Referee's Report on: Bootstrap Methods for Inference in The Parks Model by Mantobaye Moundigbaye, Clarisse Messemer, Richard W. Parks and W. Robert Reed

This paper considers inferences for the SUR model with AR(1) errors: SUR(1), say. Since this is an important model in applied research it behooves us to get this right. Let's consider the authors contribution and how it can be done better.

The authors suggest using a bootstrap percentile method to compute better critical points for the test statistics. This is fine. Comparisons to the Beck and Katz paper are made. References include Rilstone and Veall which is close to what the authors are doing. Rilstone and Veall looked at bootstrapping inferences for the basic SUR model. It should be mentioned early and contextualized by saying that the current paper does a similar analysis as Rilstone and Veall with the complications of serially correlated errors.

To provide some more context, in the preliminaries the authors should state the GLS estimator and provide its asymptotic distribution, noting that FGLS, using standard consistent estimates of the ρ_i's and Σ, will be asymptotically equivalent. Also mention variety of estimators would work and discuss why the ones that are popular are used.

This is a purely Monte Carlo paper so I would expect that a substantially wider range of results be reported (where viable or at least summarized) including different estimators. Also results at different significance levels. See the Rilstone and Veall paper for idea of range of estimation techniques.

My reading of the Beck and Katz paper is that their intent was to outline certain shortcomings of the Parks method and propose some alternatives. First and foremost is that (perhaps due to overparameterization) the standard errors from the Parks estimator can seriously underestimate variability of the estimates. Also that the Parks estimator can be inefficient to OLS (under correct specification this would only be the case in finite samples, which could be noted) and also only works when $T > N$. (Should be much greater). Their point seems to be to use OLS but with correct standard errors. (Beck and Katz is not the paper I'm reviewing but my comments are relevant in this regard). OLS always has its role and reporting it along with corrections for its distribution are important. I'm not sure if Beck and Katz

got it right wrt correct standard errors. HAC for SUR is available. That should be the point of comparison. (The formulas in the version of the Beck and Katz paper I have are unreadable.)

A couple of remarks re Beck and Katz. Their paper was written before an understanding of the higher-order accuracy of bootstrapping was widespread. Their focus (my understanding) on getting correct standard errors for OLS is justified (and was timely in 1995) (though I'm unable to verify from my copy of the paper whether their formula is correct). Their dismissal of the Parks estimator may be missplaced: they should also have considered the Parks estimator *with* robust standard errors.

If the generated data *is* SUR(1) then GLS is efficient and FGLS (Parks) will be ayymptotically efficient. One would expect an experiment constructed with such data would favor one of these methods. We're in an "it depends" situation. If the true dgp is not SUR(1) or even in finite samples OLS might dominate

Utimately one issue that should arise is with respect to power of the tests. Bootstrappinmg should give you relatively unbiased tests *provided a good pivot*. An asymptotically efficient estimate combined with proper boootstrapping should also have good power properties (this is not the focus here).

The simulations from the Grunfeld data should be rerun making sure to include the following.

- inferences based on both OLS and Parks estimators

- iinferences based on asymptotic standard errors under correct speciifcation

- inferences based on bootstrapped standard errors

- inferences based on HAC consistent standard errors

- inferences based on critical points using the χ^2 distribution

- inferences based on bootstrapped critical points

Note

- Parks should be consistent under misspeciication of covariance but even its incorrect standard errors can be robusutified

- report true GLS results as benchmark

It would be interesting to report the MSEs of OLS and Parks from the Monte Carlo. If Parks is the correct specification it may have greater variance but better MSE.

Also report average of estimated standard errors to the standard deviation of the estimates to show how wrong the estimated standard errors are.

I'm not certain that equation (7) (and hence (8)) is correct. From my reading, I believe Beck and Katz suggest using OLS along with "correct" standard errors for OLS. I may be wrong.

Most of what I suggest can be done by adding a few lines of code to currect Monte Carlo. I'd stick to the Grunfeld data set, but examine it more extensively.

It would be interesting to see what evidence that the actual data set of Grunfeld used as the basis for the Monte Carlos has the SUR(1) structure. Did the authors (or others who worked on this data) do any testing for these features?

Only need to mention "2400 Web of Science references" once.

Typo in Bruckner reference.

Remove "innovative" in Abstract. Lots of Monte Carlo experiments are designed using moments from published data sets as basis.

Remove the term "Pareto-improving" in Abstract. It's inappropriate here.