Cluster-Robust Variance Estimators for Instrumental Variables

Theory and Application in Assessing the Impact of Gender Representation on Carbon Emissions

Sam Lee

Abstract

Within applied econometrics, the practice of clustering standard errors by groups that are presumed to have errors correlated within the group but otherwise uncorrelated between other groups has invoked a series of large research on econometric methodology, especially within the past couple of decades. One recent paper in the literature outlines a guide to empirical practice using cluster-robust variance estimators (CRVEs) (MacKinnon et. al 2022). The authors of this paper assess three feasible CRVEs, with particular emphasis on a jackknife estimator. While the authors of this paper assert "for such models [with clustered data estimated by instrumental variables (IV)], neither the current state of econometric theory nor the available simulation evidence allows [the authors] to make recommendations with any confidence," this paper attempts to expand on the guide on empirical practice made by MacKinnon et. al and apply theory to an empirical analysis using instrumental variables. This paper examines what the effect of electing higher proportions of women in African and Arab nations has on yearly CO_2 emissions per capita using years since suffrage as an instrumental variable. To assess the validity of inference, I first propose three feasible cluster-robust variance estimators for a GMM framework in the just-identified case following MacKinnon et. al. I apply the proposed estimators to an empirical example to assess the credibility of inference and to evaluate their long-run performance using a Monte Carlo analysis. I find that one particular estimator may possibly perform more consistently than the others, while the others tend to under

reject.

I. Introduction

The most well-known CRVE used in econometric practice is by far the Liang-Zeger estimator (1986). This formula relies on the asymptotic properties of the OLS estimator. Furthermore, one novel critique of the difference-in-differences (DiD) estimator argues that applied econometric approaches using a DiD approach will typically underestimate the standard deviation of treatment effect without using a proper CRVE (Bertrand et. al 2004). Notably, the paper from Bertand et. al uses a Monte Carlo analysis to generate new random vectors for the response variables to illustrate the empirical coverage of CRVE estimators for both a large and small amount of clusters. In this paper, I will first introduce the CRVEs that MacKinnon et. al proposes as standard practice in applied work. Critically, since the CRVEs used in mainstream practice rely on the asymptotic properties of the OLS estimator, I attempt to parallel the work to a generalized method of moments (GMM) framework suitable for instrumental variables. The objective of this paper is to assess the performance of these proposed estimators and the implications on an important issue in society. In this paper, I analyze the effect of how electing higher proportions of women in African and Arab countries can decrease yearly per capita CO_2 emissions. I discuss why this is an important question both socially and econometrically. I will then explain the data-generating process I will use for my Monte Carlo analysis, my results, and then conclude.

II. Three Feasible CRVEs

The econometric guide from MacKinnon et. al discusses in detail three feasible CRVEs for the following linear regression model:

$$(1) \quad y_i = x_i'\beta + u_i$$

In practice, if the data is divided up into G # of disjoint clusters. Equation (1) can thus be equivalently written as,

$$(2) \quad y_g = X_g\beta + u_g, \quad g = 1, ..., G$$

While this paper serves a purpose to not reinvent the wheel, and rather, works to provide enough preliminary detail so that readers with sufficient econometric backgrounds may understand the steps I took to arrive at my own CRVE, it helps to see the groundwork that previous econometricians have laid before. That being said, without going into too much detail that MacKinnon et. al (among so many others) provide, it follows that the asymptotic properties of (1) follow the usual OLS properties. Assuming data are generated at $\beta = \beta_0$:

$$\sqrt{n}(\beta_0 - \beta) \underset{d}{\rightarrow} N(0, (X'X)^{-1}\mathbb{E}[x_i x_i' u_i^2](X'X)^{-1})$$

Since X is merely the stacked matrix of the clusters, the only difference between the asymptotic properties of β in (1) and (2) is the calculation of $\mathbb{E}[x_i x_i' u_i^2]$. Empirically, for (1), this can be estimated with the sample mean (robust only to unknown forms of heteroskedasticity):

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \hat{u}_i^2$$

Thus, assuming data are divided into G # of disjoint groups, a convenient way to cluster standard errors by each group has been to use the empirical estimation of $\mathbb{E}[X_g' u_g u_g' X_g]$, which follows as:

$$\frac{1}{n}\sum_{g=1}^{G} X_g' u_g u_g' X_g$$

Thus, the first CRVE MacKinnon provide follows directly as,

$$\mathrm{CV}_1 : \frac{G(N-1)}{(G-1)(N-k)}(X'X)^{-1}(\sum_{g=1}^{G} \hat{s}_g \hat{s}_g')(X'X)^{-1}; \quad \hat{s}_g = X_g'\hat{u}_g$$

Where a correction for degrees of freedom is applied. While MacKinnon, et. al note that this is the most widely used cluster-robust formula in use, the authors keenly mention that $X_g'\hat{u}_g$ is not always a good estimator for s_g. Two alternative CRVEs[1] have been proposed by the literature to address this issue by transforming the residual score vectors ($\hat{s}_g$) (Bell & McCaffrey, 2002).

$$\mathrm{CV}_2 : (X'X)^{-1}(\sum_{g=1}^{G} \grave{s}_g \grave{s}_g')(X'X)^{-1}; \quad \grave{s}_g = X'M_{gg}^{-1/2}\hat{u}_g; \quad M_{gg} = I_{N_g} - X_g(X'X)^{-1}X_g'$$

$$\mathrm{CV}_3 : \frac{G-1}{G}(X'X)^{-1}(\sum_{g=1}^{G} \acute{s}_g \acute{s}_g')(X'X)^{-1}; \quad \acute{s}_g = X'M_{gg}^{-1}\hat{u}_g$$

III. A Cluster-Robust Variance Estimator for Instrumental Variables

While the empirical guide outlined by MacKinnon et. al is useful and practical in its own right, it has its limitations. The cluster-robust variance estimators discussed in the paper only make use of the empirical matrix X, an $N \times k$ matrix presumed to have none of its independent variables violating endogeneity. The authors candidly assert that "for such models [with clustered data estimated by instrumental variables (IV)], neither the current state of econometric theory nor the available simulation evidence allows [the authors] to make recommendations with any confidence." Thus, as a matter of econometric exploration, I attempt to push this boundary by using the pre-existing econometric research and literature to propose a cluster-robust variance estimator for instrumental variables in the generalized method of moments (GMM). For the simplification of this paper, I only consider the just-identified case.

[1] As MacKinnon et. al pay particular interest to CV_3. As the authors note, this cluster-robust variance estimator is a jackknife estimator and is discussed thoroughly throughout their paper.

GMM Framework for Instrumental Variables

In the GMM framework, we will first consider the moment condition for instrumental variables:

$$\mathbb{E}[m_i(\beta) = Z_i(Y_i - X_i'\beta)] = 0; \quad G(\beta) := \left[\frac{\partial m_i(\beta)}{\partial \beta'}\right] = -\mathbb{E}[Z_i X_i'] \quad \text{(has rank k)}$$

Assuming conditions for identification are met[2], for data generated with $\beta = \beta_0$, the asymptotic properties of GMM estimation follow:

$$\sqrt{n}(\beta_0 - \beta) \underset{d}{\rightarrow} N(0, \left[G(\beta)\mathbb{E}[m_i(\beta)m_i(\beta)']^{-1}G(\beta)'\right]^{-1}) \implies$$
$$\sqrt{n}(\beta_0 - \beta) \underset{d}{\rightarrow} N(0, \left[\mathbb{E}[(X_i Z_i)]\mathbb{E}[Z_i Z_i' u_i^2]^{-1}\mathbb{E}[(X_i Z_i)']\right]^{-1})$$

The objective of this section will be to outline a reasonable cluster-robust variance estimator following the research that has been done on this topic. Moving from equation (1) to a two-stage least squares design, assume data is generated from the following just-identified two-stage least squares process, where W_i is endogenous. Assume exclusion, independence, and relevance conditions are all met for the instrument z_i. Let the empirical just-identified two-stage regression be characterized as

$$(3) \quad w_i = z_i'\lambda + v_i$$
$$(4) \quad y_i = [w_i|q_i]' \left[\frac{\delta}{\alpha}\right] + \epsilon_i = x_i'\beta + \epsilon_i$$

Where δ is the (local average) treatment effect of interest on w_i. Let q_i be a vector of included covariates[3] with corresponding effects, α.

Thus, let the variance estimator of interest from the GMM framework to be,

[2] While I don't discuss the assumptions and conditions for identification in detail in this paper, I do discuss in detail the assumptions for IV identification in my applied example in another paper here. However, in addition to the assumptions briefly mentioned above, I also assume that β to uniquely solves the moment condition.

[3] I notate equation (4) with augmented vectors to (i) show how I structure the data in the code and (ii) to show the connection between the first and second stages.

$$V = \left[\mathbb{E}[(X_i Z_i)]\mathbb{E}[Z_i Z_i' u_i^2]^{-1}\mathbb{E}[(X_i Z_i)']\right]^{-1}$$

Using the sample means as empirical estimates, an estimate for V reasonably follows as[4] (assuming homoskedasticity),

$$\hat{V} = \left[\left(\frac{1}{n}X'Z\right)\left(\frac{1}{n}Z'Z\hat{u}^2\right)^{-1}\left(\frac{1}{n}X'Z\right)'\right]^{-1}$$

Now, assuming the data can be classified into G # of disjoint clusters, using the same logic that by which the OLS CRVE was derived, $\hat{V}$ can be respectively rewritten as:

$$\hat{V}_g = \left[\left(\frac{1}{n}X'Z\right)\left(\frac{1}{n}\sum_{g=1}^{G} Z_g'\hat{u}_g\hat{u}_g'Z_g\right)^{-1}\left(\frac{1}{n}X'Z\right)'\right]^{-1}$$

Simplifying, we arrive at our first CRVE for instrumental variables in a GMM framework (and by including the degrees of freedom correction[5] from CV_1):

$$\overset{iv}{CV}_1 : n\frac{G(N-1)}{(G-1)(N-k)}\left[(X'Z)\left(\sum_{g=1}^{G}\hat{\zeta}_g\hat{\zeta}_g'\right)^{-1}(X'Z)'\right]^{-1} \quad ; \quad \hat{\zeta}_g = Z_g'\hat{u}_g$$

Thus, the OLS cluster-robust variance estimators from Section II of this paper can all be rewritten along the same line of theory:

$$\overset{iv}{CV}_2 : n\left[(X'Z)\left(\sum_{g=1}^{G}\grave{\zeta}_g\grave{\zeta}_g'\right)^{-1}(X'Z)'\right]^{-1} \quad ; \quad \grave{\zeta}_g = Z_g'\overset{iv}{M}_{gg}^{-1/2}\hat{u_g}; \quad \overset{iv}{M}_{gg} = I_{N_g} - Z_g(Z'Z)^{-1}Z_g'$$

$$\overset{iv}{CV}_3 : \frac{n(G-1)}{G}\left[(X'Z)\left(\sum_{g=1}^{G}\acute{\zeta}_g\acute{\zeta}_g'\right)^{-1}(X'Z)'\right]^{-1} \quad ; \quad \acute{\zeta}_g = Z_g'\overset{iv}{M}_{gg}^{-1}\hat{u}_g$$

4 By the Law of Large Numbers and the Continuous Mapping Theorem, $\hat{V}$ converges to V.

5 Interestingly enough, Stata uses this CRVE when the `cluster` option is specified in its `ivregress` command. However, its formula ignores the degrees of freedom correction.

The goal of this paper will be to show whether the asymptotic properties of these estimators hold (if any). To demonstrate this, I will now apply these formulas to an applied example that uses instrumental variables.

IV. Empirical Example

Do Increased Proportions of Women in National Legislatures of African and Arab Countries Significantly Decrease per Capita CO_2 Emissions?

I seek to identify the causal effect that electing higher proportions of women in national legislatures in African and Arab nations has on yearly per capita CO_2 emissions. I utilize a two-stage difference-in-difference approach on 64 different countries throughout Africa and the Middle East from 1998-2022. To eliminate the potential endogeneity in using the proportion of women in national legislatures as a treatment on per capita CO_2 emissions, I use years since women were granted suffrage in each respective country as an instrumental variable. One nominal paper in the literature has shown evidence that increased proportions of women in parliaments are more than likely causally related to stricter climate policies (Mavisakalyan & Tarverdi, 2019). However, this paper does not take into account how carbon emissions and women in government have evolved. To my knowledge, little research has been done to show an intertemporal causal link between women in parliaments and its effects on climate. I expand on this research by showing how carbon emissions have changed over time to elicit a local average treatment effect between changes in gender compositions at the national legislature level.[6].

Background

The current climate and economic research today overwhelmingly support the thesis that women—especially in developing countries—are disproportionately affected by the negative externalities of

[6] The full applied analysis for this project can be found on this project's Github repository: https://github.com/SamLeeBYU/Elections/Report.pdf.

climate change. During periods of increased droughts, which are exacerbated due to the effects of climate change, women often make long trips to meet agricultural, hygiene, and family needs. This has been linked to increased violence against women (UNDP, 2019). Increased risks of poverty and food insecurities have also been shown to increase as a result of increased rates of natural disasters and variable weather patterns. One article reports that "when a family is faced with the impact of the climate crisis, girls' education is one of the first things families drop" (Medlicott, 2021). Another meta-analysis consisting of 53 studies report that two-thirds of those studies find that women are more often the victims of death or injury in the case of extreme weather events (Seller, 2016).

If women are disproportionately affected by climate externalities, one would expect that female policymakers would enact effective climate policy at higher rates relative to other policymakers, assuming they have the means and power to do so. Thus, the causal question of interest reasonably follows as to whether increased proportions of women decrease CO_2 emissions. As we will see, the nature of causality of this question is rather difficult to elicit due to intricate endogeneities that persist throughout this problem.

The purpose of the empirical section of this paper seeks not to convince readers that the complex problem of endogeneity and causality can be resolved (although these issues are intricately discussed in my applied work). Rather, this empirical example is included for two motivating reasons: (1) The necessity and obligation of answering the question at hand as a result of expanding on Mavisakalyan and Tarverdi's research has great implications for climate policy, and (2) The econometric implications of using a properly adjusted cluster-robust variance estimator have similar impacts for the reliability of conclusions derived from this study and the econometrics field as a whole.

Data

The data used for this analysis was strategically gathered using publicly available sources[7]. The Inter-Parliamentary Union (IPU) maintains a consistent archive of records that monitors the number of women elected into the national legislature for each country. Not every government maintains the same structure of government, however, the IPU distinguishes between women elected into the lower house and upper house levels of government throughout multiple periods. I take the average proportion of women holding office in the national legislature overall in a given year by taking a weighted average of the two houses[8] (See Figure 1). I obtained emissions data, including total yearly (gigatons of) CO_2 emissions per capita per country, from the Emissions Database for Global Atmospheric Research (EDGAR) as maintained by the European Commission (See Figure 2).

I obtained country-year-level specific covariates, such as population and GDP, from the World Bank (See Table 1 for summary statistics on key variables). I obtained the database linking each country's year they passed a right of universal suffrage from "A Lexical Index of Electoral Democracy" on the Harvard Dataverse (Skaaning et. al, 2015). I checked country-specific anomalies using research from the Pew Research Center (Schaeffer, 2020). Not all countries in Africa and the Middle East treat women equally. Even when suffrage is granted, it is not granted universally. Thus, the instrument may not influence the proportion of women in government the same in every country. This does not discredit my econometric analysis, however, so long as the monotonicity assumption of my instrument is met. However, I exclude obvious anomalies in my analytic sample (See Table 2 for summary statistics for each country). In my sample, the United Arab Emirates (UAE) and Saudia Arabia are excluded because they don't give full suffrage to women[9]. Finally, I used the U.S. Department of State's regional definitions to define which countries I define as either

[7]Data was collected using ethical web-scraping methods or obtained via direct download. Further documentation and the scripts used to obtain the data used for this analysis can be found on this project's Github repository https://github.com/SamLeeBYU/Elections.

[8]Econometrically, this assumes that on average, the power that women are given in each house of national legislature is equal to the 'weight' of women in that chamber of legislature: In other words, assuming a fair political system, I assume no chamber can systemically dominate the other chamber.

[9]The UAE held its first national elections in 2006, but only a select amount of men and women were eligible to participate (Schaeffer, 2020). Saudi Arabia doesn't hold national elections. Women are, however, enfranchised at the local level government levels.

Figure 1 − Average Proportion of Women in National Legislatures Over Time

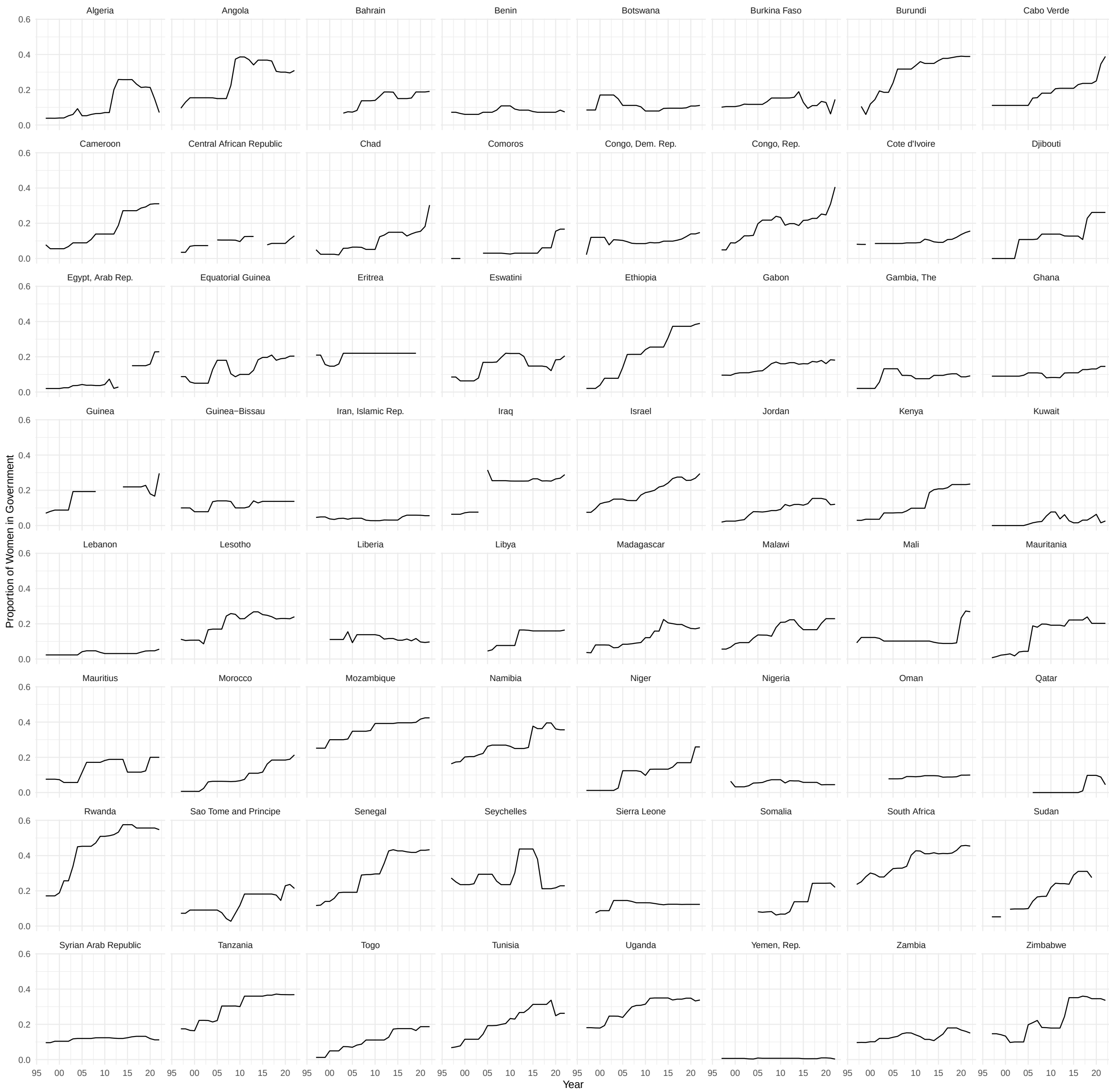
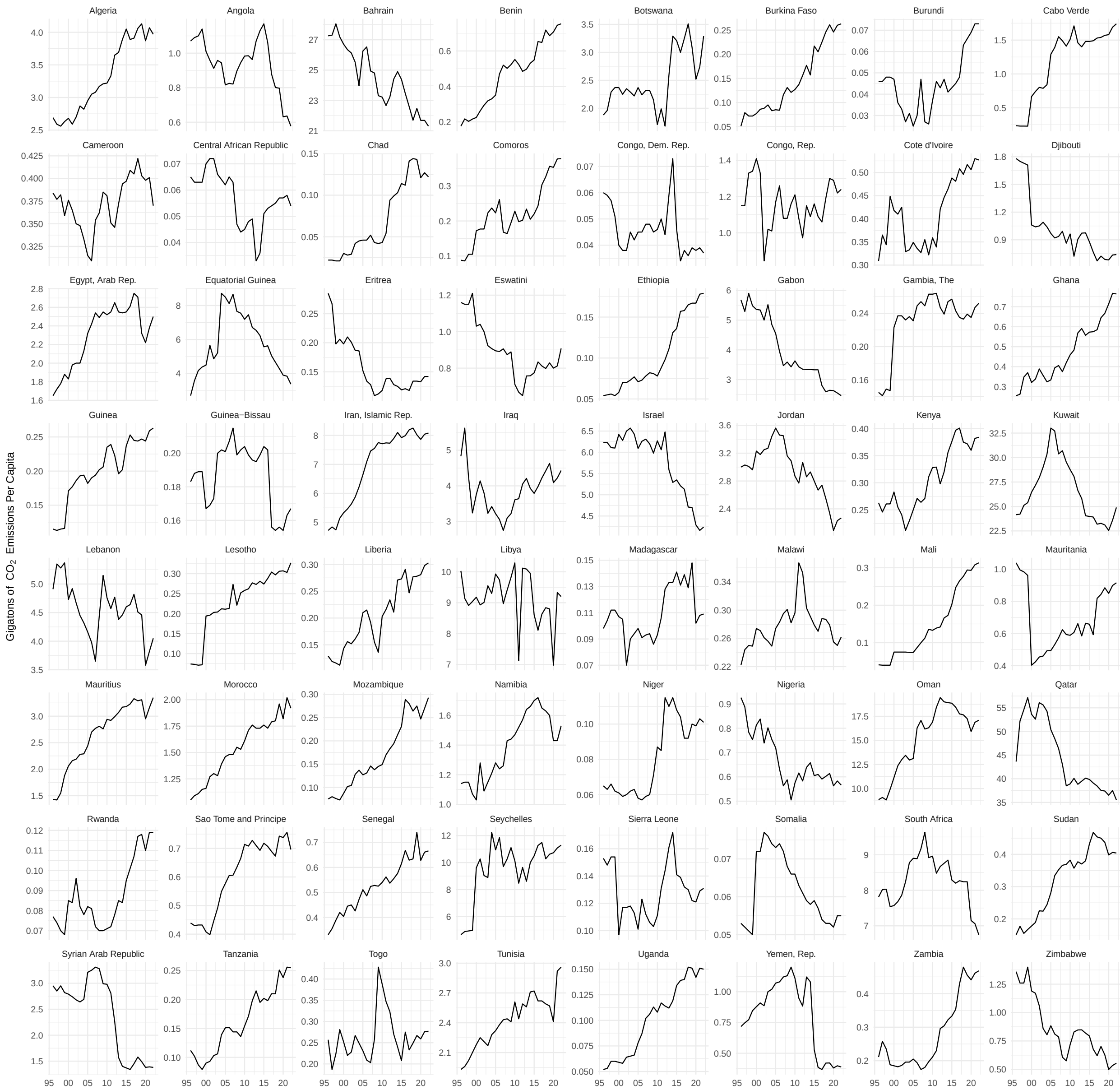


Figure 2 − Per Capita Carbon Emissions Over Time



African or Arab.

Table 1: Summary Statistics of Key Variables

Variable	Mean	SD	Min	Max
CO2 Emissions Per Capita	3.224	7.345	0.021	57.16
Proportion of Women in Gvt	0.155	0.108	0	0.575
ln(GDP)	23	1.7	18	27.2
ln(Population)	15.851	1.545	11.256	19.202
Suffrage	1968.109	13.363	1947	2006

One potential problem with data is that I do not account for country-specific government intricacies that could prevent women or a legislature from passing effective climate policy—such as not accounting for government corruption or other unique anomalies (wars between countries, coups, laws that are discriminatory against women, etc, that prevent a government from acting efficiently)—at every year[10]. Merging the suffrage data with all the IPU available data and the CO_2 emissions data from EDGAR, I was left with 64 countries spanning 1997-2022 [11].

Identification

To estimate the causal effect that electing more women in national legislatures in African and Arab countries has on reducing per capita CO_2 emissions, I employ a two-stage difference-in-difference identification strategy. In regard to curbing the negative externalities of climate change in the context of this paper, I use per capita CO_2 emissions as a response variable. Let Y_{ct} represent the total (gigatons of) carbon emissions per capita for a country (c) in a year (t). I seek to identify consistently and interpret the local average treatment effect of δ, the effect of a higher proportion of women in national legislatures in African and Arab countries on national CO_2 emissions per

[10]To assert consistency in my estimates, econometrically, I assume that on average, these factors in the idiosyncratic term of my econometric model are 0 for any given country at year t. (See equation (7)).

[11]Notably, the data is left as an unbalanced panel data set (Refer to Table 2). The World Bank does not have GDP or population data for every single year of every single country in the data, and critically, the IPU also notes that it cannot reliably retrieve the number of women in national legislatures for every year. These years for those specific countries are naturally excluded. I assume that the probability at which data for a specific year-country combination will be missing is essentially random. This assumption may be violated if countries that are systemically difficult to obtain data for (such as Somalia or the Central African Republic) is also correlated with other covariates and or CO_2 emissions.

Table 2: Summary Statistics by Country

Country	N	Mean CO2 Emissions Per Capita	Mean Proportion of Women in Gvt	Mean GDP	Mean Population	Suffrage
Algeria	26	3.33192	0.12144	133413556539	36256149.08	1962
Angola	26	0.92681	0.25617	61136510425	23819407.27	1975
Bahrain	20	24.41692	0.14652	23866885356	1110339.23	2002
Benin	26	0.48000	0.07758	9427907948	9513753.81	1960
Botswana	26	2.51423	0.10956	11679307166	2093686.46	1966
Burkina Faso	26	0.14869	0.12574	9879179362	16222351.12	1960
Burundi	25	0.04481	0.29223	1858691481	9013842.27	1962
Cabo Verde	26	1.22450	0.18505	1483184743	517811.46	1975
Cameroon	26	0.37169	0.16573	26614285194	20110188.96	1960
Central African Republic	23	0.05631	0.09069	1717825115	4516393.23	1960
Chad	26	0.07246	0.09761	8092429967	12009809.12	1960
Comoros	22	0.22754	0.04802	839948182	658580.65	1975
Congo, Dem. Rep.	26	0.04538	0.10226	26137691195	67798246.61	1967
Congo, Rep.	26	1.16296	0.19112	9890831710	4342198.92	1960
Cote d'Ivoire	25	0.41704	0.09852	37241947845	21228611.50	1960
Djibouti	26	0.98342	0.12020	1568798898	908429.23	1977
Egypt, Arab Rep.	24	2.32654	0.07382	210728516057	87759979.69	1956
Equatorial Guinea	26	5.84538	0.13341	10680310300	1098129.11	1968
Eritrea	23	0.15304	0.20727	1120484041	3025058.73	1993
Eswatini	26	0.88365	0.14917	3443873221	1098973.65	1968
Ethiopia	26	0.10427	0.21588	43426199788	89604476.46	1955
Gabon	26	3.89038	0.14328	12441574754	1731276.23	1960
Gambia, The	26	0.23292	0.08174	1276629190	1948392.35	1965
Ghana	26	0.48408	0.10488	35446770393	25480258.61	1957
Guinea	21	0.20485	0.17269	7558680162	10412517.15	1958
Guinea-Bissau	26	0.18654	0.11840	874252813	1580784.61	1974
Iran, Islamic Rep.	26	7.04769	0.04200	323334435131	75438778.31	1979
Iraq	25	3.84692	0.20797	128507863985	32491105.15	1958
Israel	26	5.72923	0.18622	250706702966	7614703.85	1948
Jordan	26	2.93385	0.09033	25772832748	7476728.62	1974
Kenya	26	0.30900	0.12401	49487400802	40920573.00	1963
Kuwait	26	26.68577	0.02475	102805583557	3032646.92	2006
Lebanon	26	4.53423	0.03430	32455205151	5123114.12	1952
Lesotho	26	0.23662	0.19993	1813006912	2071184.42	1966
Liberia	24	0.20858	0.11397	2268958422	3925143.88	1947
Libya	18	9.11577	0.12542	52387031513	5970720.88	1963
Madagascar	26	0.10985	0.12502	9465270164	21695233.31	1960
Malawi	26	0.27958	0.15362	7514886577	14841171.27	1964
Mali	26	0.15212	0.12074	10274167660	15662439.38	1960
Mauritania	26	0.67550	0.14266	5071728438	3476739.54	1960
Mauritius	26	2.70000	0.13275	9303361634	1234473.96	1968
Morocco	26	1.56115	0.09184	91087307855	32367924.31	1962
Mozambique	26	0.17431	0.35431	11612120970	23558217.61	1975
Namibia	26	1.39731	0.27434	9018993290	2113443.23	1990
Niger	26	0.08123	0.10870	7682887769	17080118.61	1960
Nigeria	24	0.66092	0.05481	313024619441	161517413.54	1977
Oman	19	15.47154	0.08991	56775810363	3281428.62	2003
Qatar	17	44.70962	0.02556	108516614223	1637385.77	2003
Rwanda	26	0.08800	0.44140	5978302715	10348245.77	1962
Sao Tome and Principe	26	0.61331	0.13043	255006484	179928.11	1975
Senegal	26	0.53638	0.29679	15447790277	12665052.88	1960
Seychelles	26	9.72346	0.28266	1080544320	89129.92	1976
Sierra Leone	25	0.12796	0.12040	2694746567	6391405.35	1961
Somalia	18	0.06238	0.14398	8216255642	12121495.23	1964
South Africa	26	8.27500	0.36477	308320233630	52097664.77	1994
Sudan	22	0.32985	0.18022	47869298018	34068352.38	1964
Syrian Arab Republic	26	2.33077	0.11750	84109722132	19402662.58	1953
Tanzania	26	0.16458	0.29519	34798787505	46143525.54	1963
Togo	26	0.26342	0.10986	4458357376	6566135.35	1960
Tunisia	26	2.44115	0.20983	37604341839	10925330.88	1959
Uganda	26	0.10385	0.28969	21152773534	32832378.15	1962
Yemen, Rep.	26	0.81108	0.00665	22320277039	24799160.46	1990
Zambia	26	0.28300	0.13306	15927797878	13906848.08	1964
Zimbabwe	26	0.83592	0.22631	13726241713	13273904.50	1979

capita. Research from Mirziyoyeva and Salahodjaev (2023) suggests that women in government promote economic growth. This could pose potential problems of reverse causality: We need to be certain that any effect we see on women's representation causes a reduction (or increase) in carbon emissions. If carbon emissions are highly correlated with economic growth and industrialization then an OLS regression (that is, using a typical difference-in-differences approach) on Y_{ct} to estimate δ would almost certainly be biased and inconsistent. To first eliminate the endogeneity for reverse causality—the supposition that perhaps lower carbon emissions cause a higher proportion of women to become elected during the same year; or rather, perhaps higher carbon emissions are correlated with other variables such as economic growth that are linked to higher proportions of women becoming elected—I lag the treatment variable of interest, W_{ct}, the average proportion of women in national legislature for country c during year t one year. This implies that any change in Y_{ct} will inherently come *after* a change in W_{ct}. In the two-stage design, I further suspect that there exists some endogeneity in W_{ct}[12]. Both CO_2 emissions and women's representation in government are highly correlated with industrialization and development. While I do not focus on specific regulatory practices within this paper, an a priori assumption could be that effective climate policy would consist of enacting particular policies and regulations that would reduce the CO_2 emissions by economic constraint. It could be that increased proportions of women in government could (causally) affect economic performance within a country at time t. Hence, I include a sequence of matrices of autoregressive economic controls, represented by $\sum_{p=2}^{P} X_{ct-p}$ (where p is the pth lag of the economic covariate[13]). Thus, to establish the causal chain of inference, I infer:

$$\sum_{p=2}^{P} X_{ct-p} \rightarrow W_{ct-1} \rightarrow Y_{ct}$$

Where X_{ct-p} is a matrix of controls related to economic growth. For the same reason the W_{ct} is lagged one year to ensure that a change in Y_{ct} comes after a change in W_{ct}, I include a series

[12]Mathematically, with endogeneity in the treatment variable, I assume $\exists \quad c_0, t_0 \quad s.t. \quad \mathbb{E}[W_{c_0 t_0 - 1} \epsilon_{c_0 t_0}] \neq 0$.
[13]A priori I set $P = 5$, the total number of lags on all the economic covariate controls.

of autoregressive lags in the economic controls (starting at lag 2) to ensure that changes in W_{ct-1} occur *after* changes in economic development. For this two-stage regression, X_{ct-p} consists of logged GDP and logged population for a country at year t[14]. While X_{ct} can eliminate some of the endogeneity of W_{ct} by including economic controls, these covariates alone are not enough to explain a country's development and how it interacts with how women are elected into national legislatures. The rate at which a country develops, both industrially and politically, in regards to the countries in this study may be highly related to which European power originally occupied the territory and thus, how the power left the state when the occupation ceased, including the political institution(s) established. Endogeneity may also arise in what natural resources are available to the country at year t. Countries with a greater propensity to use carbon-intensive resources will thus have, on average, higher CO_2 emissions. If this also impacts women disproportionately such that women in these countries are more compelled to run for office and enact effective climate policies, then the estimand for δ will be biased in the negative direction.

To eliminate this endogeneity I invoke a two-stage difference-in-difference design, using an instrumental variable for W_{ct-1}. A common instrument in the literature for women's representation in parliament has been to use the number of years since the country granted suffrage (Mavisakalyan & Tarverdi, 2019). I use a similar instrumental variable design here. Let Z_{ct-1} be the number of years since a country c has granted suffrage (measured at year t, and lagged one year to align with W_{ct-1}). In the difference-in-difference design, I use 62 country-fixed effects (omitting the Qatar and Iraq dummies due to colinearity). After lagging W_{ct} one year, I am left with data spanning 1998-2022, thus, including year-fixed effects, I also omit 1998 for colinearity. I utilize the following two-stage estimation approach:

$$(5) \quad W_{ct-1} = \pi_0 + \lambda Z_{ct-1} + \sum_{p=2}^{P} X'_{ct-p}\Psi_p +$$
$$\sum_{k=\text{Angola}\ k\neq\text{Qatar, Iraq}}^{Zimbabwe} \mu_k \mathbb{1}(\text{Country}_c = k) + \sum_{j=1999}^{2022} \eta_j \mathbb{1}(\text{Year}_t = j) + v_{ct}$$

[14] A further analysis could (and should) include further economic controls that may affect W_{ct-1} such as fertility rate, etc.

$$(6) \quad Y_{ct} = \beta_0 + \delta W_{ct-1} + \sum_{p=2}^{P} X'_{ct-p}\Omega_p +$$

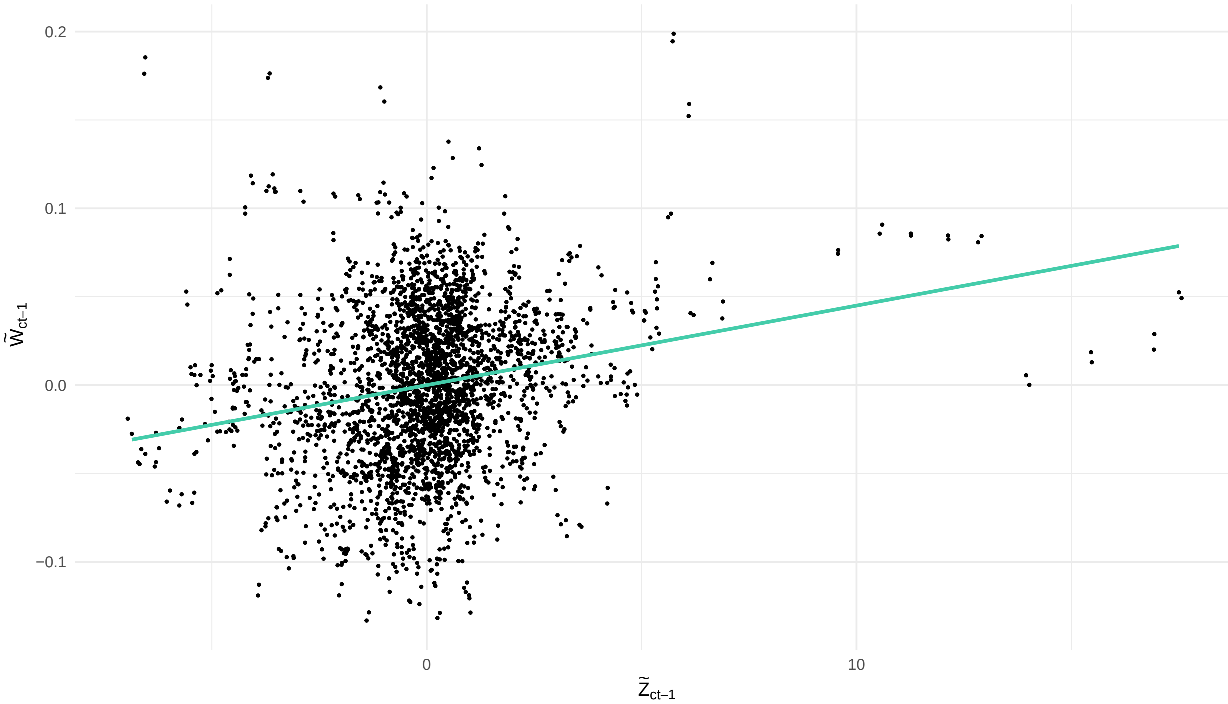
$$\sum_{k=\text{Angola; } k \neq \text{Qatar, Iraq}}^{Zimbabwe} \beta_k \mathbb{1}(\text{Country}_c = k) + \sum_{j=1999}^{2022} \gamma_j \mathbb{1}(\text{Year}_t = j) + \epsilon_{ct}$$

It is worth noting that equations (5) and (6) are of the two-stage form found in equations (3) and (4), though with a lot more added covariates. More specifically, the data in this example can divided into G # of disjoint groups. Clusters in this example could be defined as either years or countries (i.e. disjoint groups). However, it is more intuitive to think of the clusters as the countries in this example. Econometric literature would suggest to cluster by groups where the 'treatment' occurs (Bertrand et. al, 2004). We assume that standard errors are correlated within countries but not between countries.

GMM Application

With an instrumental variables framework, we will use GMM to obtain estimations and confidence intervals for δ. While the validity of the instrument has been discussed elsewhere, it is worth noting that with significant confidence, the relevance condition for the instrument has been met. By partialling out the fixed effects and covariates, we can obtain both a graphical and exact estimation of λ (the first stage coefficient):

Figure 3 − Relevance of Instrumental Variable Using Partialled−out Regressor and Regressand



The regression coefficient is significantly positive: 0.0045 (0.00343, 0.00556), which, given that this identification is just-identified, implies $-\mathbb{E}[Z_{ct-1}\mathbb{X}_{ct}]$ has rank k, where $\mathbb{X}_{\mathbb{ct}}$ is the full covariate X matrix (including fixed effects).

As for the moment condition, I assume,

$$(7) \quad \mathbb{E}[m_i(\beta)] = \mathbb{E}[Z_{ct-1}(Y_{ct} - \mathbb{X}'_{ct}[\beta_0|\delta|\Omega_1|...|\Omega_4|\beta_{\text{Angola}}|...|\beta_{\text{Zimbabwe}}|\gamma_{1999}|...|\gamma_{2022}]')] = 0$$
$$\iff \mathbb{E}[Z_{ct-1}v_{ct}|\mathbb{X}_{ct}] = 0 \quad \forall c, t$$

For identification, I also assume that the vector of parameters, β, uniquely solves (7).

Estimation

Assuming the conditions for identification and inference are met for the GMM identification strategy, we can use the standard two-stage least squares regression formula[15]. Assuming homoskedasticity (using the formula for $\hat{V}$), we arrive at the following 95% C.I. for δ, the local average treatment effect of interest:

$$-107.8425 \quad (-133.36956, -82.31544)$$

This means that (conditioned on the proportion of women being influenced by the instrument of years since suffrage was granted in the country), for every 1% increase in the average proportion of women in national legislatures of African and Arab countries, we would expect yearly per capita CO_2 emissions to decrease by 1.08 gigatons (-1.33, -0.82), on average[16].

If we use the proposed cluster-robust estimators in Section III, we get significantly wider intervals (although still *statistically* significant):

Table 3: 95% Confidence Intervals of $\hat{\delta}$

Estimator	$\overset{iv}{CV}_1$ (2.5%, 97.5%)		$\overset{iv}{CV}_2$ (2.5%, 97.5%)		$\overset{iv}{CV}_3$ (2.5%, 97.5%)	
	-164.88	-50.81	-165.661	-50.02	-166.86	-48.83

The natural question that follows is whether or not we can trust the inference results on this regression output. Empirical results from Bertrand et. al (2004) suggest that the homoskedasticity assumption on $\epsilon_c t$ causes us to drastically understate the standard errors since CO_2 emissions are correlated within a country over time. Even if higher concentrations of women in national legislatures decrease carbon emissions through regulatory practice, there are reasonable arguments to suggest that the effects of those regulatory practices may not be fruit-bearing until years later down

[15] $(Z'X)^{-1}Z'Y$ for just-identified models. This was calculated using matrix algebra in R. All scripts for regression calculations and simulations can be found in https://github.com/SamLeeBYU/Elections/CRVE.

[16] It is also worth noting that the only other covariates that were significant in the regression (given that standard errors were calculated assuming homoskedasticity) besides fixed effects were the coefficients on GDP$_{ct-2}$ and Population$_{ct-2}$, which were both significantly positive, though not nearly as high in magnitude as $\hat{\delta}$.

the road. After all, industries may need time to adjust their carbon-intensive practices into compliance. Thus, if the logic to cluster the standard errors by country prevails, which cluster-robust variance estimator should we use? Can we trust our inferences? I examine this question in the next section.

V. Monte Carlo Analysis to Assess Performance of Cluster-Robust Variance Estimators

To assess which (if any) of my proposed cluster-robust variance estimators are appropriate for a (just-identified) instrumental variables framework in a GMM setting, I will apply a Monte Carlo analysis to our empirical example. The goal of this analysis will be to see if (i) the results from our empirical analysis can be reasonably trusted. Assuming the conditions for identification hold, can we accurately assess the (local average treatment) effect that electing more women to national legislatures in African and Arab countries has on yearly per capita CO_2 emissions? This analysis also serves to (ii) analyze the long-run performance of the cluster-robust variance estimators I proposed in Section III.

Data-Generating Process

The data-generating process will proceed as follows. While it is possible to choose a pre-determined β_0 vector as other authors have previously done in the literature and generate completely new data according to a similar framework outlined by equations (3) and (4), I wish to simulate data using the same distribution of my sample data in my empirical example. To do this, I will perform a bootstrap sampling of Y_{ct}. By leaving my covariates, instrument, and treatment data the same as in the sample data, I will replace the empirical sample Y_{ct} with a new bootstrap sample of Y_{ct}, which I will define as $\mathring{Y}_{ct}$. By running GMM estimation for two-stage least squares for each of the proposed CRVEs, I will assess the long-run empirical coverage of the confidence intervals using many iterations (simulations) of $\mathring{Y}_{ct}$. Mathematically[17], I will define $\mathring{Y}_{ct}$ as,

$$\mathring{Y}_{ct} = \left\{\mathring{y}_{ct}\right\}_{c=\text{Angola},\, t=1999}^{c=\text{Zimbabwe, t=2022}} = \left\{\mathring{y}_{\text{Angola, 1999}}, \mathring{y}_{\text{Angola, 2000}} \cdots, \mathring{y}_{\text{Zimbabwe, 2022}}\right\}$$

$$Pr\left(\mathring{y}_{ct} = y_{ct}\right) = \frac{1}{n}, \quad \forall c, t$$

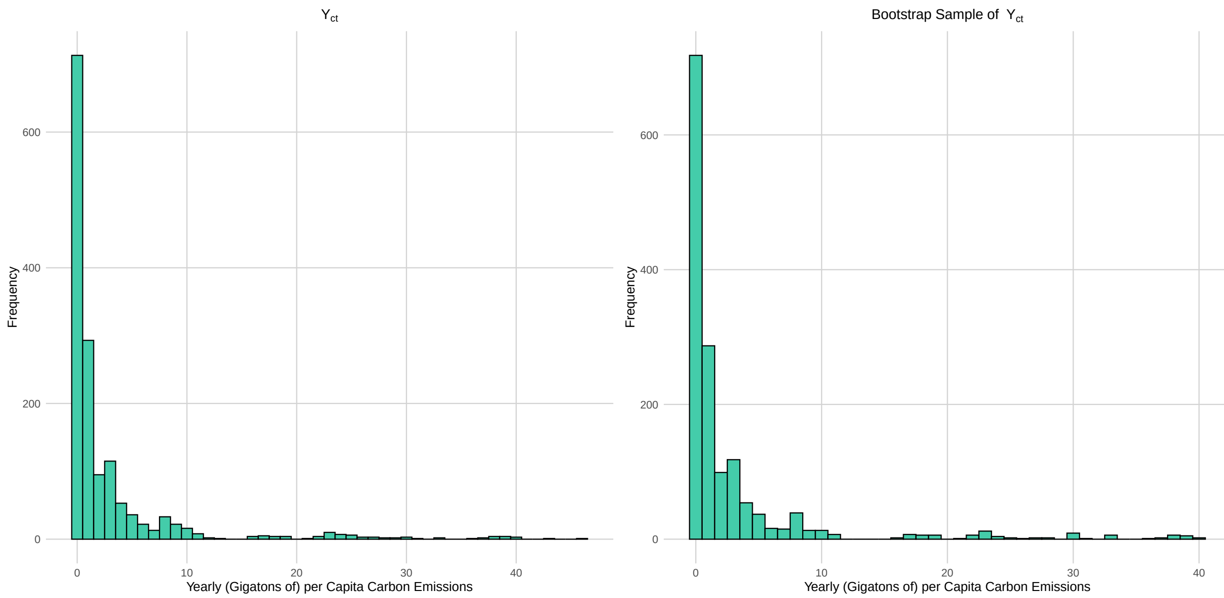


Figure 4 – Comparison between original dependent variable and bootstrapped dependent variable.

Simulation

The DGP simulates placebo vectors for Y_{ct} ($\mathring{Y}_{ct}$). If the results from the main regression were not obtained by random chance, we would expect that $\mathring{Y}_{ct}$ would testify to this effect. Thus, on average—assuming the standard errors are sized correctly—using a 95% confidence interval, we would expect the empirical coverage of a significant $\hat{\delta}$ to approach 5% using $\mathring{Y}_{ct}$ in place of Y_{ct}. However, if the results were obtained by some misspecification or some other unobserved factor, it may be that the empirical coverage may not converge to 5%, and the GMM variance estimators may fail to deliver on their promise. In such a scenario, we will not be able to put much confidence in the two-stage difference-in-differences regression results. Simultaneously, if we can put our

[17]Note, with the probability statement defined as is, this means that each element in Y_{ct} is sampled from a uniform distribution (or, equivalently, sampled with replacement with equal probability).

confidence in the regression results, we can also evaluate the empirical performance of the proposed cluster-robust variance estimators and determine which estimator may be preferential in a just-identified GMM setting that uses instrumental variables.

I run 10,000 simulations, generating a new $\overset{\circ}{Y}_{ct}$ for each simulation. I then calculate the estimates for δ (the local average treatment effect of interest) and the respective 95% confidence intervals using the three proposed CRVEs in Section III as well as the 95% confidence interval using the equation for $\widehat{V}$, which assumes a homoskedastic variance on ϵ_{ct}, as a baseline measurement. Notably, to calculate $\overset{iv}{CV}_2$ and $\overset{iv}{CV}_3$ requires the calculation of $\overset{iv}{M}_{gg}$. In order to calculate this matrix, it is requisite to partial out fixed effects to avoid singularity (MacKinnon et. al, 2022).

Simulation Results

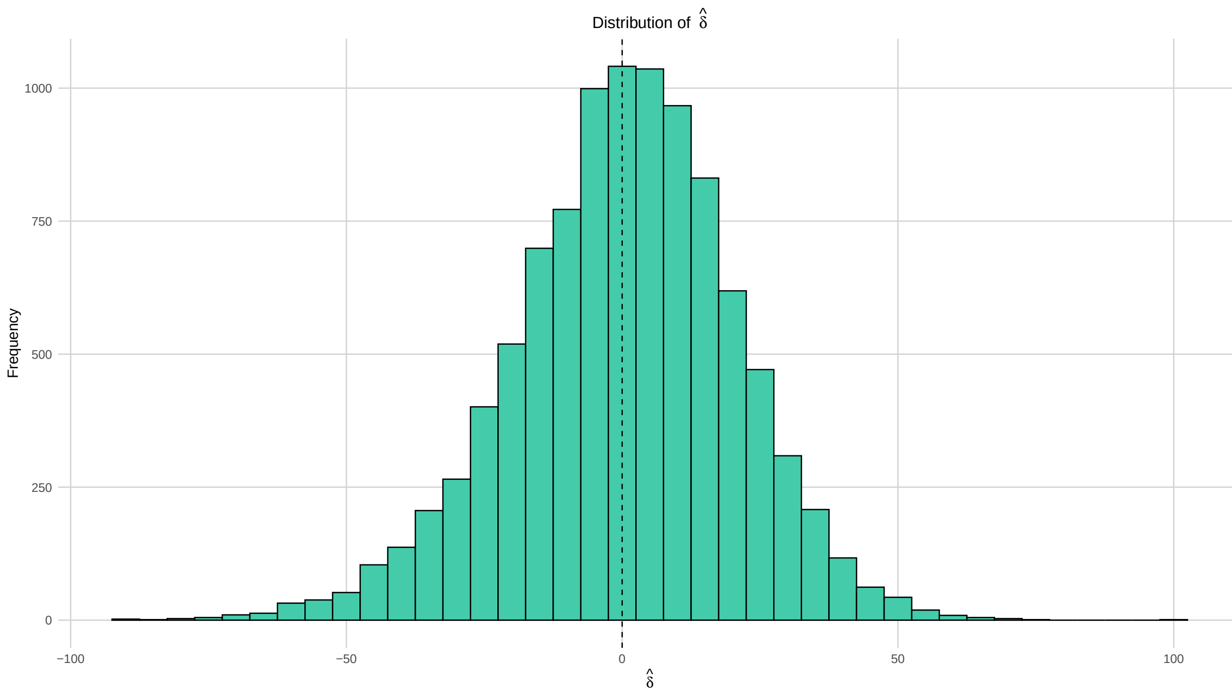


Figure 5 − Sample distribution of the simulated estimators for the local average treatment effect

Table 4: Simulation Results: Two Stage Difference-in-Difference Regression Results of $\hat{\delta}$

	$\overset{iv}{CV}_1$ (2.5%, 97.5%)		$\overset{iv}{CV}_2$ (2.5%, 97.5%)		$\overset{iv}{CV}_3$ (2.5%, 97.5%)		2SLS (2.5%, 97.5%)	
1	-20.24	13.52	-25.83	19.12	-41.90	35.18	-24.45	17.73
2	-20.50	25.35	-22.09	26.94	-24.25	29.10	-16.10	20.95
3	-23.20	9.02	-24.25	10.07	-25.28	11.10	-26.95	12.77
4	-11.07	20.30	-14.11	23.33	-22.51	31.74	-15.92	25.14
5	-10.46	18.29	-12.15	19.97	-19.37	27.19	-12.61	20.43
6	-16.58	31.21	-15.08	29.71	-12.20	26.82	-11.04	25.67
7	-17.97	8.70	-18.80	9.53	-21.25	11.98	-24.55	15.28
8	-16.23	19.49	-20.00	23.26	-31.98	35.24	-22.71	25.96
9	-20.32	13.15	-20.91	13.74	-23.88	16.71	-21.51	14.35
10	-19.66	12.13	-27.07	19.54	-46.94	39.40	-27.31	19.78
...	...	...	...	...	...	...	...	...
9996	-10.38	22.17	-10.49	22.28	-11.08	22.87	-14.57	26.36
9997	-33.48	15.01	-33.73	15.26	-34.04	15.57	-33.51	15.04
9998	-13.50	14.85	-15.36	16.71	-19.84	21.19	-18.88	20.23
9999	-43.52	5.72	-47.46	9.66	-57.51	19.71	-39.13	1.33
10000	-13.76	11.52	-13.63	11.39	-14.23	11.99	-20.46	18.22
Empirical Coverage	0.0553		0.0297		0.0118		0.0465	

VI. Implications and Conclusion

The results from the Monte Carlo simulation imply that using $\overset{iv}{CV}_1$ may be the best cluster-robust variance estimator in just-identified instrumental variables settings. $\overset{iv}{CV}_2$ and $\overset{iv}{CV}_3$ seem to routinely underreject. However, it seems that MacKinnon et. al ia also aware of this unique feature of the jackknife estimator. The authors still favor this estimator over the first two alternatives for other reasons, including computational motives. This paper reaffirms this characteristic about $\overset{iv}{CV}_3$. It may also be possible that the characteristics seen in the analysis are unique to the data-generating process used or the data itself. More research and analysis may prove beneficial on this front. Additionally, this simulation was performed with $G = 64$ (64 unique countries in the analytic sample). It may be possible that the different estimators perform differently under different conditions. Furthermore, when clustering is impractical either via theoretical analysis or empirically infeasible, this study shows that the two-stage least squares estimator (given that assumptions for identification are properly met) seems to work just as fine. Overall, while this paper cannot hope to be the final word of theoretical and empirical work on cluster-robust variance estimators, I do hope this paper has served enforce the credibility of inference employed by influential papers that cluster their standard errors.

Finally, this analysis adds credibility to the empirical findings of the applied example presented in this paper. The results of the Monte Carlo analysis imply that the test used to evaluate whether $\hat{\delta}$ is significant seems appropriately sized. Given that electing higher proportions of women in national legislatures in African and Arab countries truly has no effect on yearly per capita CO_2 emissions, using $\overset{iv}{CV}_1$[18], there is 0.021% chance (p-value < 0.001) that we would observe an effect as large or larger by random. Given these results, we can (confidently) reaffirm the effect found by Mavisakalyan and Tarverdi. While the exact mechanism through which women's presence influences climate policy remains open for further investigation, these findings hold significant policy implications. It seems clear when more women enter the national legislature, we see lower per capita CO_2 emissions.

[18]The 95% confidence interval using $\overset{iv}{CV}_1$ yields (-164.8728, -50.8122), or, on the 1% scale, (-1.65, -5.08).

References

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? https://www.nber.org/system/files/working_papers/w31063/w31063.pdf

Bell, R.M.m McCaffrey, D.F., (2002). Bias reduction in standard errors for linear regression with multi-stage samples. Surv. Methodol. 28, 169-181.

Emissions Database for Global Atmospheric Research (EDGAR). (n.d.). Edgar online. Retrieved from https://edgar.jrc.ec.europa.eu/

European Commission, Joint Research Centre (JRC). (2023). Emissions gap report 2023. Publications Office of the European Union. doi:10.2760/481442 https://edgar.jrc.ec.europa.eu/report_2023

Inter-Parliamentary Union (IPU). (n.d.). Women in national parliaments. [Data set]. Retrieved from https://data.ipu.org/womon-ranking

MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. D. (2023). Cluster-robust inference: A guide to empirical practice. Journal of Econometrics, 47(4), 414–4561

Mavisakalyan, A., & Tarverdi, Y. (2019). Gender and climate change: Do female parliamentarians make a difference? Global Labor Organization (GLO) Discussion Paper No. 221. https://www.sciencedirect.com/science/article/pii/S0176268017304500

Medlicott, E. (2021, March 8). The five devastating reasons climate change affects women more than men. Retrieved from https://www.euronews.com/green/2021/11/09/the-five-devastating-reasons-climate-change-affects-women-more-than-men

Mirziyoyeva, N., & Salahodjaev, D. (2023). Does representation of women in parliament promote economic growth? Considering evidence from Europe and Central Asia. Frontiers in Political Science, 4(2). [DOI: 10.3389/fpos.2023.1203221] https://www.frontiersin.org/articles/10.3389/fpos.2023.1120287/full

Schaeffer, Katherine (2020, October 5). Pew Research Center. Key facts about women's suffrage around the world, a century after U.S. ratified 19th Amendment. Retrieved April 1, 2024, from https://www.pewresearch.org/short-reads/2020/10/05/key-facts-about-womens-suffrage-around-the-world-a-century-after-u-s-ratified-19th-amendment/

Sellers, S. (2016). Gender and climate change: A closer look at existing evidence. Global Gender and Climate Alliance (GGCA). https://wedo.org/gender-and-climate-change-a-closer-look-at-existing-evidence-ggca/

Skaaning, Svend-Erik; John Gerring; Henrikas Bartusevicius, 2015, "A Lexicial Index of Electoral Democracy", https://doi.org/10.7910/DVN/29106, Harvard Dataverse, V6

United Nations Development Programme (UNDP). (2022, March 8). International Women's Day 2022 "Gender equality today for a sustainable tomorrow". Retrieved from https://www.undp.org/speeches/undp-administrators-statement-international-womens-day-2023-8-march

U.S. Department of State. (n.d.). Countries and Areas List. Retrieved from https://www.state.gov/countries-and-areas-list/

The World Bank. (n.d.). World Bank Open Data. Retrieved from https://databank.worldbank.org