Referee report on Manuscript # 2308,

"Which Panel Data Estimator Should I Use?: A Corrigendum and Extension"

Summary

The paper tackles a topic essential to applied panel data analyses, namely the choice of estimator and standard error to use in given constellations.

This is a follow-up paper to Reed and Ye (2011). The paper adds insight and ultimately provides applied researchers with a clear recommendation as to which estimators should be used. While the setup is restricted to static fixed-effects panels, this is general enough to cover a wide variety of applications.[1]

Comments

1. The finding that researchers should use different estimators for estimating coefficients and testing hypotheses is not really unusual. When interested in a point estimate, bias is of secondary importance and making the estimator "as efficient as possible" in MSE terms is the rational choice. Similarly, inference should be conducted using the most reliable tools. Should the tools have turned out to be the same for both goals, it would have been what some call "nice-to-have", but not more.

   The authors could perhaps have provided some explanations as to why these particular recommendations arise. This referee suspects strongly that the reason is the bias-variance trade-off (as hinted in the paper and also in Reed and Ye (2011), but not discussed thoroughly). The reasoning is as follows.

   Concretely, the inversion of the estimate of the covariance matrix Σ is neither very precise nor resulting in an unbiased estimate of the inverse Σ^{-1}. For relatively large cross-sectional dimension N (concretely, $T/N < 1.50$), this offsets the efficiency gains stemming from setting up a full-scale GLS-type estimator. For $T/N > 1.50$ this is obviously less of an issue and the Parks estimator may be used without fear. It is noteworthy that inverting the matrix Π does not cause similar issues, but this is because Π depends on one parameter only.

[1] This referee looks forward to a possible follow-up paper for dynamic panels.

For testing however, it is quite important that the test statistic be correctly centered. For instance, it is well-known in the dynamic panel data literature that the bias (more specifically the Nickell bias) is the cause of lack of size control in parameter tests. The argument obviously holds in the static setup of this paper for any source of estimation bias.

To clarify the issue, it may therefore be informative to indicate how the MSE decomposes in bias and variance. Of course, this is not primarily relevant when only recommending an estimator, since the recommendations wouldn't change, but explaining the reasons why the recommendation is as it is may increase its rate of adoption.

Also, additional estimators may be worth considering, say one that over-simplifies the cross-sectional dependence structure (for FGLS estimation), but does not assume it away. For instance, a constant cross-sectional correlation scheme may be worth a try (cf. O'Connell, 1998). Admittedly, this is not implemented in standard software, just like the bootstrap.

Minor remarks

1. The considered data generating processes match widely used empirical data sets. The simulations are however conducted such that at least one estimator is correctly specified. It may be interesting to also consider models with AR(2) models or cross-sectional heterogeneity in the AR coefficient.

2. The way DGPs are generated from existing data sets could perhaps be motivated in more detail. For instance, it is not clear why only the first N (5 in the example) are considered and not averages over several groups of N (like averaging over several contiguous length-T periods).

3. Finally, the "gaps" in the original paper of Reed and Ye (2011) need not be interpreted in such a pessimistic key; in fact, applied workers have more freedom to choose an estimator in those cases, which may then accommodate personal preferences.

Additional references

O'Connell, P.G.J. (1998) The Overvaluation of Purchasing Power Parity, *Journal of International Economics* 44, 1-19.