

The logic of our empirical studies (NHSS)

Q: what kind of inequalities of health outcomes do we care?

A: The inequality among individuals (counties)

Q: can we intuitively/descriptively observe such an inequality?

A: Yes, we observed it in the section of Descriptive Statistics

Q: OK, if we are interested in the inequality among counties/individuals & we have intuitively/descriptively observed such an inequality in the descriptive statistics section, then, naturally, can we say only what we care (socioeconomic status) affects/determines such an inequality? Or, euqivalently, what components made up with such an inequality?

A: Obviously, we can depart the total ineuqlity to:

1. inequality not related to any other factor but only determined by the individuals themselves (who the individuals are).
2. inequality not related to who the individuals are, but realted to socioeconomic status which we care in this paper
3. inequality only determined by other fators other than socioeconomic status

The 2nd kind of ineuqlity is just what we want to discuss in this paper. However, if we want to properly evaluate it, we need to distinguish the 1st & 3rd kinds from the 2nd kind of inequality.

Q: Therefore, how to control or separate the 1st & 4rd kinds of inequalities/differences from the 2nd kind?

A: Because we use a panel dataset, it is then easy to find that the **individual effect** corresponds to the 1st & 3rd kinds of inequalities: the inequality only determined by who the individuals are and/or other factors other than socioeconomic status. However, please note that if individual effects are partly related to some factors not included in our regression equations, it means that there is a type of endogeneity (missing independent variables). Therefore, it is neccessary to specially care possible problems of **heteroskedasticity** in our final specifications to expel the effects of the 3rd kind of inequality (through tests, GLS etc). When heteroskedasticity are well solved, the individual effects can be explained as the "pure" ineuqality only determiend by who the individuals are (the 1st kind). In some ways, it indicates the history results/bases in a county area.

Q: OK, now we know that we need individual effects in our model to distinguish the three kinds of inequality. But what type of individual effects to use? Fixed or random?

A: This question can be answered by Haussman test. Meanwhile, because we specially care the heteroskedasticity problem, we use both Haussman test & the robust Haussman test (Wooldridge, 2010). The latter uses an auxiliary regression.

Q: You said you care possible heteroskedasticity problem. So, how to test and solve it in your regressions?

A: We apply the FGLS estimator on all specifications. (Im et al., 1999) Therefore, BP test is nomore required.

Q: OK, so far, you have used fixed individual effects + FGLS estimators in regressions. However, do you need something about time? e.g. time effects and/or stationary tests? **(ROBUST 1)**

A: Because we use a 3-year panel, a short one, we do not need to consider time-series effects. And, to answer if we need time effects, we can simultaneously estimate a **two-way fixed effect** model as a kind of robustness check (compared to our one-way individual effect model).

Q: Common panel data models requires the **normality assumption of residuals**, can you test it? And, if not a normal distribution, how to robustly estimate your equations? **(ROBUST 2)**

A: No, not normal distributions. We performed QQ-plots & some popular normality tests on our model's residuals, then found they do not follow normal distributions (But they look similar to normal ones!) Therefore, we add the **GMM** estimator of our one-way fixed individual effect model to do robustness check.