Report on the simulation(s) performed based on the paper *Emissions, Transmission, and the Environmental Value of Renewable Energy* by Harrison Fell, Daniel T. Kaffine, and Kevin Novan*

Emma Klein, Marina Podgorneac, Siméon Zebina

November 2024

1 Research question

Across the U.S., billions of dollars have been invested in electricity transmission infrastructure over the past decade to transport renewable energy from remote areas to regions with high electricity demand. The motivation for these large investments is that grid congestion during peak electricity demand hours can hinder the process, limiting renewable energy's ability to reduce emissions where it matters most. In other words, there is less offset of pollution from fossil fuel plants in these areas, affecting the level and location of emissions avoided by renewables. Accordingly, electricity transmission infrastructure is crucial not only for enhancing the environmental value of wind energy but also for improving public health by reducing local air pollution. Therefore, the paper's main research question, which our simulation is based on, is to determine and quantify the (negative) effects that transmission congestion has on the environmental benefits of wind energy.

This paper was selected for the simulation exercise for several reasons: its progression from a simple to a more complex data-generating process (DGP), access to rich data, the inclusion of fixed effects and controls, the presence of a categorical variable for logit regression, and its instrumental variable (IV) strategy. To the best of our knowledge, no other papers examine congestion from this perspective. Additionally, its complex analysis offers a valuable opportunity to assess statistical power and model validity as part of a robustness check.

2 Data

The paper's dataset is structured at the hourly level for two U.S. electricity market regions: ERCOT (Electric Reliability Council of Texas) and MISO (Midcontinent Independent System Operator). However, we focus exclusively on ERCOT. It contains over 1,600 variables on electricity generation from various power plants, emissions (SO_2, NO_x, $PM_{2.5}$, CO_2), weather outcomes, and market conditions. The amount of electricity generated by each plant is linked to emissions and environmental damage calculations. The data are disaggregated by county for each generating unit and cover the years 2011 to 2015, with additional controls for seasonal and weekly variations.

*American Economic Journal: Economic Policy 2021, 13(2): 241–272. https://doi.org/10.1257/pol.20190258

3 Research design

The research design implemented aims to capture the effect of wind generation on environmental damages in the ERCOT electricity markets:

$$D_{hdmy} = \beta_1 W_{hdmy} + \beta_2 W_{hdmy} C_{hdmy} + \beta_3 C_{hdmy} + \sum_i \theta_i f_i(X_{hdmy}) + \gamma_{hm} + \eta_{my} + \delta_d + \epsilon_{hdmy} \quad (1)$$

Where:

- D_{hdmy}: Environmental damages for hour h, day d, month m, and year y.
- W_{hdmy}: Hourly wind generation (MWh).
- C_{hdmy}: Congestion indicator (binary).
- X_{hdmy}: Control variables (e.g., load, fuel price).
- Fixed effects control for variations by hour, month, year, and day of the week.
- Standard errors are clustered at the month-year level.

  Key coefficients:

  - β_1: Effect of wind on environmental damages when uncongested (expected negative).
  - β_2: Effect of wind during congestion (ambiguous sign). If $\beta_2 > 0$, wind generation exacerbates environmental damages during congestion; if $\beta_2 < 0$, wind generation reduces environmental damages during congestion, regardless of wind generation.
  - β_3: Effect of congestion alone on environmental damages. If $\beta_3 > 0$, congestion increases environmental damage; if $\beta_3 < 0$, congestion reduces it.

4 Replication and Tests

This part has for goal to reproduce the results of Table 2 from the paper. The table is composed of 5 regressions:

4.1 The 5 models

Model 1: Baseline Model with Minimal Controls "Column (1) is the most parsimonious specification and only includes month-year, hour-month, and day-of-week fixed effects."

Model 2: Adding Detailed Terms (Tot Load and Fuel Ratio) "Coefficient estimates for Wind and Wind Congested are similar in Column (2), which adds linear and quadratic controls for total ERCOT load and fuel price ratios."

Model 3: Introduction of Additional Controls "Column (3) adds linear and quadratic controls for average Texas temperatures as well as wind generation and load in the neighboring SPP market, with key coefficients essentially unchanged."

Model 4: Addition of Zonal Load Controls "Column (4) replaces total ERCOT load with linear and quadratic controls for the zonal loads in the four ERCOT zones."

Model 5: Fully Interacted Model "Column (5) fully interacts all controls and fixed effects from Column (4) with Congested."

4.2 Test for Heteroskedasticity

Here we want to find the limits of the original regressions. First of all, we have to question the presence of Heteroskedasticity in this model. Indeed, we have reasons to think that the presence of temporal data of wind generations in the model can influence the distribution of residuals, and we have to verify if the estimates of the model (wind, segmented) are reliable. We will use the Breusch Pagan test (linear relationship) and the White Test (more general forms of heteroskedasticity):

4.2.1 Breusch Pagan Test

We do observe a 2.2e-16 p value, meaning that we have high probability of heteroskedasticity in the 5 regressions. We do observe the same for every regression, but the t statistic is a little higher for the model 4, meaning that the addition of controls for zonal loads changed the residual variation structure.

4.2.2 White Test

We do observe great heterogeneity among the results... If the five regressions do have heterogeneity, we observe that the implementation of other variables on load and fuel price ratios is reducing heterogeneity. However, the addition of zonal load is creating higher heterogeneity.

4.3 Test for Autocorrelation

We have to verify if the residuals in the model are independent, especially because the model has time-dependent factors with higher probability of autocorrelation. We will use The Durbin Watson (test for first-order autocorrelation) and the Breusch Godfrey Test (tests for higher-order autocorrelation).

4.3.1 Durbin-Watson Test

We do observe an autocorrelation very close to 0, meaning a positive autocorrelation of the residuals. The use of time variables indicates that the observations can be correlated to each other. But we do observe a greater score for mod2 and others, meaning that the addition of controls on ERCOT load and fuel price ratios has reduced autocorrelation, and that the addition of fixed effects in the last regression reduced autocorrelation too.

4.3.2 Breusch-Godfrey Test

We do find the same results as in the previous test concerning the p-value, with more homogeneity across the regressions.

4.4 Test for Specification

We want to check whether the regression is correctly specified (omitted variables, interaction effects) with the Ramsey Test:

4.4.1 Ramsey Test

As usual the test detects problems of specification, meaning that some variables are probably missing. We do observe that the reset stat is growing with the addition of fuel and ERCOT controls, but the addition of zonal loads control is reducing the score...

4.5 Retrodesign

In addition to the tests, we will use a retrodesign analysis for the variable wind in mod1 to test the reliability and validity of the model.

The result of this retrodesign shows us a very high type S error, meaning that there is very little probability that the sign of the effect related to wind is wrong. However, the error of type M shows us that there is a very high probability that the estimated effect is exaggerated... The coefficient effect is thus probably inflated, which is a common problem in econometrics.

Finally, it is difficult to determine which regressions have the most limits because the results are changing across tests. But it seems that these 5 regressions have great limits in heterogeneity and autocorrelation in residuals. These results are probably related with the time parameters. In all cases, these results give us an additional reason to make simulations of the paper and obtain better results.

5 Fake Data Simulation

The goal of the simulation is to evaluate whether generating a synthetic dataset with fewer variables, but preserving the distribution and sample size of the original data allows for the estimation of the effect sizes (three main coefficients) obtained in the paper. When these effect sizes are incorporated, the simulation will assess whether the model achieves sufficient statistical power, ideally at least 0.8. In other words, if the model can still detect the intended effects with a high probability of rejecting the null hypothesis when it is false. The model's assumptions are as follows: congested dummy is endogenous, being influenced by unobserved factors like plant outages, which could bias the estimate of the effect on environmental damages; no reverse causality from environmental damages affecting congestion or wind generation; confounders such as electricity demand are controlled for; errors are clustered at the month-year level to account for potential autocorrelation.

All baseline parameters, along with the means and standard deviations of the variables, are derived from the ERCOT dataset.

We initially constructed a simpler data-generating process (DGP) designed to replicate the first column of Table 2 from the paper, incorporating the beta coefficients from that column but excluding time-fixed effects and time clustering. The resulting statistical power was 0.66667, with a standard deviation of approximately 0.4. Before increasing the complexity of the DGP, we varied the probability of congestion (which corresponds to the treatment probability) but observed no change in the model's power.

Subsequently, we sequentially adjusted the coefficients and found that altering β_3 (the coefficient of the congestion dummy) was the only modification that increased the power to 1. This can be interpreted as congestion on the wind electricity grid alone leads to higher gas emissions from conventional power plants, which compensate for the electricity demand than initially accounted for in the model. To further investigate, we expanded the study period from 5 to 9 years, which naturally increased the number of observations and, in turn, raised the power to 1.

In the second DGP, we introduced the three time-fixed effects as specified in the paper, but this did not affect the power. The third DGP added clustering by month-year, which also resulted in no change to the power. Finally, we included additional controls (load and fuel ratio) from the second column of Table 2, while simultaneously adjusting the effect size to align with the specification in the fifth column, which the authors identified as the preferred model.[1] This modification led to a sharp increase in the power, which reached again 1.

However, it is counterintuitive that in the fifth column, the coefficient β_3 is negative (-239), suggesting that congestion reduces pollutant emissions. In contrast, β_2 is slightly higher (from 8 to 12), indicating that the environmental value of wind decreases more during congestion. These appear to be opposite effects. Even after performing an IV on the congestion dummy, authors found similar effects. From our perspective, this could be attributed to several factors. First, congestion may lead to the use of less-polluting energy sources, reducing emissions, but this effect may not be fully captured by the model. Alternatively, multicollinearity between congestion and wind generation in certain conditions, or a peculiar pattern in the dataset in which congestion periods coincide with an unusual shift in emissions—such as a temporary reduction in the use of higher-emission plants or a switch to renewable sources that doesn't follow the typical behavior expected during congestion—could explain the results.

In actual research, we would try to check for multicollinearity using correlation matrices and VIFs, refine the model by adding relevant variables (if any) that account for the use of other less-polluting energy sources, investigate data anomalies and conduct robustness checks through sensitivity analyses.

Finally, what we have gathered from this simulation is that to increase the statistical power, it is essential to either increase the sample size longitudinally (by extending the study period) or ensure the inclusion of the appropriate coefficients from the optimal model specification, at least before implementing the instrumental variables (IV) strategy to address potential endogeneity.

6 Real Data Simulation: First Steps

Additionally to the fake data simulation, we implemented the first steps towards a real data simulation focused on the IV specification (sixth column of Table 2). The motivation for this project came from the puzzling result that the IV estimates of the parameters on the wind and congestion interaction term are larger than the OLS estimates. Suspecting a Type M error, the authors mention several explanations:

- that Congested is actually not binary;
- that they are picking a LATE rather than the overall average effect.

If the authors performed several tests to assess the robustness of their results, we thought that simulations are interesting for reasons listed hereafter.

6.1 IV replication

We started our simulation by attempting to replicate the estimates of the paper. This was more difficult than expected, and we have reason to suspect that the authors' code is flawed.

6.1.1 IV-LASSO

We were not able to replicate this part of the code on R for reasons listed in the Quarto document. We resorted to running the authors' code on Stata. Surprisingly, this methodology did not

[1] The 5th column of Table 2 in the reference paper allows the controls and fixed effects to vary depending on whether the market is congested, by interaction with the congestion dummy.

allow us to select the same number of instruments as the authors did (the LASSO selected 108 instruments instead of 56).

6.1.2 First Stage: Comparison Between Probit and Logit Models

Like the authors, we started by regressing the interaction term on the LASSO instruments. We then regressed the binary variable segmented on the previously predicted regressor and the LASSO instruments. We found that the authors did not take into account the binary nature of segmented in their first stage regression, and used a linear regression. We compared this choice to two other options, using a Logit and a Probit model. The comparison between the AIC, the BIC and the predictions actually showed that the Logit model was the best fit. We used it for the second stage of our IV.

6.1.3 Second Stage

After implementing this step, we found that the magnitude of our estimates for the three main variables of interest (wind, wind x segmented and segmented) is less important than that of the authors. Provided that out methodology is not flawed, it could provide a solution to the inflation of the IV estimates. Interestingly, our R^2 is smaller than in the paper: the explanatory power of our model is less important.

6.2 First Steps at a Real Data Simulation

6.2.1 segmented as an (ordered) multinomial variables

Following the intuition of the authors that segmented might not be binary, we wanted to create a multinomial version of the variable and change the first stage of the IV from a Logit regression to an Ordered Logit model. We were able to create the new variable, ranging from 1 to 5, based on cutoffs from the values of the average speed. We also did run successfully the Ordered Logit model. However, due to time constraints, we did not manage to run the second stage of the IV. We expected to find no significant effect on estimates, as suggested by the authors' robustness checks.

6.2.2 wind as an exogenous variable

In this simulation, we wanted to address the authors' concern that the instruments explain the variation in congestion in periods prone to a larger loss in the environmental value of wind. To do so, we generated an exogenous version of the variable wind, based on the DGP used in our fake data simulation. Our simulation was however flawed, as proven by the difference both in magnitude and in sign of the estimates from the previous regression. A further analysis could rectify the bugs in the code and draw a conclusion on the hypothesis of the estimates capturing a LATE.

6.3 Conclusion remarks on real data simulation

While working on this simulation, we realized how complicated it was to work on such a big dataset. We really made baby steps, generating segmented but not modifying the interaction term composed on segmented, for example. However, even if we were not able to implement thoroughly our real data simulations, we believe that it was an interesting exercise that shed light on several drawbacks of the authors' methodology.