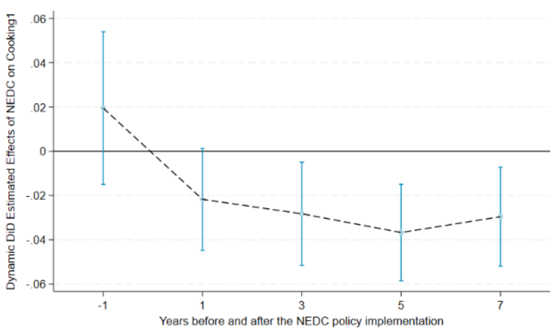
While MWZ use the EP index as their main dependent variable, the authors without explanation (perhaps, unknowingly) switch to using a single sub-component representing cooking fuels (*Cooking 1*) – see MWZ's published code. Therefore, for comparability, Figure 1(c) shows that a parallel trend is exhibited (as in MWZ) without controlling for individual FEs. On the other hand, the trend is non-parallel once controlling for individual effects as seen in Figure 1(d).



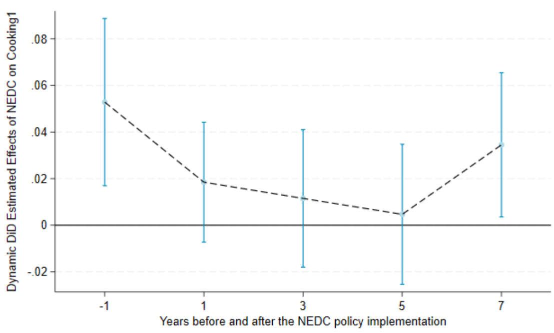
(a) EP index event study controlling for provincial and time effects, and other covariates.



(b) EP index event study controlling for unit, provincial, and time effects, and other covariates.



(c) Cooking fuel event study controlling for provincial and time effects, and other covariates.



(d) Cooking fuel event study controlling for unit, provincial, and time effects, and other covariates.

Fig. 1. Event study diagrams based on the EP index (a)-(b) and the cooking fuel sub-component "Cooking 1" (c)-(d) conditioning on provincial effects and other covariates (a)-(c), as well as unit-specific effects (b)-(d). We use 95% confidence intervals, rather than 90% as in MWZ.

Compared to Figure 1(c), overall Figures 1(a), 1(b), and 1(d) suggest that the NEDC policy may have had limited impact on EP or modern cooking fuels.

Since the parallel trends assumption is potentially violated, the analysis could end here. However, for completeness, we proceed to investigate the NEDCs' potential impact on EP using POLS (Eq. 1) as in MWZ, fixed effects (FE) controlling for individual effects (without time effects), and TWFE (Eq. 2).[5]

Table 1 Columns (1)-(3) show the POLS results, sequentially including the provincial effects, year effects, and control variables. Column (3) closely replicates the MWZ's main result that the coefficient of interest (-0.015) is statistically significant at the 1% level.

Table 1. Regression results using POLS, FE and TWFE

	POLS (Eq. 1)			FE and TWFE (Eq. 2)		
	(1)	(2)	(3)	(4)	(5)	(6)
DiD	-0.023***	0.022***	-0.015***	-0.072***	0.005	0.000
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)
Individual effects	N	N	N	Y	Y	Y
Provincial effects	Y	Y	Y	Y	Y	Y
Time effects	N	Y	Y	N	Y	Y
Controls	N	N	Y	N	N	Y
N	67905	67905	27672	67905	67905	27672

Note: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors, clustered at the individual level, are presented in parentheses. The reduction in observations in columns (3) and (6) is due to the inclusion of covariates (see MWZ for details).

We repeat this analysis in Table 1 Column (4) using unit (individual) FEs (i.e., with neither time effects nor controls), with TWFE in Column (5) without controls, and (6) using TWFE with controls. Column (5) shows that after controlling for time (year) effects the coefficient on *DiD* becomes statistically insignificant at all conventional levels and remains so upon adding further controls in Column (6). Putting statistical significance aside and focusing purely on the

5 The replication of their baseline result is shown in Appendix Table A1, where minor differences can be found in the magnitude of coefficients.

economic significance in Columns (5) and (6), on average, compared with the control group (individuals who do not reside in an NEDC), the findings suggest that EP has not decreased for those who do live in NEDCs due to the treatment.

Overall, the results, along with the above event study analyses, suggest that the NEDC policy has no statistically or economically significant impact on mitigating EP, using standard TWFE models. Robustness checks using alternative EP indices published by MWZ also support the non-existence of an effect of the NEDC policy on mitigating EP (see Appendix Table A2).

3. Conclusion

This study contributes to the literature by showing that correctly specifying fixed effects can alleviate important biases in energy and environmental policy evaluation. In contrast, under a specification that overlooks unit individual-specific FEs, relying instead on regional-specific effects, such POLS estimators may lead to biased estimates of the policy effects due to leaving the unobserved time-invariant individual characteristics unchecked in the error term. Here, we eliminated this potential source of omitted variable bias using standard TWFE defined at the unit (individual-specific) and time (year-specific) levels, rather than group (provincial) and time levels as in MWZ. In so doing, we have highlighted the importance of correctly specifying FEs in standard two-way cases.

MWZ put forward potentially valuable policy implications. For example, energy and environmental policies could be designed to coordinate multi-objectives of energy efficiency and inequality. However, the inability to verify parallel trends and the lack of representativeness of the sample (not the focus of our work), means that such recommendations cannot be relied upon by policymakers, despite their potential importance. Therefore, further research is needed to rigorously evaluate the NEDC or similar policies' ability to alleviate EP.

Overall, our study not only puts forward new evidence on the limited impact the NEDC policy may have had on mitigating EP, but also, and perhaps more importantly, shines lights on the importance of correctly specifying two-way FEs model specifications in policy evaluation.

References

Brehm, J., Pestel, N., Schaffner, S., Schmitz, L. (2025). From low emission zone to academic track: Environmental policy effects on educational achievement in elementary school. *Journal of Environmental Economics and Management*, 103165.

Callaway, B., Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200-230.

Clay, K., Severnini, E., Sun, X. (2023). Does LEED certification save energy? Evidence from retrofitted federal buildings. *Journal of Environmental Economics and Management*, *121*, 102866.

de Chaisemartin, C. D'Haultfœuille, X. (2023). Two-way fixed effects and differences-in-differences estimators with several treatments. *Journal of Econometrics*, 236(2), 105480.

Hoffmann, C., Gangadhar, S. S. B., Müsgens, F. (2025). Smells like Green Energy: The Impact of Bioenergy Production on Residential Property Values in Germany. *Energy Economics*, *145*, 108404.

Imai, K., Kim, I. S. (2021). On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis*, *29*(3), 405-415.

Kang, C., Ota, M., Ushijima, K. (2024). Benefits of diesel emission regulations: Evidence from the World's largest low emission zone. *Journal of Environmental Economics and Management*, *125*, 102944.

Ma, Y., Wan, S., Zhou, Y. (2025). Bridging energy gaps in urbanizing economies: Evidence from China's new energy demonstration city policy on multidimensional energy poverty. *Energy Economics*, 108767.

Nepal, R., Liu, Y., Dong, K. (2025). Adaptive capacity to climate change: Does energy aid matter?. *Energy Economics*, *141*, 108018.

Ren, Q., West, J. (2023). Cleaner waters and urbanization. *Journal of Environmental Economics and Management*, *122*, 102874.

Wu, H., Wang, Z., Zhong, R., Qu, Y., Hao, Y. (2024). The road to eliminating energy poverty: does renewable energy technology innovation work?. *The Energy Journal*, *45*(6), 177-201.

Zhang, H., Di Maria, C., Ghezelayagh, B., & Shan, Y. (2024). Climate policy in emerging economies: Evidence from China's Low-Carbon City Pilot. *Journal of Environmental Economics and Management*, *124*, 102943.

Appendix

Table A1. Comparison between the replication and the original baseline results

	(1)	(2)	(3)
Replicating result MWZ			
DiD	-0.0245***	-0.0219***	-0.0149***
	(-6.6336)	(-5.6737)	(-3.4789)
Published results in MWZ			
DiD	-0.0466***	-0.0211***	-0.0146***
	(-12.4702)	(-5.3565)	(-3.1333)
Individual effects	N	N	N
Provincial effects	N	N	Y
Time effects	N	N	Y
Controls	N	Y	Y
N	67,906	27,672	27,672

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. t-statistics are presented in parentheses for comparability with MWZ. The sample size decreases after including the controls used by MWZ: household size (*size*), education attainment (*eduy*), income level (*lnincome_p*), respondent age (*age*), workplace type (*workplace*), health status (*health*), regional development (*lnPGDP*), urban-rural division (urban), government intervention (*gov*). See more details in MWZ.

Table A2. Robustness results using alternative measures of EP

	POLS (Eq. 1)			TWFE (Eq. 2)		
	(1)	(2)	(3)	(4)	(5)	(6)
	EP2	EP3	ERC	EP2	EP3	ERC
DiD	-0.013***	-0.011***	-0.012***	-0.003	-0.003	0.017***
	(0.004)	(0.004)	(0.003)	(0.006)	(0.006)	(0.007)
Individual effects	N	N	N	Y	Y	Y
Provincial effects	Y	Y	Y	Y	Y	Y
Time effects	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
N	27672	27672	21089	27672	27672	21089

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the individual level, are presented in parentheses. The observations decreased in columns (3) and (6) as there are missing values of ERC. After including individual effects, the sign of the coefficient has changed, which shows the result of POLS based on Eq. 1 are biased. The sample size decreases after replacing the outcome variable MEPI with an alternative measure used by MWZ: ERC (the ratio of household energy consumption to total household income); EP2 is another MEPI using equal weights for each sub-component; and, EP3 is provided in MWZ's dataset but with its meaning is unknown, here we use it as an additional check but caution is clearly needed.