Replication of a Research Claim from Fitzgerald et al. (2018), from *Social Forces*

Replication Team: Kent Jason Cheng[1] and Daniel J. Mallinson[2]

1 Department of Social Science, Maxwell School of Citizenship and Public Affairs, Syracuse University, Syracuse, New York, USA

2 School of Public Affairs, Penn State Harrisburg, Middletown, Pennsylvania, USA

Research Scientist: Sam Field

Action Editor: Annette Brown

Project ID: Fitzgerald_SocialForces_2018_4q0L - Cheng/Mallinson - Data Analytic Replication - 3z5z

OSF Project: https://osf.io/m4yse/

Preregistration: https://osf.io/83kca

Claim Summary

The claim selected from Fitzgerald et al. (2018) is that working time is positively associated with higher state-level carbon emissions. This reflects the following statement from the paper's abstract: "Our findings suggest that over the 2007–2013 period, state-level carbon emissions and average working hours have a strong, positive relationship, which holds across a variety of model estimation techniques and net of various political, economic, and demographic drivers of emissions." The authors estimate Prais-Winsten models with panel corrected standard errors (PCSEs) for the fixed effects models. They estimate two-way fixed effects models by including both state-specific and year-specific intercepts. They also correct for first-order autocorrelation (i.e., AR(1) correction) within panels and treat that AR(1) process as common to all panels. The specification of the model containing the selected test result can be gleaned from Table 4, Model 1: Scale(FE). The selected test involves the location of the estimated coefficient in the “Working hours” term. In model 1 (Table 4), the authors find that the scale effect of working hours on carbon emissions is positive and significant. Specifically, in model 1, they find that, over time, a 1 percent increase in average working hours per worker is associated with a 0.668 percent increase in emissions, holding all else constant.

Replication Criteria

A successful replication of Fitzgerald et al.’s (2018) claim would be a positive relationship between average working hours per worker and CO_2 emissions

Replication Result

The average hours worked is positively correlated with state CO_2 emissions in the United States (b = 0.535139, SE = 0.139448, p = 0.000143). This is the All Data Model A 1 percent increase in average working hours per worker is associated with a 0.54 percent increase in emissions, holding all else constant. This result was consistent in the model with only the original data model (b = 0.425472, SE = 0.212429, p = 0.046133) and the added data model (b = 0.924023, SE = 0.461360, p = 0.048205), which split the full data on the years included in the original study and those added for this replication.

Deviations from the Original Study

The original study used real state GDP chained to 2007 dollars, but we used it chained to 2012 dollars. See the next section for one additional deviation in the estimation approach.

Deviations from Pre-registration

One deviation from the pre-registration was necessary. The pre-registration, as well as the original paper, indicated the use of panel corrected standard errors. There was no problem estimating results when using the required test sample for the preregistration. However, when the same code was applied to the full dataset, the models were no longer estimable due to computational singularity. This is often due to linearly dependent covariates. A correlation matrix, however, reveals that correlations between the independent variables were quite reasonable (highest was < 0.5). Removing the year- and state-specific intercepts allowed the model to be estimable, as did changing the panel corrected standard errors to Huber-White sandwich corrected standard errors. Turning back to the sample-based models used in the pre-registration, I re-estimated them first with only removing the year and state intercepts and then

with only changing the standard error correction method. Removing the year and state intercepts had a much larger effect on the estimation of the coefficients, standard errors, and p-values than simply changing from panel-corrected standard errors to Huber-White. Thus, I decided to make that change. Thus, in a deviation from the pre-registration and the original paper, I use Huber-White standard errors instead of panel-corrected standard errors.

Description of Materials Provided

The following materials are publicly available on the OSF project site (https://osf.io/m4yse/):

- The commented preregistration review
    - Fitzgerald_SocialForces_2018_4q0L_3z5z (Cheng_Mallinson) Preregistration.pdf
- Power Analysis
    - POWER_Fitzgerald_SocialForces_2018_4q0L.zip
- Data Files
    - Processed Files
        - compiled.dta is the final compiled dataset, save for the addition of the dependent variable, found in epa.dta, and hhsize (see next section).
        - epa.dta contains the dependent variable
    - Raw Data
        - epa: co2ffc_2017_2.xlsx
        - wrkhrs: File name (each state has its own file from FRED, see specific links above): SMU01000000500000002A.xlxs, SMU02000000500000002A.xlxs, SMU04000000500000002A.xlxs, SMU05000000500000002A.xlxs, SMU06000000500000002A.xlxs,

SMU08000000500000002A.xlxs, SMU09000000500000002A.xlxs,
SMU10000000500000002A.xlxs, SMU11000000500000002A.xlxs,
SMU12000000500000002A.xlxs, SMU13000000500000002A.xlxs,
SMU15000000500000002A.xlxs, SMU16000000500000002A.xlxs,
SMU17000000500000002A.xlxs, SMU18000000500000002A.xlxs,
SMU19000000500000002A.xlxs, SMU20000000500000002A.xlxs,
SMU21000000500000002A.xlxs, SMU22000000500000002A.xlxs,
SMU23000000500000002A.xlxs, SMU24000000500000002A.xlxs,
SMU25000000500000002A.xlxs, SMU26000000500000002A.xlxs,
SMU27000000500000002A.xlxs, SMU28000000500000002A.xlxs,
SMU29000000500000002A.xlxs, SMU30000000500000002A.xlxs,
SMU31000000500000002A.xlxs, SMU32000000500000002A.xlxs,
SMU33000000500000002A.xlxs, SMU34000000500000002A.xlxs,
SMU35000000500000002A.xlxs, SMU36000000500000002A.xlxs,
SMU37000000500000002A.xlxs, SMU38000000500000002A.xlxs,
SMU39000000500000002A.xlxs, SMU40000000500000002A.xlxs,
SMU41000000500000002A.xlxs, SMU42000000500000002A.xlxs,
SMU44000000500000002A.xlxs, SMU45000000500000002A.xlxs,
SMU46000000500000002A.xlxs, SMU47000000500000002A.xlxs,
SMU48000000500000002A.xlxs, SMU49000000500000002A.xlxs,
SMU50000000500000002A.xlxs, SMU51000000500000002A.xlxs,
SMU53000000500000002A.xlxs, SMU54000000500000002A.xlxs,
SMU55000000500000002A.xlxs, SMU56000000500000002A.xlxs

        - laborprod: lpc-by-state-and-region.xlsx
        - pop: use_pop_gdp.xlsx
        - rdgp: use_pop_gdp.xlsx
        - manuf: SAGDP2N__ALL_AREAS_1997_2019.xlsx
        - gdp: SAGDP1__ALL_AREAS_1997_2019.xlsx
        - energy: prod_btu_re_te.xlsx
        - workpop: Bridged-Race Population Estimates 1990-2018.txt
        - emppop: download.xlsx
        - NUMPREC: usa_00006.dat
    - Data preparation scripts (Script)
        - Three .do files are provided to process the component variable .dta files in "Processed files" to create the combined.dta file.
        - The data dictionary from Fitzgerald et al. (2018): Data Dictionary_Fitzgerald_SocialForces_2018_4q0L (Cheng).xlsx
        - The replication analysis script: Fitzgerald 2018 Script_clean v2.R
        - Further instructions for replicating the hhsize variable using IPUMS data: README.txt
    - Analysis script
        - Fitzgerald 2018 Script_clean v2.R is the script for reproducing the results of this report.

Proprietary Data Not in Public Files

The only piece of the replication data that cannot be shared is the proxy for average household size per state downloaded from IPUMS according to this link. Registration is required but access

is granted almost immediately after signing up. The main steps in IPUMS is to first select the sample concerned (tick the box that corresponds to the 2007 to 2016 annual ACS) then click submit sample selection. Click search and search for the variable named NUMPREC. According to this link, IPUMS data is to be cited as follows: Steven Ruggles, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas and Matthew Sobek. IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D010.V10.0.

The analyst must process the zip file according to IPUMS' instructions, and then compute for the state-level average household size. To adjust the mean estimates, use PERWT and STRATA. "PERWT indicates how many persons in the U.S. population are represented by a given person in an IPUMS sample. It is generally a good idea to use PERWT when conducting a person-level analysis of any IPUMS sample." On the other hand, "STRATA is designed for use with CLUSTER in Taylor series linear approximation for correction of complex sample design characteristics. The variable hhsize is computed by taking the survey weight adjusted state-level mean of NUMPREC for each year.

Full Replication Results

	All Data Model (2007-2016)	Original Data Model (2007-2013)	Added Data Model (2014-2016)
Working hours	0.535139*	0.425472*	0.924023*
	(0.139448)	(0.212429)	(0.461360)
	$p = 0.000143$	$p = 0.046133$	$p = 0.048205$
Employed pop. %	0.51*	0.78*	-0.03
	(0.14)	(0.22)	(0.29)
GDP per hour	-0.004	-0.16	0.43*
	(0.069)	(0.11)	(0.20)
Total Population	0.26	0.14	-0.50
	(0.27)	(0.42)	(0.67)
Manufacturing (% of GDP)	-0.02	0.01	-0.04
	(0.02)	(0.03)	(0.05)
Energy production	-0.004	0.03	-0.04
	(0.016)	(0.02)	(0.02)
Average household size	0.17	0.11	-0.04
	(0.20)	(0.29)	(0.38)
Working-age population	0.35	0.96*	1.35*
	(0.25)	(0.41)	(0.61)
Constant	-6.52*	-15.14*	-15.57*
	(2.18)	(3.46)	(6.24)
N	500	350	150
R-Squared	0.999	0.998	0.999

All continuous variables are logged (ln). All models are calculated with AR(1) correction. All models contain unreported year-specific intercepts and un-reported unit-specific intercepts. Huber-White standard errors in parentheses. * $p < 0.05$.

Citation

Fitzgerald, Jared B., Juliet B. Schor, and Andrew K. Jorgenson. 2018. “Working Hours and Carbon Dioxide Emission in the United States, 2007-2013.” *Social Forces* 96(4): 1851-1874.