False Positive Brought by Fat-tailed Weights? Re-examine the Impact of China's Retaliatory Tariff on US House Election*

Ye Sun

November 19, 2021

Abstract

I argue that an oftentimes ignored threat to the inference of shift-share regression is that the distribution of the regional size is fat-tailed. This, by my investigation, is likely to cause CLT to fail and the variance estimators, both EHW and AKM (Adão et al., 2019), to lose their power. I first computed the tail index of the population of three common spatial units (commuting zone, county, and MSA). The results suggested a tail index of less than 2 for all choices of spatial unit. I then verified that under the fat-tailed regional weights some assumptions in EHW and AKM are not met. To better showcase the impact of the fat-tailed regional weights, I did a Monte Carlo exercise using real weights that were applied in actual researches, showing that the most impacted aspect of these regressions is the variance estimator. I also proposed four remedies to alleviate the influence of fat-tailedness in the data: (i) using a family of Z estimator with a parameter b that controls the influence of the tail on the regression (ii) rezone the regression with spatial units that are of roughly even sizes (e.g Congressional District) (iii) using longer panel data provided that the regional weight structure stays roughly the same (iv) non-parametric approach to estimate the variance estimator. I applied these methods to Blanchard et al. (2019) and the result suggested that the impact of China's retaliatory tariff on the US House Election might not be statistically significant.

*The code for replicating this paper is available on GitHub at https://github.com/caibengbu/MA_thesis. Any feedbacks are welcomed. I thank Benjamin Soltoff, Elizabeth L Huppert, Christian B Hansen and Qin Wen for their brilliant advice. This project wouldn't have been possible without their kind support.

1 Introduction

Weighted least squares regression (WLS), paired with the a shift-share regression (referred to as SS-WLS throughout the paper), has powered a series of empirical trade literature since Autor et al. (2013). This method constructs a regional level "exposure" variable which maps sectorial shocks to regions, followed by a WLS regression with the regional population as the regression weights (we hereinafter refer to the population of the region that is used to weight the regression as "regression weights" or "regional weights", and the weights used to construct the shift-share variable, usually the employment shares, as "exposure weights"). In the case of Autor et al. (2013), the event they investigate is the surge of Chinese import after China entered WTO in the 2000s. They constructed the commuting-zone (CZ) level trade shock exposure as the key regressor to explain the regional unemployment rate and wage changes. They found out that CZs with high exposure to Chinese import competition faces significantly worsened labor market for US workers. In the case of Blanchard et al. (2019), the event they investigate is the US-China trade war shock in 2018. They constructed the county level trade exposure as the key regressor and used it to explain the House election result in 2018. They found out that counties with high exposure to Chinese retaliatory tariff show significant tendencies of shifting from "red" to "blue" in the 2018 House election.

To better organize this paper, the regression in a classical setting takes the following general form throughout the paper

$$y_i = x_i'\beta + \varepsilon_i \tag{1}$$

i is the index of the smallest spatial unit. N is the total number of spatial units in our population. y_i is the dependent variable and is a scalar. x_i is a P-by-1 vector of regressors, ε_i is the noise and e_i is the residuals. For each spatial unit i, its population size is denoted as λ_i. we mainly look into the simplest form (without control or IV or heterogenous treatment effect) of the AKM regression. It takes the following general form throughout the paper

$$y_i = \beta \sum_{s=1}^{S} w_{is} \mathcal{X}_s + \varepsilon_i \tag{2}$$

I unfold my analysis in two settings: firstly the classical regression model with the noise ε_i being the probabilistic component of the model, and secondly, following Adão et al. (2019) with the sectorial shocks $\mathcal{X}_s$ being the probabilistic component.

SS-WLS is widely used in numerous empirical studies: it dates back to Bartik (1991) and Blanchard and Katz (1992). However, the caveats of the econometric theory behind it start to unfold quite recently. This approach is studied by a growing literature (Adão et al., 2019; Goldsmith-Pinkham et al., 2020; Borusyak et al., 2018). While previous discussions have been focusing on possible endogeneity in the employment shares (Borusyak et al., 2018) and autocorrelation among regions with similar employment profile (Adão et al., 2019), this paper joins the discussion by scrutinizing the possible violation of some WLS assumptions in such practice. In the classical setting, for the WLS estimator to be consistent, asymptotically normal and Eicker-Huber-White's heteroskedastic standard error to converge to a non-random limit, one needs to make sure that the first moment and second moment of the interaction of weights, regressors, and noise are finite. This is required by the Law of Large Numbers (LLN) and Central Limit Theorem (CLT). The regression weights are usually assigned to be the population of the spatial unit: CZ in Autor et al. (2013) and county in Blanchard et al. (2019). It might not be an exhaustive list of the spatial units that were ever used with SS-WLS method but our finding is that both choices lead to highly skewed and fat-tailed distribution of the regression weights, which gives us the reason to suspect whether the finite moment assumptions are violated. This can lead to possible inconsistent point estimates of the coefficient and inconsistent estimation of the standard error of the point estimate. I found out that, with real data, the latter is more pronounced and will cause distortion that is not negligible to the statistical testing. In the setting of Adão et al. (2019), fat-tailed regional weights also jeopardize the validity of Assumption 2 (ii) and (iii)[1]. Even though the use of regional weights was not mentioned throughout the paper, it can be shown that the use of regional weights is numerically equivalent of weighting the exposure weights, the residuals, and the dependent variable by the square root of the regional weights, which is also how they implemented their Stata package `reg_ss`. Intuitively, Assumption 2 (ii) and (iii) of Adão et al. (2019) implies that there is no "dominating" sector on the aggregated country level. Another assumption stated in Appendix. A.1.(ii) requires that each region has "non-negligible" sectors. This would potentially give rise to a contradiction because, for regions that were assigned big regional weights, their "non-negligible" sectors would also be blown up. These "non-negligible" sectors after being amplified by the regional weights can very well likely

[1]Following the indexing of the QJE version of the paper and the QJE version of the Appendix when refering to the content of Adão et al. (2019)

be a nationwide "dominating" sector.

Why do we need to assign regional weights given that it might contaminate the data? The motive of using WLS is generally understood as a way to estimate a weighted average of the heterogeneous treatment effects, as in Autor et al. (2013) and Blanchard et al. (2019). However, without considering IV, as Solon et al. (2015) points out in their section V, WLS does not produce such estimate: it does not "identify a particular average of heterogeneous effects". Also put in Adão et al. (2019) in detail, the OLS point estimate is not targeting the "total treatment effect" τ, but a weighted average of $\{\beta_{is}\}_{i=1,s=1}^{N,S}$ with the weights being $\pi_{is} = w_{is}^2 \mathrm{Var}[\mathcal{X}_s]$. Therefore, it is not very likely to interpret weighting as a way to obtain "total treatment effect". Taking into consideration of IV, Goldsmith-Pinkham et al. (2020) suggests that some point estimates receive negative Rotemberg weights meaning that it is not very likely to have a LATE-like interpretation. However, we can still interpret the regional weights as correcting heteroskedasticity as our last resort. It is plausible because the shift-share variable can be interpreted as "shock per capita", typically for the way shift-share variable is constructed in Autor et al. (2013) and Blanchard et al. (2019):

$$x_i = \sum_{s=1}^{S} \frac{L_{is}}{L_s L_i} \cdot Shock_s$$

The underlying assumption is that the fraction of shock in sector s allocated to region i is equal to the employment ratio of sector s in region i out of the total employment in sector s, meaning that $\frac{L_{is}}{L_s} \cdot Shock_s$ equal to the magnitude of shocks from sector s that region i is exposed to. Further dividing this value by L_i which equals to all the working population would give us "shock per capita". In addition, the dependent variable and the controls are all selected to be a "per capita" or "on average" quantity (e.g. wage rate, support rate of the political party, the share of the black population, etc). Aggregating individual-level data to regional-level data, if there is no misspecification, would result in the noise ε_i having a variance proportional to $1/\lambda_i$ by CLT. However, though looking into the residuals and their correlation with the weights, using $1/\lambda_i$ for heteroskedasticity will severely overcorrect the heteroskedasticity, despite the observed negative correlation between residuals and λ_i.

After discussing the consequences and interpretation of fat-tailed weights, I also found support from a variety of urban economics papers that document and theorize the emergence of fat-tailedness of the population distribution (Gabaix, 1999; Rossi-Hansberg and

Wright, 2007; Gabaix and Ibragimov, 2007). In most countries, the size distribution of cities strikingly fits a power law. Power law probability distribution can sometimes have unbounded moments depending on the value of the exponent (Pareto distribution has an unbounded first moment when the exponent is less than one and has an unbounded second moment when the exponent is less than two). According to Gabaix and Ibragimov (2007), their estimation of the exponent is 1.050, which suggests the unboundedness of the second moment of the city size distribution[2]. They define a city as the metropolitan statistical area (MSA), which is different from what most SS-WLS literature adopts – county and CZ. A different definition of the smallest spatial unit would potentially make the population distribution to be different. Despite that, I still show that with the smallest spatial unit being CZ and county, the second moment is still very likely to be unbounded.

This paper itself does not falsify the main message from all SS-WLS literature because (i) not all SS-WLS literature suffer from fat-tailed weights and (ii) even if the weights are fat-tailed, it can still, even highly probable, produce point estimates and variance estimates that does not deviate a lot from the true targeted value. In addition to Autor et al. (2013), Autor et al. (2014) analyzed exposure to international trade on earnings and employment of U.S. workers using the Annual Employee-Employer File extract from the Master Earnings File of the U.S. Social Security Administration to study longitudinal earnings histories for a randomly selected 1% of workers in the United States. Worker-level evidence is not subject to fat tails in the data but they still found a significant impact from Chinese import competition on American workers. However, other external evidence suggests that we might still need to give it more attention. Chetverikov et al. (2016) uses quantile regression to study the distributional effect of the import competition from China. Quantile regression is known as an effective way to get around the fat-tailedness in the data. It is not sensitive to outliers created by the weights at the tail. The result from Chetverikov et al. (2016) accidentally provides external validity for my concern. Their median regression produces a smaller impact from Chinese import competition and a wider confidence interval (CI) that includes 0. This might result from the fat-tailedness of the weights.

The rest of the paper is organized as the following: I would like to first motivate the analysis by estimating the tail index of the regional weights to get a sense of how fat-tailed the regional weights are. Secondly, I show that in both the classical setting and the

[2]Here we are talking about the distribution of the superpopulation. For more discussion on population and superpopulation please refer to Adão et al. (2019).

shift-share setting, the adoption of weights caused the implausibility of some assumptions. Thirdly, I show, in a Monte Carlo exercise, the consequences of using real regional weights when the data-generating process (DGP) is known. I then propose four possible remedy candidates to counter fat-tailed weights. I first create a family of estimators that are a version Z-estimator proposed by Huber (1967, 1964) with a customized moment function. This family of parameterized Z estimators includes OLS as its special cases and LAD as its limit case. This family of estimators has a parameter b that controls the sensitivity to tails. I show that, in the case of Blanchard et al. (2019), when I gradually tune the parameter b to make the tails less influential, the estimated effect becomes smaller and not significant. I also re-examine Blanchard et al. (2019)'s result with the smallest spatial unit being the congressional district (CD). CDs are designed to have similar demographic features especially the total population. We can see how much the regression result has changed through re-zoning our observations. Lastly, I propose two other ways to get around the fat-tailedness in the data: using information from panel data and using a non-parametric approach to estimate the standard error.

2 Drawbacks of $HC0$ and AKM with fat-tailed weights

In this section, I demonstrate firstly how "fat" the regional weight is for the US for different smallest spatial units (County, CZ, and MSA). This is important in understanding the impact of the fat-tailed weights because different levels of "fatness" might have different effects (i.e. non-existence for the first moment and above, non-existence for the second moment and above, etc). I then demonstrate the violated assumptions in a classical WLS setting and the shift-share setting proposed by Adão et al. (2019). Lastly, I use a minimal Monte Carlo exercise to show possible consequences of using these weights.

2.1 How fat are the regional weights?

In this section, we mainly adopt the "rank-1/2" method developed by Gabaix and Ibragimov (2007) to examine the three choices (Metropolitan Statistical Area, County, Commuting Zone) of possible spatial units in SS-WLS papers (Presented in Table 1, Panel 1). We also observed that by plotting the logarithm of rank against the logarithm of population size (Figure 1, the other two scatter plots are similar and available upon requests), the populations are not strictly subject to Power Law, because if they are, the log(size)-log(rank) plot would have a linear appearance. At least in terms of the full sample, it is not very likely that the populations are subject to Power Law. However, since we are mainly interested

Spatial Unit	MSA	County	CZ
Panel 1: Full observations			
Estimated Tail Index	0.9173	0.6335	0.5583
Standard Error	(0.0662)	(0.0159)	(0.0299)
Number of Units	384	3142	722
Panel 2: 50% quantile and above only			
Estimated Tail Index	1.0652	0.9565	0.9837
Standard Error	(0.1087)	(0.0341)	(0.0732)
Number of Units	192	1751	361
Panel 3: 75% quantile and above only			
Estimated Tail Index	1.2088	1.1159	1.1774
Standard Error	(0.1745)	(0.0563)	(0.1241)
Number of Units	96	785	180

Table 1: Estimation of Tail Index

in the behavior of the tails. We can always truncate the sample to the tails only and do our estimation (Table 1, Panel 2 and 3). According to Figure 1, the truncated estimation has a much better approximation of the tails than the full sample estimation. If we set our cut-off level at the 75% quantile, the estimated tail index is 1.2088 for MSA, 1.1159 for counties, and 1.1774 for CZ. The estimation suggests that these choices of spatial units are all very likely to bring fat-tailedness into projects that adopt SS-WLS. With a tail index $\zeta \in (1, 2)$, one can assert that $\mathrm{E}[\lambda_i] < \infty$ and $\mathrm{E}[\lambda_i^2] \to \infty$. Note that this assertion treats the regional weights as a random variable, and is based on an underlying assumption that there exists a "superpopulation" beyond the entirety of the United States. This only works in the classical WLS setting but not in the shift-share setting where regions and characteristics of regions are treated as deterministic, and the population of interest is fixed. It is unclear how the unboundedness of the moment of the superpopulation would impact the methodology that only cares about the observed population of interest, I demonstrate in Section 2.2.2 that even though the characteristics of regions are being "conditioned on"

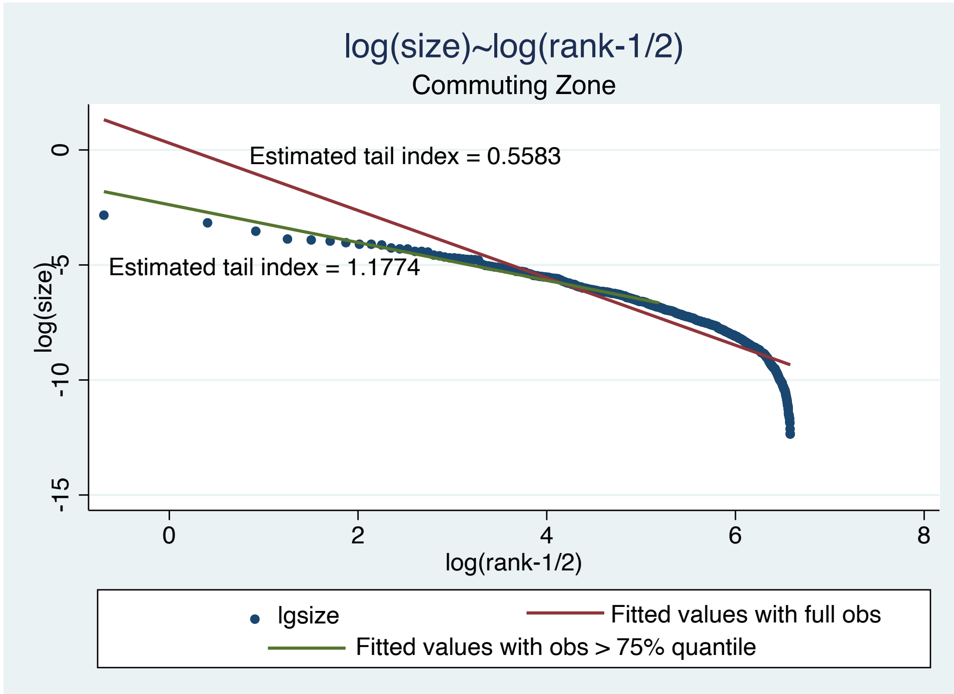

Figure 1: Scatter plot of the populations of commuting zones

throughout Adão et al. (2019), the fat-tailed regional weights still have a similar impact on the AKM standard error.

These estimations of tail index suggest that SS-WLS method might have a consistent, unbiased point estimate but not asymptotic normality and an effective variance estimator. Our following analysis follows this hint from the estimation of the tail index and put emphasis on asymptotic normality and variance estimators.

2.2 Violated Assumptions

2.2.1 Classical setting

For a linear model described as in Equation (1), Econometrics textbooks like Hayashi (2000) assume the existence of the fourth moment for the regressor and residuals. This is a sufficient condition given by Cauchy-Schwarz inequality. In this section, we examine two necessary condition (i) $\mathrm{E}[\lambda_i x_i \varepsilon_i] = 0$ and (ii) $\mathrm{E}[\lambda_i^2 x_{ip}^2 \varepsilon_i^2] < \infty$ for all $p = 1, \cdots, P$.

Condition (i) is essential for $\widehat{\beta}_{WLS}$ to be a consistent estimator and condition (ii) is a necessary condition for $\widehat{\beta}_{WLS}$ to be asymptotically normal and Eicker–Huber–White's variance covariance matrix estimator $\widehat{\Omega}_{HC0}$ to be a consistent estimator for the real variance covariance matrix. Although the unboundedness of $\mathrm{E}[\lambda_i]$ and $\mathrm{E}[\lambda_i^2]$ does not directly leads to the unboundedness of $\mathrm{E}[\lambda_i x_{ip} \varepsilon_i]$ and $\mathrm{E}[\lambda_i^2 x_{ip}^2 \varepsilon_i^2]$, this imposes much tighter constraints for x_i and ε_i in order for (i) and (ii) to hold.

To verify whether an expectation is finite, I use randomized cumulative average[3] to test if the expectation is convergent. By LLN, for a random variable z such that $\mathrm{E}|z| < \infty$, the cumulative average of the iid series $\{z_i\}$ will converge to $\mathrm{E}[z]$ almost surely (in this case, convergence should be irrelevant to the order of the series. No matter how I randomize the series it will always show convergence). I focus on the point estimate and the statistical inference of the intercept (meaning that $x_{ip} = 1$ for all i). I use the residuals from Blanchard et al. (2019) as a proxy for $\{\varepsilon_i\}$ and the 2016 US county population gathered from American Community Survey (ACS) as the weights $\{\lambda_i\}$. I plot the randomized cumulative average of $\{\lambda_i x_{ip} \varepsilon_i\}$ and $\{\lambda_i^2 x_{ip}^2 \varepsilon_i^2\}$ in Figure 2 and Figure 3. Figure 2 converges fast where Figure 3 remains stochastic even when N is large[4]. This implies that $\mathrm{E}[\lambda_i^2 x_{ip}^2 \varepsilon_i^2]$ might be infinite. Therefore, we can make a conjecture that (i) $\widehat{\beta}_{WLS}$ is not asymptotically normal and (ii) $\widehat{\Omega}_{HC0}$ is not a consistent estimator for the variance of $\widehat{\beta}_{WLS}$.

2.2.2 Shift-share setting

In this section, we explain in a minimal example how weights interact with other variables in a shift-share setting. Using the introductory example from Adão et al. (2019) with no controls (Y is a $N \times 1$ vector. β is a scalar. W is a $N \times S$ exposure weight matrix. $\mathcal{X}$ is a $S \times 1$ vector that represents shocks. ε is a $N \times 1$ vector that signifies the untreated state, Λ is a diagonal matrix with the regional weights being on the diagonal), no heterogeneous treatment effect, the same linear model, and the same data generating process. The estimated model can be written as:

$$\sqrt{\Lambda} Y = \beta \sqrt{\Lambda} W \mathcal{X} + \sqrt{\Lambda} \varepsilon$$

[3] The cumulative average of a series $\{z_i\}_{i=1}^N$ is defined as $\{\mathbb{E}_n[z_i]\}_{n=1}^N$

[4] This evidence also eliminates the possibility of interpreting regional weights as correcting heteroskedasticity. If it is true that $\mathrm{Var}[\varepsilon_i] \propto \frac{1}{\lambda_i}$, Figure 3 should converge, too.

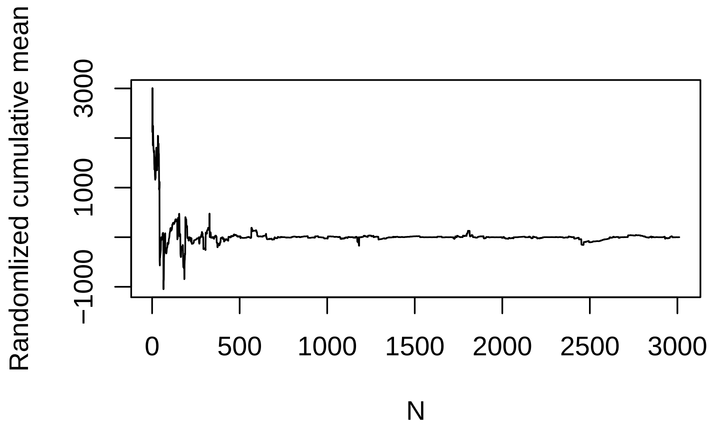

Figure 2: Cumulative average of randomized $\lambda_i x_{ip} e_i$

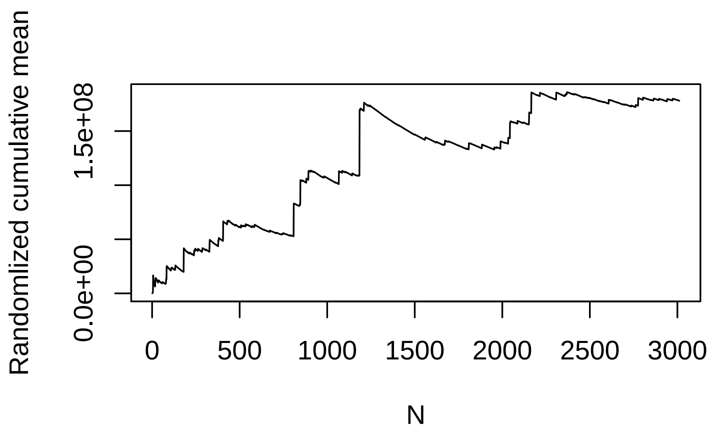

Figure 3: Cumulative average of randomized $\lambda_i^2 x_{ip}^2 e_i^2$

where $\sqrt{\Lambda}$ is a diagonal matrix with the square root of regional weights being on the diagonal. The WLS point estimate is:

$$\begin{aligned} \hat{\beta}_{WLS} &= \frac{\mathcal{X}'W'\Lambda Y}{\mathcal{X}'W'\Lambda W\mathcal{X}} \\ &= \beta + \frac{\mathcal{X}'W'\Lambda\varepsilon}{\mathcal{X}'W'\Lambda W\mathcal{X}} \end{aligned}$$

The analytical variance is

$$\begin{aligned}\mathrm{Var}[\hat{\beta}_{WLS}|\mathscr{F}_0] &= \mathrm{Var}[(\mathcal{X}'W'\Lambda W\mathcal{X})^{-1}\mathcal{X}'W'\Lambda\varepsilon|\mathscr{F}_0] \\ &= \frac{\varepsilon'\Lambda W \mathrm{Var}[\mathcal{X}|\mathscr{F}_0]W'\Lambda\varepsilon}{(\mathcal{X}'W'\Lambda'W\mathcal{X})^2}\end{aligned}$$

where $\mathscr{F}_0 = (\varepsilon, W, \Lambda)$. And it can approximated by

$$\widehat{V_{AKM}} = \frac{\hat{\epsilon}'\Lambda W \mathrm{diag}(\mathcal{X}\mathcal{X}')W'\Lambda\hat{\epsilon}}{(\mathcal{X}'W'\Lambda'W\mathcal{X})^2}$$

In order for $\hat{\epsilon}'\Lambda W \mathrm{diag}(\mathcal{X}\mathcal{X}')W'\Lambda\hat{\epsilon}$ to converge to $\varepsilon'\Lambda W \mathrm{Var}[\mathcal{X}|\mathscr{F}_0]W'\Lambda\varepsilon$, we need to do a little modification to Assumption 2 in Adão et al. (2019)

$$\max_s \frac{\sum_{i=1}^N w_{is}\lambda_i}{\sum_{t=1}^S \sum_{i=1}^N w_{it}\lambda_i} \to 0, \quad \max_s \frac{\left(\sum_{i=1}^N w_{is}\lambda_i\right)^2}{\sum_{t=1}^S \left(\sum_{i=1}^N w_{it}\lambda_i\right)^2} \to 0$$

Intuitively, the original Assumption 2 (ii) and (iii) imply that there is no “dominating” sector on the aggregated country level. The modified version implies that, after weighing the regions, there is no “dominating” sector on the aggregated country level. At the same time, in order for $\mathcal{X}'W'\Lambda'W\mathcal{X}$ to be bounded away from zero, the assumption A.1.(ii) in their Appendix suggests that each region has “non-negligible” sectors. This would potentially give rise to a contradiction because, for regions that were assigned big regional weights, their “non-negligible” sectors would also be blown up. These “non-negligible” sectors after being amplified by the regional weights can very well likely be a nationwide “dominating” sector. By looking into the real ADH data[5], we found out that in finite sample[6]

$$\max_s \frac{\sum_{i=1}^N w_{is}\lambda_i}{\sum_{t=1}^S \sum_{i=1}^N w_{it}\lambda_i} = 0.0640, \quad \max_s \frac{\left(\sum_{i=1}^N w_{is}\lambda_i\right)^2}{\sum_{t=1}^S \left(\sum_{i=1}^N w_{it}\lambda_i\right)^2} = 0.3676$$

For a value that is ranged between $[0, 1]$, 0.3676 is generally not considered a negligible quantity, suggesting that our concern is not an unrealistic scenario. Moreover, for

[5] Only for the year 2000. I focus on the cross-sectional scenario first. The result in this section is not contradictory to AKM's re-examination of ADH because ADH uses panel data. I also claim that using panel data can alleviate the influence of fat-tailedness. See Section 3.3

[6] On the contrary, in the unweighted case, $\max_s \frac{\sum_{i=1}^N w_{is}}{\sum_{t=1}^S \sum_{i=1}^N w_{it}} = 0.0328$, and $\max_s \frac{\left(\sum_{i=1}^N w_{is}\right)^2}{\sum_{t=1}^S \left(\sum_{i=1}^N w_{it}\right)^2} = 0.1111$

data $\{\sum_{i=1}^{N} w_{is}\lambda_i\}_{i=1}^{N}$, the top 1 percentile takes up 21.38% of the total, and for data $\{\sum_{i=1}^{N} w_{is}^2\lambda_i^2\}_{i=1}^{N}$, the top 1 percentile takes up 67.01% of the total.

2.3 Monte Carlo

I presented in the previous two sections the evidence for potential violation of the assumptions, but we are still agnostic of the real impact of a regression. To further illustrate, I use a Monte Carlo approach.

2.3.1 Classical setting

For the classical setting, I set the number of regressors P to 1 for simplicity, and the only regressor we include is the intercept. The true model can be expressed as

$$y_i = \beta + \varepsilon_i$$

The Monte Carlo exercise follows the step describes in the following

1. Take the real county population data as the weights λ_i, draw fake weights $\lambda_i^{compare}$ generated from $U(0,1)$ for comparison.

2. Draw ε_i independently from $N(0,1)$. Set $\beta = 1$, compute y_i using the true model.

3. Run WLS with real weights λ_i, obtain residuals e_i, compute $\hat{\Omega}_{HC0} = \frac{1}{n}\sum_{i=1}^{n} \lambda_i^2 e_i^2$

4. Run WLS with fake weights $\lambda_i^{compare}$, obtain residuals $e_i^{compare}$ and can compute $\hat{\Omega}_{HC0}^{compare} = \frac{1}{n}\sum_{i=1}^{n}(\lambda_i^{compare})^2(e_i^{compare})^2$

5. Repeat step 2-4 10,000 times, compute $\Omega = \frac{1}{n}\sum_{i=1}^{n} \lambda_i^2 1^2$, $\Omega^{compare} = \frac{1}{n}\sum_{i=1}^{n}(\lambda_i^{compare})^2 1^2$ and compare the distribution of $\hat{\Omega}_{HC0}/\Omega$ and $\hat{\Omega}_{HC0}^{compare}/\Omega^{compare}$.

I plotted the histogram of $\hat{\Omega}_{HC0}/\Omega$ and $\hat{\Omega}_{HC0}^{compare}/\Omega^{compare}$ in Figure 4 and 5. The distribution of $\hat{\Omega}_{HC0}^{compare}/\Omega^{compare}$ is asymptotically normal, is centered at 1 (sample mean = 0.9997), and has a small variance (sample variance = 0.0012). It can be observed that even though the weights are randomly generated and therefore "misspecified", Eicker–Huber–White's variance covariance matrix estimator is still a sharp estimator of Ω. The result echoes with the findings discussed in Romano and Wolf (2017). On the contrary, the distribution of $\hat{\Omega}_{HC0}/\Omega$ is left-skewed, deviated from 1 (sample mean = 0.9722) and has a relatively big variance (sample variance = 0.1637). Out of 10000 simulations, the largest $\hat{\Omega}_{HC0}$ can be 5.4981 times of the real Ω. the smallest $\hat{\Omega}_{HC0}$ is 0.3511 times of the real Ω. Comparing to the simulation with fake weights, the most overstated $\hat{\Omega}_{HC0}^{compare}$ is 1.1475 times of the $\Omega^{compare}$, and the most understated $\hat{\Omega}_{HC0}^{compare}$ is 0.8689 times of $\Omega^{compare}$. To show how sta-

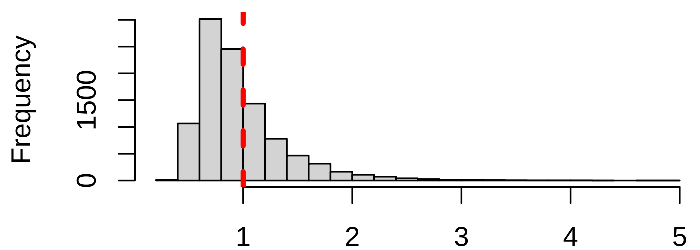

Figure 4: Histogram of $\hat{\Omega}_{HC0}/\Omega$, the true value equals to 1 (red dashed line)

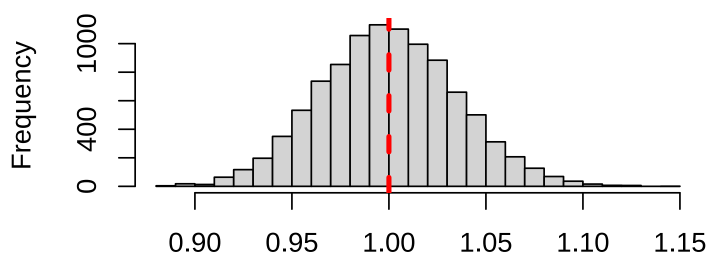

Figure 5: Histogram of $\hat{\Omega}_{HC0}^{compare}/\Omega^{compare}$, the true value equals to 1 (red dashed line)

tistical test loses its power, I use a new approach to compute rate of Type I error and use it as our indicator. My null hypothesis is that $\hat{\beta}$ equals to its true value β. The real distribution of our point estimate $\hat{\beta}$ is $N(\beta, \Omega)$, and the inferred distribution of $\hat{\beta}$ is $N(\beta, \hat{\Omega}_{HC0})$. For any realized $\hat{\beta}$ that lies in $(-\infty, \beta + \sqrt{\hat{\Omega}_{HC0}} \cdot \Phi^{-1}(0.025)) \cup (\beta + \sqrt{\hat{\Omega}_{HC0}} \cdot \Phi^{-1}(0.975), +\infty)$, it will lead to Type I error. Since we also know the real analytical distribution of $\hat{\beta}$, we are able to compute the probability of $\hat{\beta}$ lying in such region.

$$Type\_I(\hat{\Omega}_{HC0}) = 2 \times \Phi\left(\sqrt{\frac{\hat{\Omega}_{HC0}}{\Omega}}\Phi^{-1}(0.025)\right)$$

This measure serves better for our purpose than the conventional way to get the rate of Type I error: for every single simulation, compute the confidence interval (CI), decide rejection based on whether β_0 is in the CI, and finally find the mean of the binary rejection results as the rate of Type I error. The output of the conventional way is a point estimation of the rate of Type I error. Our approach produces a vector of the rates of Type I error, which allows us to see for one simulation the deviation of the rate of Type I error from the real rate of Type I error (which is set to 0.05).

It turns out that the mean rate of Type I error using $\hat{\Omega}_{HC0}$ is 0.0712, 42.4% higher than the confidence level $\alpha = 0.05$. Not only the size of the test is distorted in general, but the test is also inefficient due to the inefficiency of $\hat{\Omega}_{HC0}$. The histogram of $Type\_I(\hat{\Omega}_{HC0})$ is plotted in Figure 6. For every single simulation, the size of the test can deviate considerably from the confidence level. As a comparison, the regression with uniformly distributed weights (Figure 7) has a much convergent and normal distribution.

It can also be shown that $\hat{\Omega}_{HC0}$ under fat-tailed weights is not only not a consistent estimator for Ω but also a downwardly biased estimator in finite sample. For more explanation please refer to the appendix.

2.3.2 Shift-share setting

For the shift-share setting, I ignore controls for simplicity. We also assume homogeneous β across sectors and regions. In this section, we use the data from ADH collected by Adão et al. (2019). The true model can be expressed as

$$y_i = \beta \sum_s w_{is} \mathcal{X}_s + \varepsilon_i$$

The Monte Carlo exercise follows the step describes in the following

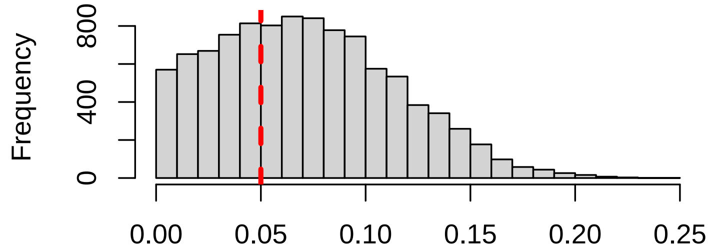

Figure 6: The distribution of $Type\_I(\hat{\Omega}_{HC0})$, the true value equals to 0.05 (red dashed line)

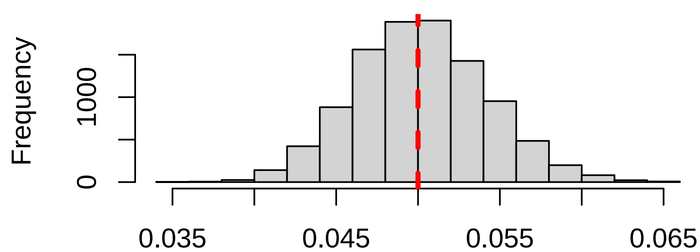

Figure 7: The distribution of $Type\_I(\hat{\Omega}_{HC0})$ for thin-tailed fake weights, the true value equals to 0.05 (red dashed line)

1. Draw ε_i independently from $N(0,1)$. Set $\beta = 1$. Use the real exposure weights for w_{is}. Use these values throughout the simulation.

2. Take the real county population data as the weights λ_i, draw fake weights $\lambda_i^{compare}$ is set to 1 for comparison.

3. Draw $\mathcal{X}_s$ independently from $N(0,1)$. Compute y_i using the true model.

4. Run WLS with real weights λ_i, obtain residuals e_i, compute the numerator of $\widehat{V_{AKM}}$: $e'\Lambda W \text{diag}(\mathcal{X}\mathcal{X}')W'\Lambda e$ denoted as $\widehat{\Omega_{AKM}}$

5. Run WLS with fake weights $\lambda_i^{compare}$, obtain residuals $e_i^{compare}$ and can compute $\widehat{\Omega_{AKM}^{compare}} = (e^{compare})'\Lambda^{compare} W \text{diag}(\mathcal{X}\mathcal{X}')W'\Lambda^{compare} e^{compare}$

6. Repeat step 2-4 10,000 times, compute real $\Omega = \varepsilon'\Lambda W W'\Lambda\varepsilon$, $\Omega^{compare} = \varepsilon'\Lambda^{compare} W W'\Lambda^{compare}\varepsilon$ and compare the distribution of $\widehat{\Omega_{AKM}}/\Omega$ and $\widehat{\Omega_{AKM}^{compare}}/\Omega^{compare}$.

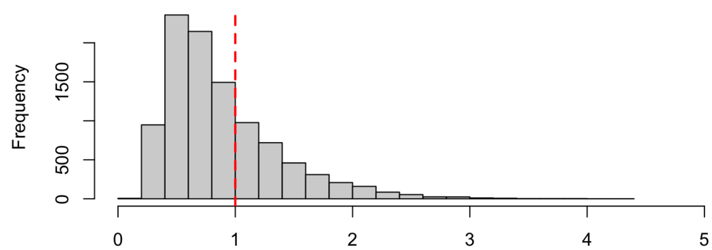

Figure 8: Histogram of $\hat{\Omega}_{AKM}/\Omega$, the true value equals to 1 (red dashed line)

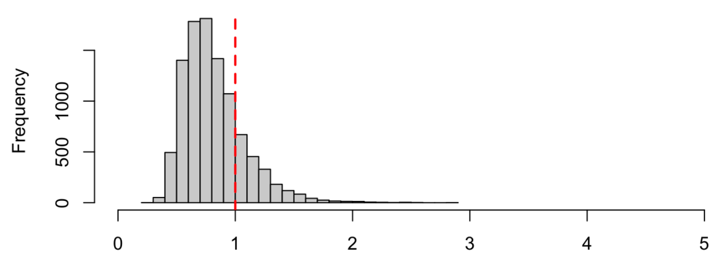

Figure 9: Histogram of $\hat{\Omega}_{AKM}^{compare}/\Omega^{compare}$, the true value equals to 1 (red dashed line)

The histogram of $\hat{\Omega}_{AKM}^{compare}/\Omega^{compare}$ and $\hat{\Omega}_{AKM}/\Omega$ is plotted in Figure 8 and 9. We can observe the same divergence for the histogram that uses the real regional weights. In the unweighted AKM case, we didn't get a normal distribution either, because the exposure weights are already fat in tail (Recall that $\max_s \frac{\left(\sum_{i=1}^N w_{is}\right)^2}{\sum_{t=1}^S \left(\sum_{i=1}^N w_{it}\right)^2} = 0.1111 > 0$). The histogram of the unweighted case still shows a shape that resembles a Chi-squared distribution instead of a normal distribution, suggesting the existence of a fat tail and the failure of CLT. However, this doesn't directly invalidate Adão et al. (2019), the reason are the following: (i) since the variance of the shocks in not known and also takes an important role in whether the variance estimator converges (i.e. it might be the case that the dominant sector receive a shock of a less variance). We are generally agnostic of the variance of the real shock, while in our simulation, we assume homoskedastic shocks. This can be the reason why in the real application the inference method can still be valid. (ii) In the ADH model that AKM is mainly investigating, they used a panel data. Using panel data is a way to alleviate the influence of fat-tailedness (See my discussion in Section 3.3).

3 Remedies for Fat-tailed Weights

In this section, we propose four candidates to counter the effect of fat-tails. Our first suggestion is to adopt Least Absolute Deviation (LAD) and other Z estimators that emphasize less on the tail behavior. The second suggestion we propose is rezoning our regression so that the regional weights are not fat-tailed. Our third suggestion is to exploit the information on panel data to provide a better variance-covariance matrix estimator. Lastly, we suggest that non-parametric approach can be helpful when we assume continuity of $\mathrm{E}[\varepsilon_i x_i x_i' | \lambda_i]$ in λ_i.

3.1 Z estimators

Under the one-dimensional classical setting, I develop a family of parameterized Z estimators which includes OLS as its special cases and LAD as its limit case. This family of estimator has a parameter b that controls the sensitivity to tails. According to Huber (1967, 1964), the Z-estimator takes the following form.

$$\hat{\theta} = \text{the solution to } \frac{1}{N}\sum_{i=1}^{N} \Psi(y_i, x_i, \theta) = 0$$

For example, one of the most popular members of Z-estimators OLS has a moment function Ψ that takes the form of the interaction of the regressor and residual, and Ψ of LAD takes the form of the interaction of the regressor sign of residuals. They all take the form of

$$\Psi(y_i, x_i, \theta) = x_i K(y_i - \theta x_i)$$

For example, $K_{OLS}(\epsilon) = \epsilon$, and $K_{LAD}(\epsilon) = \text{sgn}(\epsilon)$. Our family of estimator is constructed as a family of Z estimators whose moment function can be written in a form of $x_i K(y_i - \theta x_i)$ where $K(\cdot)$ is a differentiable function that takes the form of

$$K(\epsilon) = \text{sgn}(\epsilon)[(a + |\epsilon|)^b - a^b], \quad a, b \in (0, +\infty)$$

When the value of b is set to 1, it is reduced to OLS. When the value of b is set to 0, it is reduced to LAD. In practice, we choose the value of b first, then decide the value of a given b and the fact that K is continuous and differentiable. Since we require our K function to be differentiable, LAD is not part of the family, because no matter what a we choose we cannot have a continuous and differentiable K given $b = 0$. With regional weights, the definition is similar except that all x_i are substituted with $\sqrt{\lambda_i} x_i$ and so are y_i and ε_i. A plot of our K function can be found in Figure 10. It can be seen in the figure that the bigger the value of b takes, the more emphasis the model puts on the tails. In order for an member of the estimator family with a parameter $b \in (0, 1)$ to be consistent and unbiased, we need the following assumption

Assumption (i) $\varepsilon_i | \lambda_i, x_i$ is symmetrically distributed and centered at 0.

Assumption (ii) OLS/WLS estimator is consistent.

Assumption (i) can be verified by investigating the residuals after regression and assumption. Since our family of estimator is a subset of Z estimator, following the property of Z estimator, the estimator is asymptotic normal

$$\sqrt{N}(\hat{\theta} - \beta) \sim N(0, A^{-1}(y_i, x_i, \lambda_i, \beta) B(y_i, x_i, \lambda_i, \beta) A^{-1}(y_i, x_i, \lambda_i, \beta)) \tag{3}$$

where $A(y_i, x_i, \lambda_i, \beta)$ and $B(y_i, x_i, \lambda_i, \beta)$ can be approximated by

$$\begin{aligned} A(\widehat{y_i, x_i, \lambda_i}, \beta) &= -\frac{1}{N} \sum_{i=1}^{N} \frac{\partial \Psi(y_i, x_i, \lambda_i, \theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}} \\ B(\widehat{y_i, x_i, \lambda_i}, \beta) &= \frac{1}{N} \sum_{i=1}^{N} \Psi(y_i, x_i, \lambda_i, \hat{\theta})^2 \end{aligned}$$

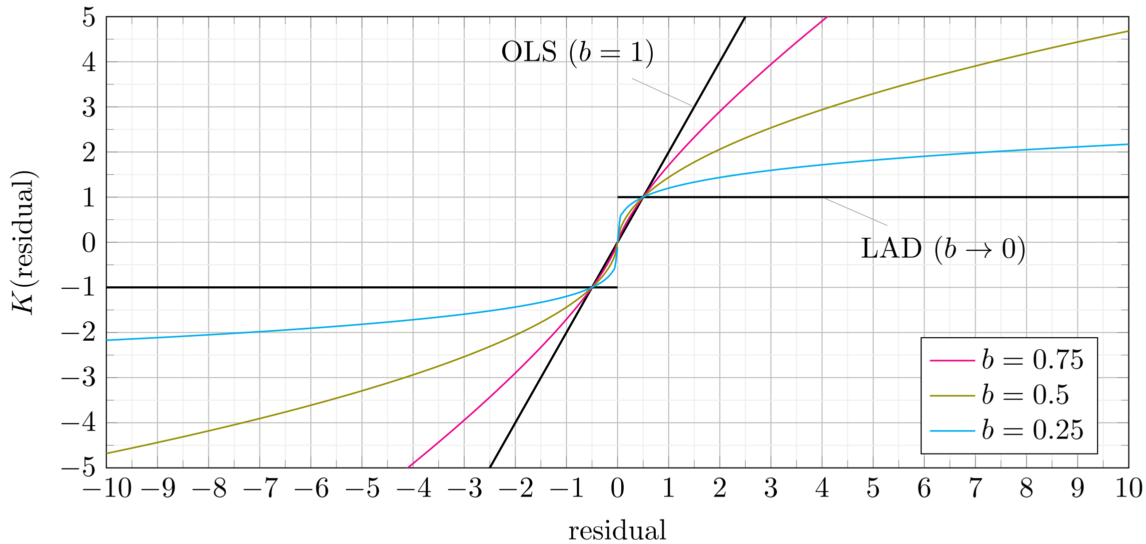

Figure 10: K function with different choice of b

if $\mathrm{E}[\Psi(y_i, x_i, \lambda_i, \beta)^2] < \infty$ and $\mathrm{E}\left[\frac{\partial \Psi(y_i,x_i,\lambda_i,\theta)}{\partial \theta}\Big|_{\theta=\beta}\right]$ is bounded away from 0.

Assumption (iii) OLS/WLS estimator is asymptotically normal, and EHW variance-covariance matrix is a consistent estimator for the real variance-covariance matrix

If Assumption (iii) is satisfied, any member from the family with $b \in (0, 1)$ is also asymptotically normal and the asymptotic variance can be approximated by Equation 3. Please refer to the Appendix for proof. Now, we obtained a family of estimator $b \in (0, 1)$ who is consistent, asymptotically normal and has a valid variance-covariance estimator if the OLS/WLS case is the same. The perks of using them, compared to using OLS/WLS, is that they are less likely to be subjected to the behavior of the tails.

I applied the estimators to Blanchard et al. (2019). I first partial out the dependent variable and the retaliatory tariff exposure on the controls and FEs, then apply our estimator to the residuals after partialling out. The result is plotted in Figure 11. The red line denotes the value of the point estimate, the upper/lower edge of each box is the CI under confidence level 5% and the upper/lower end of each whisker is the CI under confidence level 1%. The absolute value of point estimate is shrinking as the choice of b approaches 0. It fails to reject $\beta = 0$ when b is less than 0.3. The trend again suggests that our SS-WLS

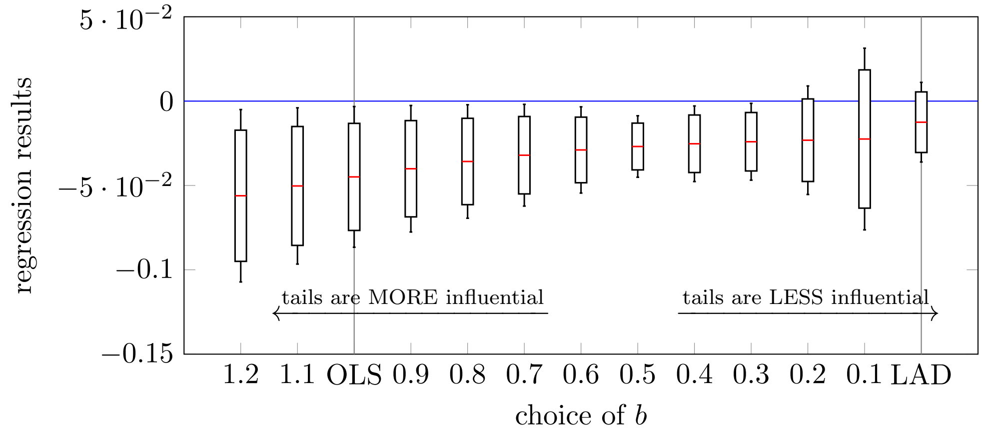

Figure 11: The result of the Z estimators on Blanchard et al. (2019)

regressions are very likely to be subject to fat-tailedness of the data. When the data are fat-tailed, using LAD and customized Z estimator can be a convenient way to produce a more reliable result.

3.2 Rezoning

In this section, we explore the same regression when the weights have more regulated behavior. We adopt the zoning of the congressional districts (CD) for our regression. CDs are designed to be of equivalent population and demographic features to avoid gerrymandering. If we can run the regression on a CD-level we should be able to filter the influence of fat-tailed weights.

3.2.1 Data

The voting data is from MIT's Election Data and Science Lab (MEDSL). This data provides US House election results from 1978-2018. In line with Blanchard et al. (2019), we truncate the data to the time range of 2012-2018[7]. Our sample comprises all American CDs. We

[7]Since the trade war mainly happened in 2018 and 2019. we do not need a 40-year-long panel to show the pre-trend. Still, the main purpose is to replicate Blanchard et al. (2019) as accurately as possible. There was also redistricting happening throughout the time which we want to avoid because it will complicate the situation.

construct our dependent variable as the change of the share of the Republican votes out of all effective votes between the 2016 US House Election and 2018 US House Election ($\Delta RVoteSh_i^{18,16}$). We also use the voter shift during 2012-2014 ($\Delta RVoteSh_i^{14,12}$) and 2014-2016 ($\Delta RVoteSh_i^{16,14}$) as our controls.

I define the trade shock in the same way as Blanchard et al. (2019), but our data source is different. Instead of using the tariff data collected by Bown (2019), I obtain the change in tariff level from the official website of the State Council of China: four lists contain HS 8-digit level commodities upon which the tariff is raised 5%, 10%, 20%, and 25%. [8]. The change in tariff for good g is denoted as $\Delta(\tau_g)$. I also obtained the China-US trade data on HS 8-digit level. I collect all the US import data from China in 2017 X_g, just one year before the trade war shock. The impact of the tariff on product g in dollar terms is then captured as: $TS_g = X_g \Delta(\tau_g)$. I also used the concordance between NAICS and 8 digit HS code from US Census [9]. The second step is to map the trade shocks from the NAICS space to the CD space. We use data from Zip Code Business Patterns (ZBP) from US Census and HUD USPS Zip Code Crosswalk files from HUD's Office of Policy Development and Research (PD&R). From ZBP data, we can obtain the employment in NAICS industry g in zip code area i: $L_{i,g}$. Although most zip codes can be fit into one CD, there are still a considerable amount of zip code areas that span multiple CDs. HUD USPS Zip Code Crosswalk provides data that maps zip code areas to CDs. For zip codes that span multiple CDs, this crosswalk provides the share of business addresses in one zip code area that is located in a certain CD. We can map $L_{i,g}$ to CDs with this crosswalk. The CD-level retaliatory tariff shocks is constructed as:

$$TS_i = \sum_g \frac{L_{i,g}}{\sum_i L_{i,g}} \frac{TS_g}{\sum_g L_{i,g}}$$

This is the same shift-share specification à la Blanchard et al. (2019). Intuitively, the trade shock in industry g (denoted as TS_g) maps proportionately to all American CDs by the CD's employment share in such industry out of the entire United States (the share is denoted as $\frac{L_{i,g}}{\sum_i L_{i,g}}$). I sum up all the mapped trade shocks from all industries and obtained the total trade shock of the event of the 2018 US-China trade conflicts. We further divide this value by the working population to arrive at the trade shock per person on average. This is the same operation mentioned at the beginning of the paper.

[8] See the original notice at http://www.gov.cn/xinwen/2018-08/03/content_5311619.htm

[9] See also https://www.census.gov/foreign-trade/schedules/b/2017/exp-code.txt

Similar to Blanchard et al. (2019), I also added the following controls to absorb other CD-level variations. I obtained the mean income data, population share of different age groups (25-34, 35-44, 45-54, 55-64, 65 and above), population share of different ethnic groups (black, white, non-white-Hispanic), population share of female citizens, share of people with a bachelor degree and above, share of people graduated from high school and share of non-institutionalized citizens that are insured from American Community Survey (ACS)'s API. I added the two lag terms of the change in republican support as my controls as well.

3.2.2 Empirical Model and the Result

The baseline regression specification is:

$$\Delta RVoteSh_i^{18,16} = \beta_1 \log(TS_i) + \Gamma_i' \gamma + \varepsilon_i$$

where i is CD's index and Γ_i is a vector of controls including state fixed effects. In Table 2, we gradually alter the controls from column 1 to column 9. The controls are (1) no controls, (2) education control only, (3) sex control only, (4) health insurance control only, (5) age controls only, (6) past election controls only, (7) economic controls only, (8) race controls only, (9) all available controls. Initially, the trade shock is negatively correlated with the dependent variable (voter support shift to the Republican Party). However, the negative correlation can be absorbed by age controls, past election controls, race controls, and economic controls. Therefore, contrary to Blanchard et al. (2019), we found no negative evidence of the influence of China's retaliatory tariff on US House Election through the lens of CDs. They are negatively correlated but the negative correlation can be absorbed by other covariates.

However, it is worth noting that it is oftentimes considered as one common market within one commuting zone, which is one of the biggest reasons for which researchers stick to county/CZ. When rezoning the smallest spatial unit to CD, we lose this nice property. The zoning of CD emphasizes similar demographics across CDs and neglects geographical and commuting continuity.

Table 2: Regression result after rezoning to CD

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
lgTS	-0.0192*	-0.0219**	-0.0192*	-0.0149	-0.00869	-0.0106	-0.0117	-0.00205	0.00447
	(0.0102)	(0.0107)	(0.0101)	(0.0103)	(0.0111)	(0.00793)	(0.0105)	(0.0108)	(0.00884)
popsh_2534					0.521				0.917*
					(0.581)				(0.543)
popsh_3544					-0.373				-0.697
					(1.217)				(1.256)
popsh_4554					-1.678*				-0.943
					(0.998)				(1.061)
popsh_65					-0.307				0.667
					(0.360)				(0.524)
popsh_bachelor		-0.000132							-1.35e-05
		(8.56e-05)							(0.000138)
popsh_fem			-0.0146						-1.583
			(0.967)						(0.987)
popsh_hi				-0.572***					-0.337
				(0.217)					(0.401)
voterchange_2012_2014						-0.406***			-0.388***
						(0.0746)			(0.0726)
voterchange_2014_2016						-0.699***			-0.737***
						(0.0791)			(0.0781)
unemp_rate							0.0143**		0.0158**
							(0.00661)		(0.00737)
mean_income							0.00978		0.0328
							(0.0481)		(0.0556)
popsh_white_non_hisp								-0.189	-0.206
								(0.171)	(0.175)
popsh_black								-0.0511	-0.161
								(0.170)	(0.192)
popsh_hisp								0.0156	-0.138
								(0.158)	(0.183)
State FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	428	428	428	428	428	428	428	428	428
R-squared	0.202	0.207	0.202	0.213	0.227	0.478	0.219	0.232	0.539

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3.3 Panel Data: Use Additional Information on $\mathrm{E}[\varepsilon_i^2 x_i x_i' | \lambda_i]$

As discussed in Section 2, the efficiency of $\widehat{\Omega}_{HC0}$ is dependent disproportionately on whether ε_i^2 is an accurate representation of $\mathrm{E}[\varepsilon_i^2|\lambda_i]$ for some i with the biggest weights. In a more generalized setting where we don't fix the value and the dimension of x_i, if we can find additional information on $\mathrm{E}[\varepsilon_i^2 x_i x_i'|\lambda_i]$ for important observations, we can get a better estimation of Ω. A direct way to gain more information is through expanding the cross-sectional regression to a panel regression. In a panel regression setting, $\widehat{\Omega}_{HC0}$ is defined as

$$\widehat{\Omega}_{HC0} := \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \lambda_{it}^2 e_{it}^2 x_{it} x_{it}'$$

In a panel setting, $\widehat{\Omega}_{HC0}$ will be a better estimator for Ω than that in a cross-sectional setting. Regional population λ_{it} is almost invariant across time, especially for important observations. The rise of a region usually takes a long time, λ_{it} is highly positively auto-correlated across time within regions. If we ignore the change of regional population along with time, we can write

$$\widehat{\Omega}_{HC0} \approx \frac{1}{n} \sum_{i=1}^{n} \lambda_i^2 \frac{1}{T} \sum_{t=1}^{T} e_{it}^2 x_{it} x_{it}'$$

This is equivalent of replacing the $e_i^2 x_i x_i'$ in $\widehat{\Omega}_{HC0}$ in cross-sectional setting with $\frac{1}{T}\sum_{t=1}^{T} e_{it}^2 x_{it} x_{it}'$. Due of LLN, the latter is a better approximation of $\mathrm{E}[\varepsilon_i^2 x_i x_i'|\lambda_i]$ if $\varepsilon_{it}^2 x_{it} x_{it}'$ is stationary and ergodic across t. In a panel regression with thin-tailed weights, the major difference is that $\widehat{\Omega}_{HC0} \xrightarrow{p} \Omega$ when $nT \to \infty$, while in a regression with fat-tailed weights, $\widehat{\Omega}_{HC0} \xrightarrow{p} \Omega$ when $T \to \infty$. For small T, the improvement is limited.

3.4 Nonparametric Approach to Estimate Ω

Another candidate to estimate the value $\mathrm{E}[\varepsilon_i^2 x_i x_i'|\lambda_i]$ is to exploit the information from nearby observations. Assuming $\mathrm{E}[\varepsilon_i^2 x_i x_i'|\lambda_i]$ is continuous in λ_i, we can utilize non-parametric approach to estimate $\mathrm{E}[\varepsilon_i^2 x_i x_i'|\lambda_i]$ to obtain a better estimation than $\widehat{\Omega}_{HC0}$. Here, I adopt KNN to estimate $\mathrm{E}[\varepsilon_i^2 x_i x_i'|\lambda_i]$,

$$\mathrm{E}[\widehat{\varepsilon_i^2 x_i x_i'}|\lambda_i] = \frac{1}{k} \sum_{j \in KNN(i)} \varepsilon_i^2 x_i x_i'$$

and the new estimator can be written as

$$\widehat{\Omega}_{KNN} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i^2 \mathrm{E}[\widehat{\varepsilon_i^2 x_i x_i'}|\lambda_i]$$

I choose KNN because λ_i at the tail is distributed sparsely. Other non-parametric like kernel regression and splines might not perform well with little observations.

3.5 Results

In this section, I present the results obtained via KNN and panel data regression and also other mentioned methods throughout the paper. It can be observed that only the original regression in column (10) and KNN estimation in column (12) are significant at 0.05 confidence level. Other methods including KNN ($k = 10$), quantile regression, and standard error bootstrapping all failed to reject the null hypothesis. Regarding the KNN method specifically, I also found out that as k grows, the standard error starts to grow (see column (13) where KNN is implemented with $k = 10$). When $k = 1$, KNN estimation is reduced to the original regression. As k grows, the residuals are smoothed gradually. When k goes to infinity, we get rid of any patterns that might potentially be contained in the $\mathrm{E}[\varepsilon_i x_i x_i'|\lambda_i]$ skedastic function. If the model is homoskedastic across regional weights (i.e $\mathrm{E}[\varepsilon_i x_i x_i'|\lambda_i]$ is a constant), then $k \to \infty$ would be the optimal choice. If the model is heteroskedastic, which is often the case, a big k might cause the over-smoothing problem, which leads to bias in the estimation of the standard error. Another perspective to look at this is the bias-variance trade-off: when $k = 1$, the $\widehat{\Omega}_{KNN}$ is reduced to $\widehat{\Omega}_{HC0}$ and has a big variance and at the same time understates the real value of Ω. As k increases, the understatement is corrected and the variance is getting smaller, which is what we want to see. As k continues to increase to infinity, the understatement gets over-corrected and the variance is vanishing. At this point, the estimation is also not reliable. LAD regression in column (14) differs in both point estimates and standard errors from the benchmark regression since it adopts a different point estimate as well. Columns (11) - (15) all adopt additional remedies to alleviate the contamination of fat-tailed weights.

Table 3: Result after alleviating the influence of fat-tailedness of the weights

	(10) Benchmark	(11) Panel Data	(12) KNN(k=5)	(13) KNN(k=10)	(14) Quantile	(15) Bootstrap
TradeShock	-0.0337✓	-0.0337×	-0.0337✓	-0.0337×	-0.0136×	-0.0337×
Standard Error	(0.0152)	(0.0176)	(0.0161)	(0.0185)	(0.00823)	(0.0178)
p-value	0.031	0.056	0.0366	0.069	0.098	0.058
Observations	3,011	3,011	3,011	3,011	3,011	3,011
R-squared	0.6622	0.6622	0.6622	0.6622	0.4891	0.6622

✓ $p<0.05$, × $p\geq 0.05$

4 Conclusion

This paper scrutinizes the possible violation of some WLS presumptions in applied empirical studies that adopt shift-share regression. In general, for the WLS estimator to be consistent, asymptotically normal and Eicker-Huber-White's heteroskedastic standard error to converge to a non-random limit, one needs to make sure that the first and the second moment of the interaction of weights, regressors, and noise are finite. This paper aims to raise the concern of the contamination of fat-tailed weights in a series of papers that adopts US county or US CZ as its observations and weights the observation by its population. Once being weighted, some observations are blown up by the fat-tailed weights. This would compromise the consistency and asymptotic property of the estimator.

I used a Monte Carlo experiment to see the potential effects of fat-tailed weights on regressions and proved that under fat-tailed weights, the Eicker-Huber-White's variance-covariance matrix estimator generally understates the real variance-covariance matrix, and is inefficient. I then take Blanchard et al. (2019) as an example and replicated their work but rezoned their regression observations. In my replication, I found no evidence for the claim that the retaliatory trade shock from China shifts American voters to the Republican Party and this could very likely be a consequence of the fat-tailed weights. Considering using county/CZ has other good properties like being regarded as a common market, I proposed three methods that can alleviate the influence of the fat-tailed weights without rezoning: Z-estimator as well as LAD, exploiting information from panel data and exploiting information from similar-sized observations (non-parametric approach). I then applied these methods to the data set provided by Blanchard et al. (2019). Results all showed a rise in p-value and the coefficient is no longer statistically significant under 0.05 confidence level.

Several limitations of this paper are: (i) it is not verified how the procedure of partialling out would impact non-linear Z-estimator and LAD. (ii) all the methods proposed to alleviate fat-tailedness in the data have some compromise: for Z-estimator, it's the additional assumption of the symmetry of the noise distribution; for rezoning the regression, we lose nice properties like how a CZ can be considered a common market; panel data is only valid when T is large; in non-parametric approaches, we are limited to the bias-variance trade-off. All these remedies provide some degree of "correction", but we are not sure if it is over-corrected or under-corrected, and if new threats are brought in. These are the questions that are remained to be discussed in future works. Albeit not the perfect method, they produce enough evidence to show how fat-tailed weights can affect our regression.

References

Rodrigo Adão, Michal Kolesár, and Eduardo Morales. Shift-Share Designs: Theory and Inference*. *The Quarterly Journal of Economics*, 134(4):1949–2010, 08 2019. ISSN 0033-5533. doi: 10.1093/qje/qjz025. URL https://doi.org/10.1093/qje/qjz025.

David H. Autor, David Dorn, and Gordon H. Hanson. The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review*, 103(6):2121–68, October 2013. doi: 10.1257/aer.103.6.2121. URL https://www.aeaweb.org/articles?id=10.1257/aer.103.6.2121.

David H Autor, David Dorn, Gordon H Hanson, and Jae Song. Trade adjustment: Worker-level evidence. *The Quarterly Journal of Economics*, 129(4):1799–1860, 2014.

Timothy J. Bartik. *Front Matter*, pages i–iv. W.E. Upjohn Institute, 1991. ISBN 9780880991131. URL http://www.jstor.org/stable/j.ctvh4zh1q.1.

Emily J Blanchard, Chad P Bown, and Davin Chor. Did trump's trade war impact the 2018 election? Technical report, National Bureau of Economic Research, 2019.

Olivier Jean Blanchard and Lawrence F. Katz. Regional Evolutions. *Brookings Papers on Economic Activity*, 23(1):1–76, 1992. URL https://ideas.repec.org/a/bin/bpeajo/v23y1992i1992-1p1-76.html.

Nicholas Bloom, Andre Kurmann, Kyle Handley, and Philip Luck. The Impact of Chinese Trade on U.S. Employment: The Good, The Bad, and The Apocryphal. Technical report, 2019.

Kirill Borusyak, Peter Hull, and Xavier Jaravel. Quasi-experimental shift-share research designs. Working Paper 24997, National Bureau of Economic Research, September 2018. URL http://www.nber.org/papers/w24997.

Chad P Bown. The 2018 us-china trade conflict after forty years of special protection. *China Economic Journal*, 12(2):109–136, 2019.

Denis Chetverikov, Bradley Larsen, and Christopher Palmer. Iv quantile regression for group-level treatments, with an application to the distributional effects of trade. *Econometrica*, 84(2):809–833, 2016. doi: https://doi.org/10.3982/ECTA12121. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA12121.

MIT Election Data and Science Lab. U.S. House 1976–2020, 2017. URL https://doi.org/10.7910/DVN/IG0UN2.

Xavier Gabaix. Zipf's Law for Cities: An Explanation*. *The Quarterly Journal of Economics*, 114(3):739–767, 08 1999. ISSN 0033-5533. doi: 10.1162/003355399556133. URL https://doi.org/10.1162/003355399556133.

Xavier Gabaix and Rustam Ibragimov. Rank-1/2: A simple way to improve the ols estimation of tail exponents. Working Paper 342, National Bureau of Economic Research, September 2007. URL http://www.nber.org/papers/t0342.

Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift. Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624, August 2020. doi: 10.1257/aer.20181047. URL https://www.aeaweb.org/articles?id=10.1257/aer.20181047.

Fumio Hayashi. *Econometrics*. Princeton Univ. Press, Princeton, NJ [u.a.], 2000. ISBN 0691010188. URL http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+313736715&sourceid=fbw_bibsonomy.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964. ISSN 00034851. URL http://www.jstor.org/stable/2238020.

Peter J Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 221–233. University of California Press, 1967.

Joseph P. Romano and Michael Wolf. Resurrecting weighted least squares. *Journal of Econometrics*, 197(1):1–19, 2017. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2016.10.003. URL https://www.sciencedirect.com/science/article/pii/S030440761630197X.

Esteban Rossi-Hansberg and Mark L. J. Wright. Urban Structure and Growth. *The Review of Economic Studies*, 74(2):597–624, 04 2007. ISSN 0034-6527. doi: 10.1111/j.1467-937X.2007.00432.x. URL https://doi.org/10.1111/j.1467-937X.2007.00432.x.

Gary Solon, Steven J. Haider, and Jeffrey M. Wooldridge. What are we weighting for? *Journal of Human Resources*, 50(2):301–316, 2015. doi: 10.3368/jhr.50.2.301. URL http://jhr.uwpress.org/content/50/2/301.abstract.

Ron Wilson and Alexander Din. Understanding and enhancing the us department of housing and urban development's zip code crosswalk files. *Cityscape*, 20(2):277–294, 2018.

Appendix A More explanation on the failure of HC0

To further disentangle what caused the failure of $\widehat{\Omega}_{HC0}$, I decomposed $\widehat{\Omega}_{HC0}$ into two additive component ς and $\tilde{\Omega}$.

$$\varsigma := \frac{1}{n} \cdot \underbrace{\frac{\sum_{i=1}^n \lambda_i \varepsilon_i}{\sum_{i=1}^n \lambda_i}}_{\text{Term A}} \cdot \sum_{i=1}^n \lambda_i^2 \cdot \underbrace{\left(\frac{\sum_{i=1}^n \lambda_i \varepsilon_i}{\sum_{i=1}^n \lambda_i} - \frac{2 \sum_{i=1}^n \lambda_i^2 \varepsilon_i}{\sum_{i=1}^n \lambda_i^2} \right)}_{\text{Term B}} \quad \text{and} \quad \tilde{\Omega} := \frac{1}{n} \sum_{i=1}^n \lambda_i^2 \varepsilon_i^2$$

I argue that the bias of $\hat{\Omega}_{HC0}$ comes from ς and the loss of efficiency mainly comes from $\tilde{\Omega}$.

Claim $\mathrm{E}[\widehat{\Omega}_{HC0}] < \Omega_{HC0}$ when λ_i are fat-tailed.

Proof. Term A can be written as $u'\varepsilon$, $u_i = \lambda_i / \sum_{i=1}^n \lambda_i$. Term B can be written as $v'\varepsilon$, $v_i = \lambda_i / \sum_{i=1}^n \lambda_i - 2\lambda_i^2 / \sum_{i=1}^n \lambda_i^2$. We can write ς as $\varsigma = \frac{1}{n}\sum_{i=1}^n \lambda_i^2 (u'\varepsilon\varepsilon' v)$. Therefore $\mathrm{E}[\varsigma] = \frac{1}{n}\sum_{i=1}^n \lambda_i^2 (u'\mathrm{E}[\varepsilon\varepsilon']v)$. It can be shown that $u_i > 0$, $v_i < 0$, $\forall i$. Define matrix D as a diagonal matrix generated from the vector $\{u_i/v_i\}_{i=1}^N$. We have $u = Dv$. $-D$ is positive definite. As a variance-covariance matrix, $\mathrm{E}[\varepsilon\varepsilon']$ is positive semi-definite. Therefore, $-\mathrm{E}[\varepsilon\varepsilon']D$ is positive semi-definite. $u'\mathrm{E}[\varepsilon\varepsilon']v = u'(\mathrm{E}[\varepsilon\varepsilon']D)u < 0$. Therefore, $\mathrm{E}[\varsigma] < 0$. When weights are viewed as deterministic, $\mathrm{E}[\tilde{\Omega}] = \Omega$, therefore,

$$\mathrm{E}[\hat{\Omega}_{HC0}] = \mathrm{E}[\varsigma] + \mathrm{E}[\tilde{\Omega}] < \Omega$$

With fat-tailed weights, $\hat{\Omega}_{HC0}$ will produce conservative robust standard errors. The same bias is negligible and harmless with thin-tailed weights: both term A and term B would be very close to 0 as a consequence of Law of Large Numbers (LLN). $\hat{\Omega}_{HC0}$ still converges to Ω in probability. Our claim would not cause severe distortion in statistical inference. It will have an impact on the size of the t-test when weights are fat-tailed. The magnitude of the bias is also related to heteroskedasticity. If the skedastic function is increasing in weights, this bias is then amplified. If the skedastic function is decreasing in weights, this bias is undermined.

The efficiency loss comes from $\tilde{\Omega}$. Without LLN, even though we have $\mathrm{E}[\tilde{\Omega}] = \Omega$, the variance of $\tilde{\Omega}$ is big. The fat-tailed weights make some ε_i^2 far more important than others in the weighted summation. Therefore, the accuracy of $\tilde{\Omega}$ is dependent disproportionately on whether ε_i^2 is an accurate representation of $\mathrm{Var}[\varepsilon_i]$ for some i with the biggest weights.

Looking further into the distribution of the two components $\tilde{\Omega}$ and ς in the simulation, it is also observed that the main deviation in sample mean of $\hat{\Omega}_{HC0}$ comes from ς (sample mean = -0.0259, sample variance = 0.0019), and the main portion of variance comes from $\tilde{\Omega}$ (sample mean = 1.0004, sample variance = 0.1802). This exercise provides evidence for our claim that the bias of $\hat{\Omega}_{HC0}$ comes from ς while the loss of efficiency mainly comes from $\tilde{\Omega}$.

Appendix B Proof for the consistency for our customized family of Z estimator

To prove that a family member with a parameter $a, b \in (0,1)$ is consistent with regression weights, it suffices to prove that

$$\mathrm{E}[\sqrt{\lambda_i} x_i K(\sqrt{\lambda_i}\varepsilon_i)] = 0$$

We first prove that $\mathrm{E}[|\sqrt{\lambda_i} x_i K(\sqrt{\lambda_i}\varepsilon_i)|] < \infty$.

$$\begin{aligned}
\mathrm{E}[|\sqrt{\lambda_i} x_i K(\sqrt{\lambda_i}\varepsilon_i)|] &= \mathrm{E}[|\sqrt{\lambda_i} x_i K(\sqrt{\lambda_i}\varepsilon_i)|] \\
&= \mathrm{E}[|\sqrt{\lambda_i} x_i \cdot [(a + |\sqrt{\lambda_i}\varepsilon_i|)^b - a^b]|] \\
&\leq \mathrm{E}[|\sqrt{\lambda_i} x_i \cdot ba^{b-1}|\sqrt{\lambda_i \varepsilon_i}||] \quad \text{use the inequality } (a+x)^b - a^b \leq ba^{b-1}x, \forall x \geq 0 \\
&= ba^{b-1}\mathrm{E}[|\lambda_i x_i \varepsilon_i|]
\end{aligned}$$

By Assumption (ii), WLS is consistent, we can assert that $\mathrm{E}[\lambda_i x_i \varepsilon_i]$ is 0, hence $\mathrm{E}[|\lambda_i x_i \varepsilon_i|] < \infty$. We can conclude that $\mathrm{E}[|\sqrt{\lambda_i} x_i K(\sqrt{\lambda_i}\varepsilon_i)|] < \infty$. Secondly, we prove that $\mathrm{E}[\sqrt{\lambda_i} x_i K(\sqrt{\lambda_i}\varepsilon_i)] = 0$. By law of iterated expectation, we have

$$\begin{aligned}
\mathrm{E}[\sqrt{\lambda_i} x_i K(\sqrt{\lambda_i}\varepsilon_i)] &= \mathrm{E}[\sqrt{\lambda_i} x_i \mathrm{E}[K(\sqrt{\lambda_i}\varepsilon_i)|x_i, \lambda_i]] \\
&= \int_{-\infty}^{\infty}\int_{0}^{\infty} \sqrt{\lambda_i} x_i \left[\int_{-\infty}^{\infty} K(\sqrt{\lambda_i}\varepsilon_i) f_{\varepsilon_i|x_i,\lambda_i}(\varepsilon_i) d\varepsilon_i\right] f_{x_i,\lambda_i}(x_i, \lambda_i) d\lambda_i dx_i
\end{aligned}$$

We can prove that the inner integral inside the bracket is always zero given Assumption (i), because $f_{\varepsilon_i|x_i,\lambda_i}(\varepsilon_i)$ is an even function and $K(\sqrt{\lambda_i}\varepsilon_i)$ is an odd function in ε_i. Therefore, the entire RHS is equal to zero and the estimator family for $b \in (0,1)$ is consistent. With our conditions, the unbiasedness also follows.

Appendix C Proof for the asymptotic normality for our customized family of Z estimator

As long as $\varepsilon_i | x_i, \lambda_i$ is not degenerate and x_i is not degenerate, the condition that $\mathrm{E}\left[\left.\frac{\partial \Psi(y_i, x_i, \lambda_i, \theta)}{\partial \theta}\right|_{\theta=\beta}\right]$ is bounded away from 0 is satisfied. We inspect mainly the condition that $\mathrm{E}[\Psi(y_i, x_i, \lambda_i, \beta)^2] < \infty$.

$$\mathrm{E}[\Psi(y_i, x_i, \lambda_i, \beta)^2] = \mathrm{E}[(\sqrt{\lambda_i} x_i)^2 [(a + |\sqrt{\lambda_i}\varepsilon_i|)^b - a^b]^2]$$

Denote $\sqrt{\lambda_i}x_i$ as $\tilde{x_i}$ and $\sqrt{\lambda_i}\varepsilon_i$ as $\tilde{\varepsilon_i}$. An outline of proof would be that at the tail (when $\tilde{\varepsilon_i}$ is big), $\tilde{x_i}^2[(a+|\tilde{\varepsilon_i}|)^b - a^b]^2 < \tilde{x_i}^2[(a+|\tilde{\varepsilon_i}|)^1 - a^1]^2 = \lambda_i^2 x_i^2 \varepsilon_i^2$. Find the threshold points of $\tilde{\varepsilon_i}$ such that $[(a+|\tilde{\varepsilon_i}|)^b - a^b]^2 = [(a+|\tilde{\varepsilon_i}|)^1 - a^1]^2 = \tilde{\varepsilon_i}^2$. Name them as $+v$ and $-v$. We can show that

$$\begin{aligned}
\mathrm{E}[\Psi(y_i, x_i, \lambda_i, \beta)^2] &= \int_{-\infty}^{\infty} \tilde{x_i}^2 f_{\tilde{x_i}}(\tilde{x_i}) \left[\int_{-v}^{+v} [(a+|\tilde{\varepsilon_i}|)^b - a^b]^2 f_{\tilde{\varepsilon_i}|\tilde{x_i}}(\tilde{\varepsilon_i}) d\tilde{\varepsilon_i} + \int_{(-\infty,-v)\cup(v,+\infty)} [(a+|\tilde{\varepsilon_i}|)^b - a^b]^2 f \right. \\
&< \int_{-\infty}^{\infty} \tilde{x_i}^2 f_{\tilde{x_i}}(\tilde{x_i}) \left[\int_{-v}^{+v} [(a+|\tilde{\varepsilon_i}|)^b - a^b]^2 f_{\tilde{\varepsilon_i}|\tilde{x_i}}(\tilde{\varepsilon_i}) d\tilde{\varepsilon_i} + \int_{(-\infty,-v)\cup(v,+\infty)} \tilde{\varepsilon_i}^2 f_{\tilde{\varepsilon_i}|\tilde{x_i}} d\tilde{\varepsilon_i} \right] d\tilde{x}
\end{aligned}$$

We can show that for the first term

$$\int_{-\infty}^{\infty} \tilde{x_i}^2 f_{\tilde{x_i}}(\tilde{x_i}) \int_{-v}^{+v} [(a+|\tilde{\varepsilon_i}|)^b - a^b]^2 f_{\tilde{\varepsilon_i}|\tilde{x_i}}(\tilde{\varepsilon_i}) d\tilde{\varepsilon_i} d\tilde{x_i} < 2|v| \sup_{\tilde{\varepsilon_i}\in[-v,v]} \Big([(a+|\tilde{\varepsilon_i}|)^b - a^b]^2 f_{\tilde{\varepsilon_i}|\tilde{x_i}}(\tilde{\varepsilon_i}) \Big) \times \int_{-\infty}^{\infty} \tilde{x_i}^2 f_{\tilde{x_i}}(\tilde{x_i}) d$$

If we assume $\mathrm{E}[\tilde{x_i}]$ exist, then the LHS integral above is also bounded because

$$2|v| \sup_{\tilde{\varepsilon_i}\in[-v,v]} \Big([(a+|\tilde{\varepsilon_i}|)^b - a^b]^2 f_{\tilde{\varepsilon_i}|\tilde{x_i}}(\tilde{\varepsilon_i}) \Big)$$

is bounded.

For the second term,

$$\int_{-\infty}^{\infty} \tilde{x_i}^2 f_{\tilde{x_i}}(\tilde{x_i}) \int_{(-\infty,-v)\cup(v,+\infty)} \tilde{\varepsilon_i}^2 f_{\tilde{\varepsilon_i}|\tilde{x_i}}(\tilde{\varepsilon_i}) d\tilde{\varepsilon_i} d\tilde{x_i} > 0$$

It can be show that

$$\int_{-\infty}^{\infty} \tilde{x_i}^2 f_{\tilde{x_i}}(\tilde{x_i}) \int_{(-\infty,-v)\cup(v,+\infty)} \tilde{\varepsilon_i}^2 f_{\tilde{\varepsilon_i}|\tilde{x_i}}(\tilde{\varepsilon_i}) d\tilde{\varepsilon_i} d\tilde{x_i} < \int_{-\infty}^{\infty} \tilde{x_i}^2 f_{\tilde{x_i}}(\tilde{x_i}) \int_{-\infty}^{\infty} \tilde{\varepsilon_i}^2 f_{\tilde{\varepsilon_i}|\tilde{x_i}}(\tilde{\varepsilon_i}) d\tilde{\varepsilon_i} d\tilde{x_i} = \mathrm{E}[\tilde{x_i}^2 \tilde{\varepsilon_i}^2]$$

Using Assumption (iii), $\mathrm{E}[\tilde{x_i}^2\tilde{\varepsilon_i}^2] < \infty$. Therefore, the second term is also bounded. $\mathrm{E}[\Psi(y_i, x_i, \lambda_i, \beta)^2]$ is less than the summation of two bounded value. It is therefore, also bounded.