Dear Charles,

Many thanks for your insightful comments. We have looked at them and performed further checks of our analyses and results. Our conclusion is still that an **aggregation bias exists, and it is large**. Below we explain why, and why we get different results from you. We hope that this clarifies the matters and that you will agree that the aggregation bias is indeed a serious concern.

1. We do bootstrapping in our first experiment because we draw (with replacement) from the estimated cross-section of sectoral rhos and assume some disturbances at the sectoral level. Given these assumptions on the DGP, we have to bootstrap. Your proposed method seems to actually assume that the rhos are deterministic, and perfectly observed. Is that more plausible?

2. You criticize our second experiment on grounds that we have lowered the average rho to 0.8. There are two separate issues here. One is whether to include explosive roots at the sectoral level, the other is a measure of the cross-sectoral heterogeneity in persistence. The paper demonstrates that the bias increases with this heterogeneity. Thus, if the dispersion in cross-sectoral persistence is calibrated so as to make sure no process is explosive, one may throw the baby with the bath water, and obtain a lower bias, simply because heterogeneity is much lower. We have acknowledged that latter phenomenon in the paper all along.

So then there are two "free" parameters: the variance of the cross-sectoral distribution, and the truncation interval. In our experiments, we have played around with both. Our second experiment was meant to illustrate this fact, by reproducing your exercise with a wider truncation interval (and, naturally, a lower mean rho). You are right in saying that expanding the range of possible rhos jacks up the bias. That's proved in the paper. That said, however, 0.8 is indeed not the mean rho as implied in the data, which is why we ran experiment 3.

3. There are, once again, two issues here. First, you don’t seem to reproduce our results. Second, you say “the small sample bias always dominates the aggregation bias”. The two points are related.

We argue that you don’t find the same results we do because you include a (fixed effects) intercept in your aggregate estimation, even though you assume none in the Data Generating Process (DGP) at the sectoral level. This is also the reason why the small sample bias appears to dominate in your exercises (see below). In our experiments, we always include fixed effects (as well as stochastic autoregressive parameters) in the DGP when looking at potential problems with the FE estimator. Given the estimated model, this is the correct approach.

The difference may seem irrelevant, but it is actually far from it in small samples (you use T=150 in your experiments). As is well known, OLS estimates of the autoregressive parameter are negatively biased in small samples. Tanizaki (available at http://ht.econ.kobe-u.ac.jp/~tanizaki/cv/working/unbiased.pdf) shows that this bias is substantially aggravated for the case in which a constant term does not appear in the DGP but is included in the empirical model. This happens since de-meaning the data introduces correlation between the regressor and the error term and leads to a larger downward bias in AR models. This is also the point in a recent paper by Choi, Mark and Sul (2003). To check how this explains the discrepancy in our results, just revise your GAUSS program by including a fixed effect in the DGP. You'll find a large and dominant aggregation bias.

Our impression here is that explosive roots do not matter, contrary to what you claim. If you preserve the variance of heterogeneity, but truncate the distribution to exclude explosive roots, you still get a bias. In your experiment, you truncate *and* reduce cross-sectoral dispersion. If you were to impose truncation on the original experiments in the paper, you would still get a bias.[1] Actually, for the sample size that you consider, heterogeneity bias sets in for slightly higher variance than what you assume. We attach Figure 1, which reproduces your experiment, for T=150 *and* with fixed-effects in the DGP, for various values of the variance parameter. The aggregation bias kicks in for a standard deviation larger than 0.03.

The intuition in Tanizaki or Choi et al carries through in a panel setting. There are two sources of bias, the aggregation bias and the standard FE bias. Again the constant terms are important. As mentioned above, our DGP includes fixed effects, while we guess that Shiu-Sheng's does not. The FE bias has much the same consequences as the OLS small sample bias for single time-series. In particular, if one includes the FE in the DGP, the aggregation bias of the FE estimator is evident while it seems to be much less evident when no FE is included in the DGP. Actually, we were able to reproduce Shiu-Sheng's result by setting the fixed effects to zero in all the cross-sections (and taking a "within" transformation of the data) when estimating the empirical distribution of the FE estimate of the autoregressive parameter. We attach the GAUSS programs for the panel estimates that show this. (Using our seed, i.e. setting rndseed 2^20, the program should generate an estimated persistence under fixed effects equal to 0.97722328, whereas mean persistence equals 0.94000108).

Hence, we believe our results are correct. Your negative results on the aggregation bias would be appropriate if the true value of the fixed effects were zero. Furthermore, for the sample sizes in our paper, which are longer than the T=150 considered in this discussion, the aggregation bias becomes much more important relative to the FE bias.

4. As mentioned above, we believe Shiu-Sheng's results fall victim to our criticism, as we were only able to replicate his results in our (attached) program when including *no* intercept at the sectoral level, but demean in the aggregate.

---

[1] Actually, in a Random Coefficient Model, series with explosive roots would probably get a *low* weight, contrary to what you claim in your previous mail, since they are weighted by the inverse of the variance-covariance matrix. Also, there is a difference between estimating half-lives, which are infinite for series with explosive roots, and estimating autoregressive coefficients, which is what we do.

Regarding the Monte Carlo experiments with 10 and 100 panel elements, there is heterogeneity in both cases and both show there is an aggregation bias. One cannot compare the estimates across the two experiments and use the fact that the fixed effects estimates of rho are similar as an argument against aggregation bias (which is what we understand you do). These are two different experiments, with different realizations of uncertainty. The proper comparisons are between FE estimates and the true values. The bias is unquestionably large, whether the panel has size 100x100, or 10x100.

We did not report RCM estimates there because the simulated sample is short, and thus, as we have acknowledged all along, estimates potentially fall victim to a small sample bias. However, since you are asking, Figure 2 compares the performance of RCM and FE where N=221, T= 150 and the heterogeneity is drawn from a uniform(+-0.059) with mean rho=0.94. There is a small downward bias in RCM (around -0.01) due to high persistence, but the bias in FE is substantial.

<u>References:</u>

Tanizaki, Hisashi (2003), On Least-Squares Bias in the AR(*p*) Models: Bias Correction Using the Bootstrap Methods, mimeo Kobe University.

Choi, Chi-Young, Nelson Mark and Donggyu Sul (2003), The Dominance of Downward Bias in Half-Life Estimates of PPP Deviations, mimeo.

Figure 1

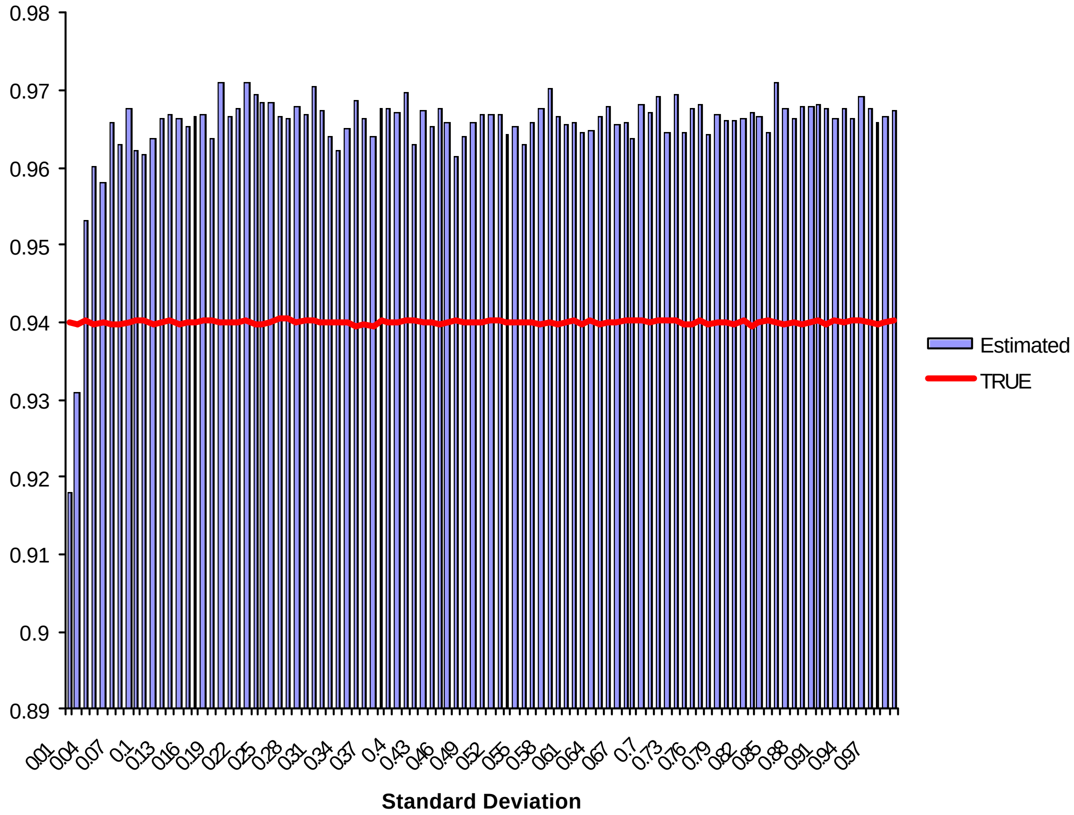
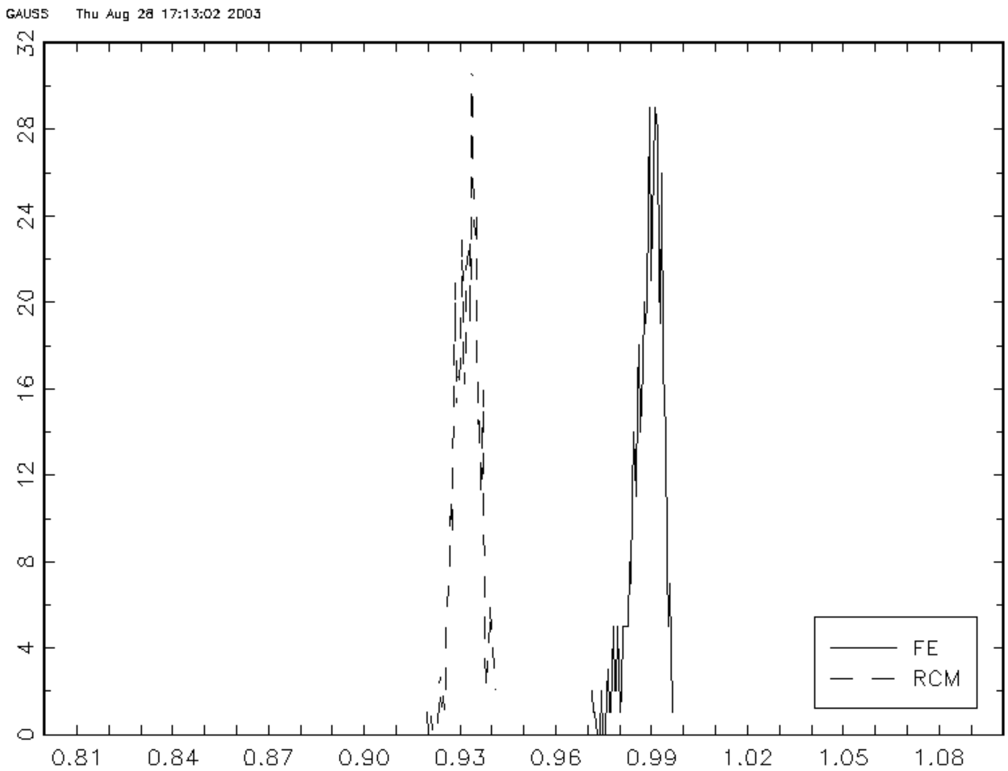

0.98
0.97
0.96
0.95
0.94
0.93
0.92
0.91
0.9
0.89
0.01 0.04 0.07 0.1 0.13 0.16 0.19 0.22 0.25 0.28 0.31 0.34 0.37 0.4 0.43 0.46 0.49 0.52 0.55 0.58 0.61 0.64 0.67 0.7 0.73 0.76 0.79 0.82 0.85 0.88 0.91 0.94 0.97
Standard Deviation
Estimated
TRUE

Figure 2


GAUSS Thu Aug 28 17:13:02 2003
0
4
8
12
16
20
24
28
32
0.81
0.84
0.87
0.90
0.93
0.96
0.99
1.02
1.05
1.08
FE
RCM