Possibilities in process standardization

Christopher F. Parmeter

University of Miami

September 24th, 2024

- SFA methods/approaches/issues have grown in the past 25 years.

- Many of these “improvements” have yet to permeate into regulatory frameworks.

- Robust estimation/outlier detection;
- Nonparametric estimation/inference;
- Spatial spillovers;
- Generalized panel data;
- Dependence between noise and efficiency;
- Endogeneity;
- Big data concerns

- Ok, so we have all of these “advanced” methods, but what can they do for us in a regulatory framework?
- **Depends**.

- Spatial example:
  - If companies share information, experiences, insights, or just through observation, based on location, this could lead to comovement in patterns of inefficiency.
  - Weather events, geographical conditions and other unobserved cost drivers (e.g. population structure and electricity demand patterns) are likely to be spatially correlated.
- How would these spatial spillovers impact estimation of the frontier itself and levels of inefficiency?

- Unmodelled spatial effects manifest as heteroskedasticity;
- When concern is over the frontier, this might have minimal effects;
- However, when aim is to decompose error term into noise and inefficiency, unmodelled heteroskedasticity can have large effects on inefficiency which would subsequently impact benchmarking.

- Big Data Example:
  - Suppose regulator wants to include many determinants of the production process and of the utility distributors characteristics.
  - This leads to a large dataset $K \to N$.
  - Inefficiency will effectively vanish from the model.

- Big Data Example:
  - Can use machine learning methods: LASSO, elastic net, random forests, etc.
  - Have to be careful as there is bias introduced.
  - Apply methods step-wise: output, each input, then combined.

- If we do not account for the right kind of heterogeneity, then benchmarking is drawn into question.
- If we account for too much heterogeneity, then we eliminate inefficiency, no need for benchmarking since every distributor is essentially different from everyone else.
- Machine learning methods can balance between these two extremes.

Table: Average of skewness of OLS residuals over 1,000 Monte Carlo replications. The columns show the proportion of irrelevant covariates (generated as independent standard Normal) that are added to the regression based on the sample size (ex: $0.5 * 400 = 200$ irrelevant variables added).

n	c=0	0.01	0.1	0.2	0.3	0.5	0.9
100	-0.494	-0.488	-0.420	-0.342	-0.267	-0.143	-0.001
200	-0.525	-0.517	-0.445	-0.375	-0.299	-0.177	-0.011
400	-0.536	-0.530	-0.454	-0.380	-0.308	-0.186	-0.012
800	-0.547	-0.539	-0.466	-0.391	-0.319	-0.193	-0.016
1,600	-0.549	-0.542	-0.468	-0.391	-0.319	-0.189	-0.016

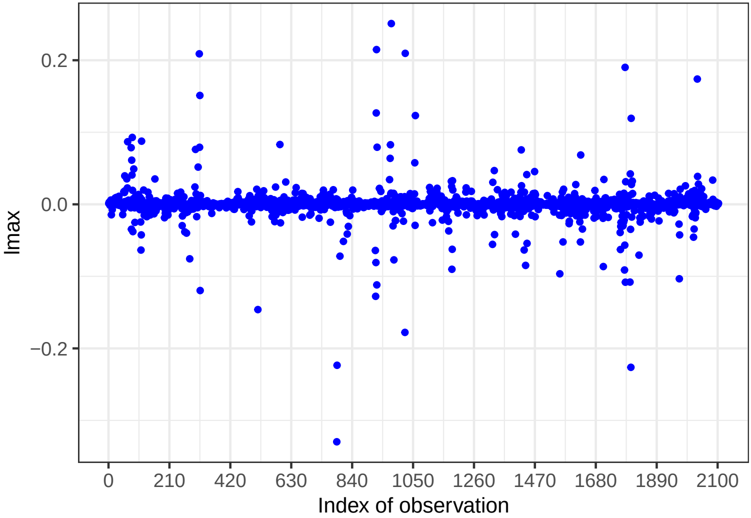
Table: Average of skewness of LASSO residuals over 1,000 Monte Carlo replications. The columns show the proportion of irrelevant covariates (generated as independent standard Normal) that are added to the regression based on the sample size (ex: $0.5 * 400 = 200$ irrelevant variables added).

n	c=0	0.01	0.1	0.2	0.3	0.5	0.9
100	-0.503	-0.404	-0.386	-0.374	-0.367	-0.359	-0.350
200	-0.520	-0.452	-0.436	-0.430	-0.425	-0.420	-0.413
400	-0.536	-0.479	-0.470	-0.465	-0.463	-0.459	-0.455
800	-0.546	-0.506	-0.500	-0.498	-0.497	-0.494	-0.492
1,600	-0.552	-0.522	-0.519	-0.517	-0.516	-0.516	-0.514

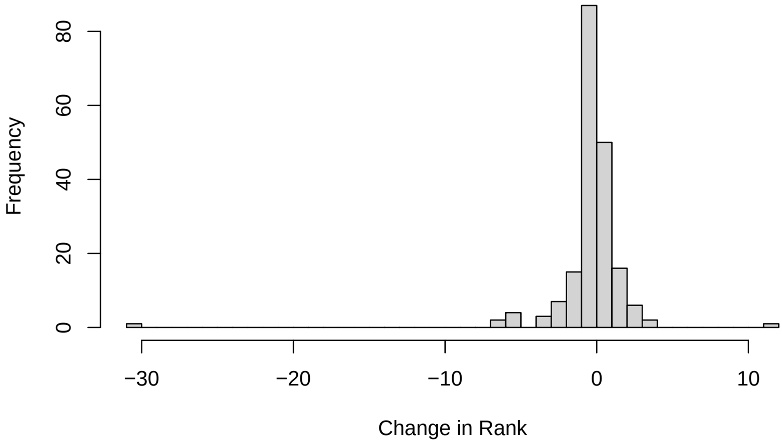
- Robust estimation/Outliers example:
  - Quite often SFA might produce **Type I** errors: failure to find inefficiency.
  - A leading cause of these failures is due to outliers in the data.
  - What can be done?

- Robust estimation/Outliers example:
  - Identify outliers and address (but how to identify?)
  - Use a stochastic frontier specification that is robust (what do we mean by robust?)
  - Use a robust estimator of the desired stochastic frontier model (what is a robust estimator?)

- An approach to identify outliers is to use an **influence graph**.
- Standard regression diagnostics are ill-suited for SFA since they typically rely on Normality of error terms.



0.2
0.0
−0.2
lmax
0
210
420
630
840
1050
1260
1470
1680
1890
2100
Index of observation


Frequency
0
20
40
60
80
−30
−20
−10
0
10
Change in Rank

- A stochastic frontier specification that is robust to outliers is the Students-t-Half Normal specification.

- This likelihood function has bounded influence functions when the degrees of freedom parameter, ν is held fixed (note that as $\nu \to \infty$, the Students-t converges to a Normal).

- Estimation of this model requires either numerical integration, simulated methods or a COLS approach.

- Alternatively, could switch to quantile methods, which are also robust to a fixed percentage of outliers for fixed quantile, τ.

- Quantiles have the benefit of retaining the initial composite error specification (Normal-Half Normal for example) but require finding the τ that is consistent with the location of the stochastic frontier.

- Another option is robust estimation for the stochastic frontier model of choice:
  - Minimum Hellinger Distance;
  - Ψ-divergence;
  - ML$_q$E;
  - Minimum Density Power Divergence.

- The field of stochastic frontier analysis is constantly changing.
- Variety of methods and approaches beyond traditional Normal-Half Normal specification to allay various concerns.
- How useful these approaches are for benchmarking and providing deep insights remains to be seen.