Effect of Hurricane Harvey on Wages and Employment in Texas

Hunter C. Tyszka

December 13, 2024

1 Introduction

On October 25, 2017, the southern coast of Texas was battered by Hurricane Harvey, the most powerful hurricane to make landfall in the United States since 2004 and the first hurricane to make landfall in Texas since 1970 (NESDIS, 2024). Hurricane Harvey ravaged Texas communities, leaving widespread flooding, record-breaking rainfall, and a path of destruction in the Texas south, particularly in Houston. The estimated damages were well over $100 billion USD, and over 200,000 homes and businesses were flattened. Hurricane Harvey left a massive and tragic mark in many counties in Texas.

In the aftermath of this devastating hurricane, this paper will study two outcomes of interest.

First, whether there were major impacts on wage income in the affected counties immediately after Harvey rolled through the states. I will ask the question whether the affected counties had decreased wage earnings compared to the unaffected counties in the year and a half following the hurricane. I will use a difference in difference approach to examine these impacts, focusing on the outcome variable of quarterly wages per worker in each county.

Secondly, whether there were impacts on the level of employment in the affected counties. I will similarly estimate this effect using a difference in difference approach, focusing on employment as a percentage of population in each county.

This paper is an addition to previous work that has analyzed the impact of hurricanes on economic indicators in North America. Previously, (Strobl, 2011) investigated whether the destruction of property, impacting production, created enduring decreases in growth in American counties affected by hurricanes, after compensation for damages was provided from federal relief. The result was a notable decrease in the growth rate by about 0.45 percentage points compared to an average county-level growth rate. This paper claimed that the largest decreases in growth were a result of wealthy individuals and households vacating the affected counties. The method to derive this conclusion was through a panel fixed effect estimator, adjusted for spatial correlation across economic growth rates.

Another paper, (Belasen, A.R., Polachek, S.W., 2009) investigated the relationship between hurricanes and labour markets at the county level in Florida. This paper used a difference in difference model to examine changes in average quarterly earnings per worker, allowing the researcher to eliminate time-invariant county-specific effects. This method has been adopted in my research into hurricane effects in Texas. The researchers also used data from the Bureau of Labor Statistics' Quarterly Census of Employment and Wages, for wage data at the county level in Florida.

My research will add to this collection of literature by investigating a state which generally has a strict subset of counties not affected by Hurricanes, with some that have tendency to experience them. I will investigate if there is a relationship between per worker wages and employment as a proportion of population, and hurricane exposure.

2 Data

Data for this project have been collected from several sources, to build a dataset that contains observations of economic indicators for each county in Texas from the years 2014 through 2019.

The Quarterly Census of Employment and Wages from the U.S. Bureau of Labor Statistics was used to collect this economic information (U.S. Bureau of Labor Statistics, n.d.). A CSV file for each county in Texas for each year from 2014 to 2019 was compiled into one dataset. The economic variables in this data include quarterly aggregate wages, average weekly wage, quarterly count of establishments, employment levels for the three months in the quarters, and before and after tax values of wages. The data I have elected to use is at the aggregate level for the county, meaning all industry specific information has been consolidated into one value for each variable.

I have appended to this data population values for each county for the years of interest, collected from the U.S. Census Bureau (United States Census Bureau, n.d.). These population values have been used to compute per capita data values for some of the variables in the consolidated CSV dataset. The population figures are also used to create observations of the outcome variable of interest, employment as a percentage of county population.

To determine which Texas counties have been affected by Hurricane Harvey, I consulted the Federal Emergency Management Agency (FEMA, n.d.) page which outlines the counties in Texas that are eligible for individual assistance. This individual assistance program is designed to provide emergency assistance to individuals who have under-insured or uninsured losses and allows them to access resources that are necessary for their survival after a disaster occurs (FEMA, 2024). I have designated counties as 'affected' by the Hurricane, and therefore 'treated' counties, if FEMA has declared individuals within a county to be eligible for individual assistance. While other counties which have not been declared eligible may have some hurricane damages, they have not been affected to the same degree as those counties which have been designated by FEMA. Counties that have been designated by FEMA have been assigned a dummy value of one, to indicate that they are the treated group within this analysis.

I have elected to focus my analysis on the six quarters before Hurricane Harvey, and the six quarters after, including the quarter where Texas was affected. This constrained time frame introduces less opportunity for factors or major shocks independent of Hurricane Harvey to influence the results of my analysis. Given that there are over 250 counties in Texas, and twelve quarters of time periods, I am confident that this refinement still provides a large enough sample size for a thorough investigation without compromising on necessary statistical assumptions, tools and asymptotic theory.

3 Methods

3.1 Hypothesis of Treatment Outcomes

I hypothesize that as a result of the Hurricane, and the broad path of destruction that it brought, the counties affected by the hurricane will have a lower level of quarterly wage per worker, compared to unaffected counties. I speculate that this decline is the result of reduced demand in the affected regions, as individuals and households work to repair their homes and businesses, and a decrease in workers at the higher wage level, as they are more likely to have higher mobility in times of disaster, as claimed by (Strobl, 2011).

In terms of the second outcome of interest, employment as a percentage of county population, I suspect that this value will be lower in affected counties compared to those unaffected. This decline is likely due to the direct destruction of workplaces, leading to a lower level of employment in these counties.

3.2 Matching Methods

There are 254 counties in the state of Texas, of which, 41 were designated by FEMA as eligible for individual assistance. Within the state, economic variables such as employment levels, number of establishments, average wages, vary broadly across counties. As a consequence of this, the characteristics of counties that were affected by Harvey, differ on average with respect to those that were not. In particular, there were more than five times as many unaffected counties than affected ones. This has the potential to add noise to my regressions, which may cloud the effect of the hurricane on the outcomes of interest and skew the coefficients that represent causal inference.

To overcome this issue, I have employed matching within my data analysis to identify unaffected counties that have similar levels of population and number of business establishments as affected counties. These two values are likely to be confounding variables, as population and the number of establishments are indirectly related to wages and levels of employment. Similar to the work in the

Causal Inference for the Brave and True (Facure, 2022), I used a nearest neighbour algorithm to derive a group of untreated counties with relatively similar characteristics to the treated counties.

To adjust for possible issues in standard errors as a result of running a difference in difference regression, on the matched counties, I have presented regression tables that cluster standard errors on matched pairs of counties. This is to account for errors in standard errors that result due to the matching having been done with replacement, as was outlined in (Abadie, Spiess, 2021). I have accomplished this by creating a unique index which identifies when two counties have been matched through the nearest neighbour algorithm, and then clustered standard errors on this index value.

The result of this matching process provides a subset of Texas counties which share similar characteristics before the hurricane impacted the state. This smaller dataset contains 78 unique counties, 41 treated and 37 untreated, as some of the matched untreated counties were matched to multiple treated counties. I believe that this number of sample counties is still high enough to rely on asymptotic theory to derive the regression estimates.

Throughout the results section, I will present my findings using this set of matched counties, with several regression equations for the two outcome variables, in order to analyze the causal effect.

3.3 Difference in Difference Setup and Assumptions

To analyze the effect of Hurricane Harvey on the counties ravaged by the hurricane, I have conducted two difference in difference analyses with the outcomes of interest, quarterly wages per worker and employment as a percentage of county population. Similar to (Belasen, A.R., Polachek, S.W., 2009), the per worker wages allow for the elimination of some of the time invariant and county invariant factors that could result in wages being different across counties.

The difference and difference model fits well for this analysis, because there is a sample of counties that were affected by the hurricane, and a sample of unaffected counties, for quarterly time periods over the span of 2016 to 2018, or 12 time periods. This results in a total of 254 counties, with 41 treated counties and 213 untreated counties. After completing the matching of treated and untreated counties, there will be 41 treated counties compared to 37 untreated counties.

I have also elected to incorporate county and time fixed effects into the regression models. These additional terms will help to account for some of the county and time invariant differences that are present within and between both unaffected and affected counties. The use of these fixed effects will help to better pin down the effects of the hurricane by eliminating factors independent from it, that have an effect on the two outcome variables.

Finally, within the difference in difference regression equations, I have elected to use clustered standard errors at the county match level. I have created a unique index for each county-match pair to account for errors in standard errors that may arise as a result of using nearest neighbour matching. By clustering on this index, I endeavour to remove the issues of incorrect standard errors after matching.

3.4 Difference in Difference Equations

The difference in difference regression equation used to analyze the effect of the hurricane on quarterly wages per worker, without fixed effects is as follows:

$$QtrlyWagePerWorker_i = \alpha + \beta_1 HurricaneCounty_i + \beta_2 Post_t + \beta_3 HurricaneCountyPost_{it} + \epsilon_i$$

$QuaterlyWagePerWorker_i$ represents the outcome of quarterly wage per worker computed at the county level. $HurricaneCounty_i$ is a dummy variable which takes a value of one if the county was affected by the hurricane and a value of zero if it was not. $Post_t$ is a dummy variable which takes the value of one if the date was quarter three 2017, or after, and zero if before. Finally, $HurricaneCountyPost_{it}$

is the interaction of $HurricaneCounty_i$ and $Post_t$, which takes a value of one if both other dummies also have a value of one. The coefficient of interest is thus β_3 which represents the effect of the hurricane on quarterly wages per worker.

Including fixed effects, the regression equation becomes:

$QtrlyWagePerWorker_i = \alpha+\beta_1 HurricaneCounty_i+\beta_2 Post_t+\beta_3 HurricaneCountyPost_{it}+\sum_{i=2016Q1}^{n=2018Q4} \gamma_i Quarter+ \sum_{j=Anderson}^{m=Zavala} \beta_j County + \epsilon_i$

The difference in difference regression equation used to analyze the effect of the hurricane on employment as a percentage of population, without fixed effects is as follows, with the terms defined similarly:

$EmpByPop_i = \alpha + \beta_1 HurricaneCounty_i + \beta_2 Post_t + \beta_3 HurricaneCountyPost_{it} + \epsilon_i$

The fixed effects equation is similar, as follows:

$EmpByPop_i = \alpha+\beta_1 HurricaneCounty_i+\beta_2 Post_t+\beta_3 HurricaneCountyPost_{it}+\sum_{i=2016Q1}^{n=2018Q4} \gamma_i Quarter+ \sum_{j=Anderson}^{m=Zavala} \beta_j County + \epsilon_i$

3.5 A Note on the Parallel Trends Assumption

In order for this difference and difference analysis to be reasonable, I must address the parallel trends hypothesis. I think that the parallel trend assumption is quite reasonable in this case, as a result of the distribution of damages across the states. The southern states of Texas, along the gulf coast were most affected by the storm, and Texas in general had not seen a hurricane in almost fifty years. Two of the United States' most productive and populated urban areas are within Texas, Dallas and Houston. While Houston was affected by the hurricane, Dallas was not. I think that these large regions will help to balance the trends between the two cohorts.

Furthermore, by using the matching process, and balancing the treated and untreated groups on similar characteristics, it is more likely that the parallel trends assumption will hold.

To make this assessment of the parallel trends assumption more rigorous, I have conducted an event study which analyzes the difference in trends between the treated and untreated counties to determine if there is a statistically significant difference in trends before the hurricane occurred. This event study will be presented in the following results section.

4 Results

4.1 Assessment of Parallel Trends Assumption

I have conducted several event studies which analyze whether or not the difference in trends between the treated and untreated Texas counties for both outcomes of interest are statistically different than zero.

This first plot describes the difference between employment as a percentage of county population between the treated and untreated counties.

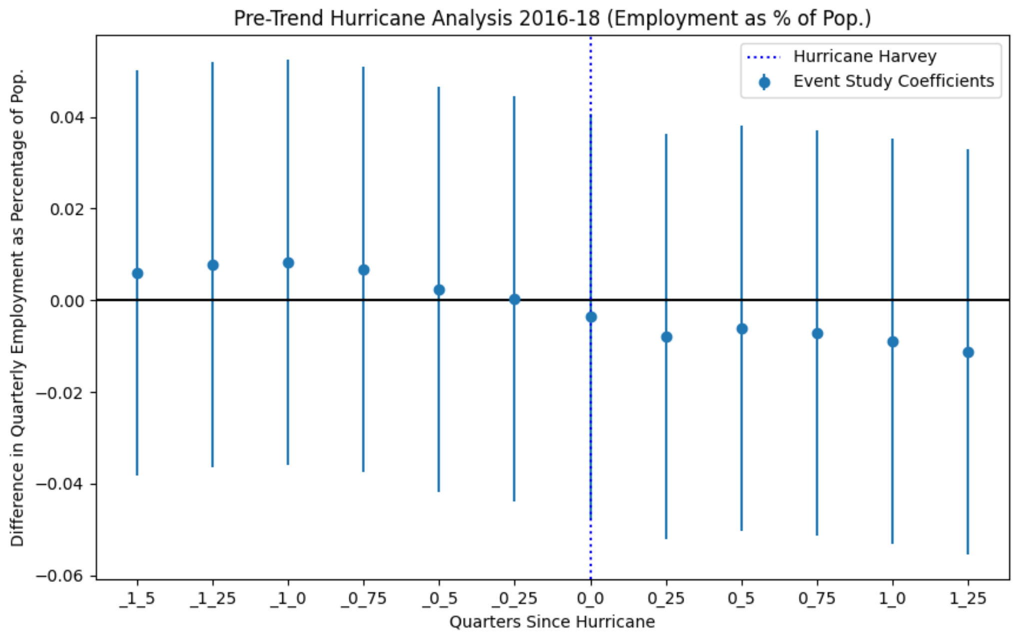

Figure 1: Employment as % of Population

The confidence intervals are between about [-4%,4%], with the estimate of difference around 0.00%. This fluctuates slightly over time around 0.00%. This is a useful indication that the trends are relatively similar before the intervention and that the parallel trends assumption is reasonable for use in this case.

Similarly, for the outcome variable of quarterly wage per worker, the event study plot is as follows:

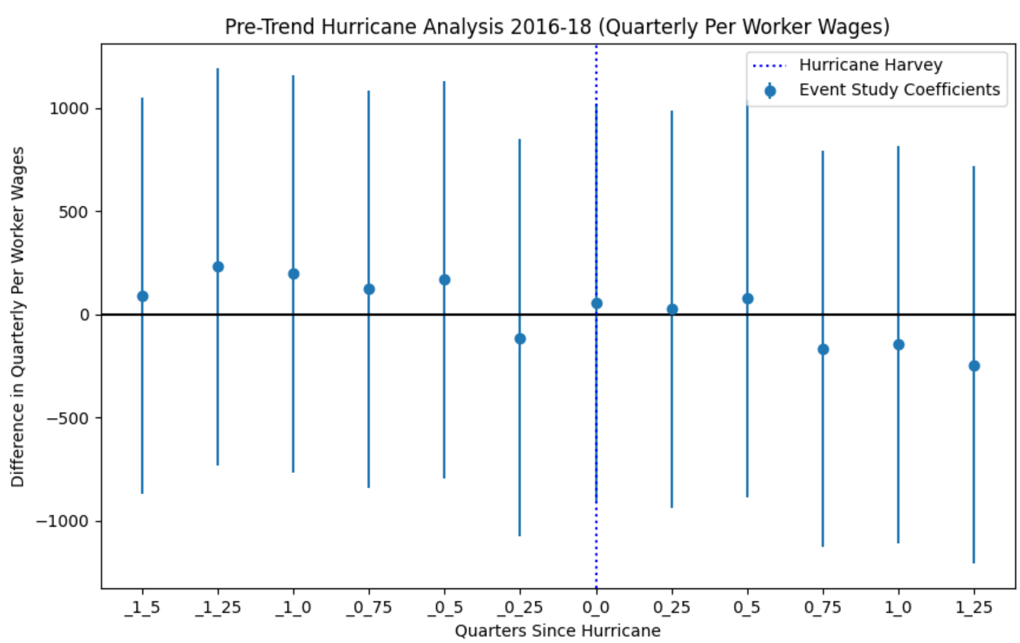

Figure 2: Quarterly Wages Per Worker

For this outcome variable, once again, the estimate values are hovering around zero, with difference estimates fluctuating just above and below over time. The confidence interval for the estimates is [-1,000 , +1,000].

Based on these results, I believe that there is a strong signal that parallel trends is still a reasonable assumption for this case.

4.2 Difference in Difference Regression Analysis

I have conducted difference in difference analysis for both the outcome variables according to the aforementioned regression equations.

4.2.1 Outcome 1: Quarterly Wages per Worker

I have conducted the following difference in difference analysis according to the former regression equation:

$QtrlyWagePerWorker_i = \alpha + \beta_1 HurricaneCounty_i + \beta_2 Post_t + \beta_3 HurricaneCountyPost_{it} + \epsilon_i$.

The standard errors are clustered at the treated-untreated county match level, where each matched pair is assigned a unique index to represent its match. This method improves the accuracy of the estimates as it reduces issues with the standard errors as a result of the matching process.

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
Intercept	1.046e+04	308.809	33.879	0.000	9857.013	1.11e+04
hurricane_counties	410.8028	435.251	0.944	0.345	-442.274	1263.879
post_hurricane	810.7308	140.065	5.788	0.000	536.209	1085.252
hurricane_counties:post_hurricane	-182.7400	139.206	-1.313	0.189	-455.579	90.099

While there is a difference of about -$182.74 USD in the counties affected by the hurricane, the result is not highly statistically significant, as it has a p-value of 0.189. However, the confidence interval is [-455.579, 90.099], which indicates that there is much greater likelihood that there is a negative effect on wages after the hurricane, rather than a positive effect.

The results of the regression with fixed effects are as follows:

$QtrlyWagePerWorker_i = \alpha + \beta_1 HurricaneCounty_i + \beta_2 Post_t + \beta_3 HurricaneCountyPost_{it} + \sum_{i=2016Q1}^{n=2018Q4} \gamma_i Quarter + \sum_{j=Anderson}^{m=Zavala} \beta_j County + \epsilon_i$

Once again, I have clustered standard errors at the county-match level.

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
Intercept	1.503e+04	84.693	177.444	0.000	1.49e+04	1.52e+04
hurricane_counties	-4143.9853	70.940	-58.416	0.000	-4283.024	-4004.946
post_hurricane	390.0455	106.914	3.648	0.000	180.498	599.593
hurricane_counties:post_hurricane	-182.7400	145.517	-1.256	0.209	-467.949	102.469

Now, with the fixed effects, there is still a decline in quarterly average wage per worker by $182.74 with a p-value of 0.209. This could be an indicator that the combination of clustering the standard errors and using time and county fixed effects introduces some collinearity which reduces the quality of the estimates.

The following is a graphical representation of the trends in average quarterly wages per worker in the unaffected and affected Texas counties. I have included a counterfactual line which represents the trend that the affected counties would have taken, had they not been affected by the hurricane. This counterfactual trend depends on the parallel trends assumption, and assumes that these counties would have followed the same trend as the unaffected counties, all else equal.

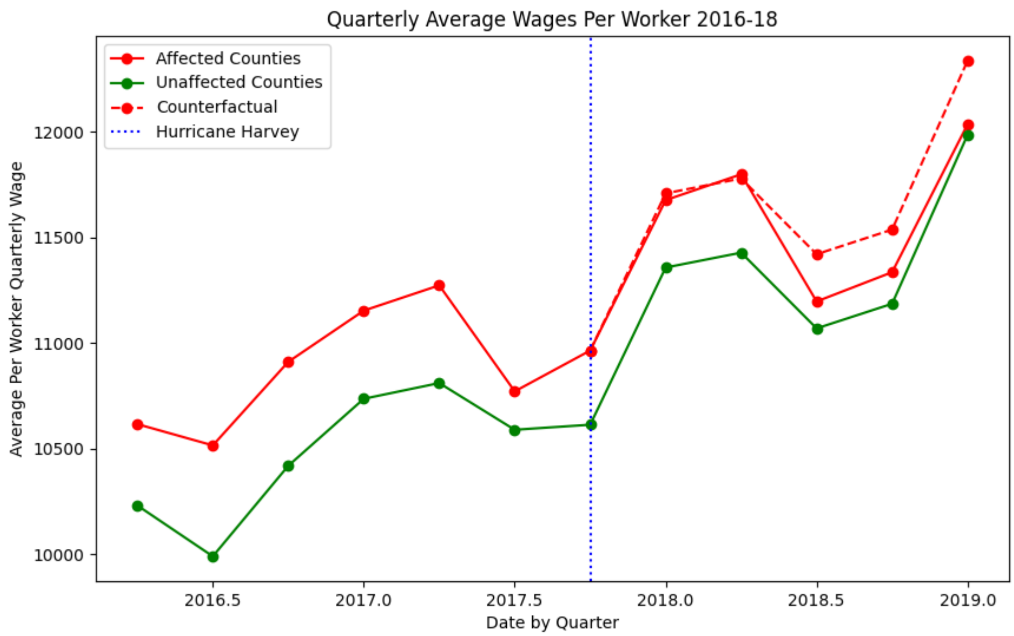

Figure 3: Qtrly Wage Per Worker - Trend

As seen in the figure, immediately after the hurricane, it is not clear if there is any immediate effect on average quarterly wages. Possible reasons for this difference could include: lag in wage effects, especially given that the hurricane was at the end of a quarter, the jobs and workers affected the most are lower earners who do not have as much of a pull on the general trend. In the longer time frame, between six months and two years after the hurricane, the counterfactual line passes the trend line of affected counties. This suggests that there could be long run effects of the hurricane which take a few quarters to take hold and appear in the wage data.

4.2.2 Outcome 2: Employment as a Percentage of County Population

I have conducted the following difference in difference analysis according to the former regression equation:

$$EmpByPop_i = \alpha + \beta_1 HurricaneCounty_i + \beta_2 Post_t + \beta_3 HurricaneCountyPost_{it} + \epsilon_i$$

Once again, the standard errors are clustered at the treated-untreated county match level, where each matched pair is assigned a unique index to represent its match.

	coef	std err	z	P> |z|	[0.025	0.975]
Intercept	0.3088	0.015	20.437	0.000	0.279	0.338
hurricane_counties	-0.0081	0.016	-0.499	0.618	-0.040	0.024
post_hurricane	0.0169	0.007	2.576	0.010	0.004	0.030
hurricane_counties:post_hurricane	-0.0127	0.007	-1.915	0.056	-0.026	0.000

The regression result shows that there is a decline in employment by about 1.27% as a percentage of county population in the counties affected by the hurricane. This result is statistically significant, at the 0.10 level, with a p-value of 0.056, and confidence interval of [-0.026, 0.00]. Hence, it is highly plausible that the hurricane had a negative impact on employment in counties affected by the hurricane.

Next, the results of the regression with fixed effects are as follows:

$$EmpByPop_i = \alpha + \beta_1 HurricaneCounty_i + \beta_2 Post_t + \beta_3 HurricaneCountyPost_{it} + \sum_{i=2016Q1}^{n=2018Q4} \gamma_i Quarter + \sum_{j=Anderson}^{m=Zavala} \beta_j County + \epsilon_i$$

Again, I have clustered standard errors at the county-match level.

	coef	std err	z	P> |z|	[0.025	0.975]
Intercept	0.3949	0.004	105.528	0.000	0.388	0.402
hurricane_counties	-0.0924	0.003	-27.412	0.000	-0.099	-0.086
post_hurricane	0.0109	0.005	2.150	0.032	0.001	0.021
hurricane_counties:post_hurricane	-0.0127	0.007	-1.831	0.067	-0.026	0.001

With the fixed effects incorporated into the regression, there is still an observed decline in employment as a percentage of county population by the same 1.27%. However, this result is less statistically significant, taking on a P-value of 0.067. This could be an indication that the combination of clustering standard errors at the county match level, while introducing fixed effects, introduces some collinearity which reduces the quality of the estimates.

The following is a graphical representation that shows the trends in employment as a percentage of county population for the affected and unaffected counties in Texas. I have included a counterfactual line which shows the trend that the affected counties would have taken, had they not been affected by the hurricane. This once again, relies on the parallel trends assumption.

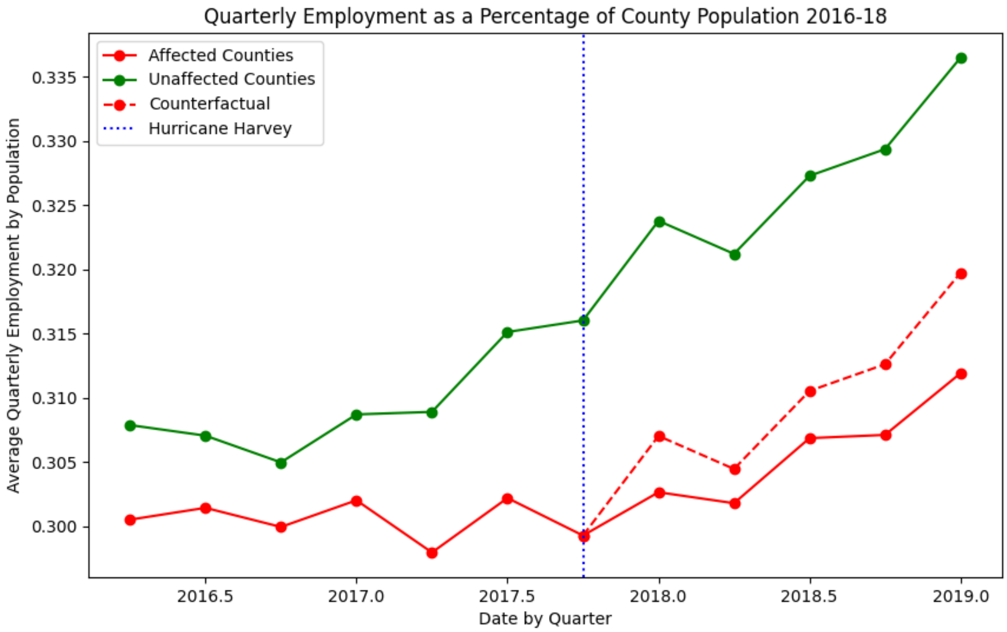

Figure 4: Employment as a Percentage of Population - Trend

In this case, there is a noticeable difference in employment as a proportion of population for the unaffected counties and affected counties. Moreover, it appears that the counterfactual for the treated counties would have been much higher than the actual values, without Harvey, if the parallel trends assumption holds.

5 Conclusion

By using a difference in difference analysis on the two outcome variables, quarterly wages per worker and employment as a percentage of population, I have observed the effect of hurricane Harvey on the affected counties in Texas.

Through the matching method, wherein untreated counties were matched to treated counties on the basis of population and number of business establishments, and through the use of time and county fixed effects, I determined that there was a decline in both outcome variables. I argued and showed, through event study, that the parallel trends assumption is a plausible hypothesis and foundation for this difference in difference analysis. By using standard errors that were clustered on the level of county-matches between treated and untreated groups, I accounted for some issues in standard errors that resulted from matching.

However, although the results did show a decline, their statistical significance was not resounding. This can suggest that there is still a plausible causal effect of the hurricane on wages and employment in the treated counties, compared to untreated counties that had similar characteristics.

6 References

Abadie, A., Spiess, J. (2021). Robust Post-Matching Inference. Journal of American Statistical Association. 117(538), 983-995. https://doi.org/10.1080/01621459.2020.1840383

Belasen, A.R., Polachek, S.W. (2009). How disasters affect local labor markets: the effects of hurricanes in Florida. The Journal of Human Resources, 44(1), 251-276. https://www.jstor.org/stable/20648894

Facure, M.A. (2022) 10-Matching. Causal Inference for the Brave and True. https://matheusfacure.github.io/python-causality-handbook/landing-page.html

FEMA. (n.d.) Designated areas: disaster 4332. Federal Emergency Management Agency. https://www.fema.gov/disaster/4332/designated-areas

FEMA. (2024, October 23). FEMA individuals and households program application eligibility, registration and appeals. Federal Emergency Management Agency. https://www.fema.gov/fact-sheet/fema-individuals-and-households-program-application-eligibility-registration-and-appeals#:~:text=FEMA%27s%20Individuals%20and%20Households%20Program,losses%20caused%20by%20a%20disaster

NESDIS. (2024, August 26). Hurricane Harvey: a look back seven years later. National Environmental Satellite, Data, and Information Service, https://www.nesdis.noaa.gov/news/hurricane-harvey-look-back-seven-years-later

Strobl, E. (2011). The economic growth impact of hurricanes: evidence from U.S. coastal counties. The Review of Economics and Statistics, 93(2), 575-589. https://www.jstor.org/stable/23015955

U.S. Bureau of Labor Statistics. (n.d.) Quarterly Census of Employment and Wages, CSVs By Area, 2014-2019. [Data Set]. U.S. Bureau of Labor Statistics. https://www.bls.gov/cew/downloadable-data-files.htm

United States Census Bureau. (n.d.) County Population Totals: 2010-2019 – Texas [Data set]. United States Census Bureau. https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html