Misleading Inferences from Panel Unit-Root Tests: A Comment

GEORGE S. FORD, *President, Applied Economic Studies, Tampa, Florida, gford@aestudies.com.*

JOHN D. JACKSON, *Professor of Economics, Auburn University, Alabama, jjackson@business.auburn.edu.*

AUDREY D. KLINE, *Assoc. Dean of Business, University of Louisville, Kentucky, audrey.kline@louisville.edu.*

I. Introduction

In an interesting paper published in this journal in 2001, Professors Boucher-Breuer, McNown and Wallace (BMW) argue that panel unit root tests such as Levin and Lin (1992) are flawed because "they are incapable of determining the mix of I(0) and I(1) series in a panel setting (at 482)" and that the conclusion of the test applies to all members of the panel even though the panel may include a mix of I(0) and I(n) series. The implication, the authors contend, is that the "all or nothing nature of these procures may lead to serious misinterpretations of tests applied to mixed panels (at 483). In response to this shortcoming, BMW propose an alternative test called the SURADF test, which is an Augmented Dickey Fuller (ADF) test for each element of the panel but estimated as a system using Seemingly Unrelated Regression (SUR). Based on Monte Carlo simulations, the authors conclude that the SURADF is a more powerful test than the ADF on individual series.

In this comment, we do not contest the particular criticisms of traditional panel unit root tests levied by BMW. What we attempt to show is that while the SURADF test may be fruitful for some problems, the test suffers from a serious shortcoming that limits, perhaps to a large extent, its applicability. The limitation arises from the fact that SUR accounts for the correlation of errors among panel members. Therefore, the results of the SURADF test depend on the selection of panel members and, as we show, the test results may not be robust for any particular member of the panel if the other members are altered. This problem is significant because in many cases panel members are selected or excluded for somewhat arbitrary reasons, including a) data availability; b) researcher interest in particular countries; or c) membership in some organization that collects and reports data such as the OECD. If the results of a test are highly dependent on the selection of panel members, then arbitrary selection criteria will lead to fragile conclusions from statistical testing.

II. The SURADF Test and Mixed Panels of Mixed Panels

Put simply, the SURADF test estimates the parameters of the ADF test for each member of a panel using SUR (and fixed effects). In the case of a single lag term, each ADF test includes three coefficients including a constant term and coefficients for q_{t-1} and Δq_{t-1} for a total number of estimated coefficients equal to $3m$, where m is the number of panel members. Simulations were employed to estimate the critical values for each member of the panel. To shorten this comment, we direct the reader to the original paper for the details.

BMW begin their analysis with 45 years (1950-95) of exchange rate data for 14 OECD countries (arbitrarily selected from 25 OECD members) from the IFS CD-ROM. We replicate their data and extend the sample to include 20 countries. From the original sample of 14 countries, BMW select 10 with which to perform their SURADF test. Results from our replication of the SURADF are summarized in Panel Group A of Table 1. Of the ten members, the null of a unit root is rejected for only three.

In Panel Groups B and C in Table 1, we present two sets of results that differ only in the selection of panel members. In Panel B, the only change is that Spain is replaced with Iceland. The consequence of the replacement is the loss of statistical significance on Norway and Denmark, both of which were statistically significant only at the 10% level for the BMW group. Greece remains statistically significant at the same 5% level. By replacing four of the original ten countries in the panel (Panel Group C), the statistical significance of Norway, Denmark and Greece is eliminated (even at the 10% level). None of the series is determined to be stationary in this last grouping.

The fundamental precept of statistical sampling and a cornerstone of econometric analysis is to produce results from a small subset of data that can be generalized to a much broader (both across observational units and over time) class of outcomes relating to similar observational units. As shown in these few examples, the proposed procedure does not permit generalization at all. That is, the SURADF test is not a robust method to assess the presence or absence of a unit root when panel members are not uniquely determined. At least when the ADF is applied to each panel member individually for a given time period, unambiguous results are obtained – the series is stationary or it is not. For the SURADF test, however, our examples show that the conclusions concerning stationarity are ambiguous, depending on which countries are taken as panel members. Furthermore, because of the role played by cross-equation error covariances in SUR estimation, this lack of generality is itself a general result. And it is a particularly problematic result because with time series data, rarely do we have the luxury of including all potentially relevant panel members if for no other reason than a lack of data. Even in cases where all OECD countries could be included for an "OECD analysis," we are left with the question of why OECD membership is relevant for the determination of exchange rate behavior. Since missing data likely excludes some OECD countries from the panel, even a member-wide analysis is precluded.

Another potential problem with SURADF is when the number of potential panel members exceeds (or, in some cases, even approximates) the number of time periods.[1] In such cases, SUR will not estimate thereby requiring panel members to be excluded or split into groups. Even if a plausible method is employed to make such decisions, say cluster analysis or some arguably exogenous attribute (e.g., fixed versus floating

[1] It is this particular problem that initiated our interest in the properties of SURADF, since we attempted to apply the test to a panel that was wider than deep.

exchange rates), it would seem prudent to evaluate the robustness of the results, which, depending on the potential number of panel members, may be a daunting task.

III. Conclusion

In an effort to resolve the “all or nothing” nature of panel unit root tests, Professors Boucher-Breuer, McNown and Wallace propose the SURADF test that can determine the mix of stationary and non-stationary series in a panel. We reveal a practical shortcoming of the test. Specifically, the results of the test appear to be highly sensitive to the selection of panel members. Since member selection is often driven by somewhat arbitrary rules such as missing data or membership in an organization, this sensitivity of the test is perhaps a significant shortcoming greatly limiting both its applicability and its generality.

References:

Boucher Breuer, Janice, Robert McNown and Myles. S. Wallace. "Misleading Inferences form Panel Unit-Root Tests with an Illustration from Purchasing Power Parity," *Review of International Economics*, 9(3), 482-493 (2001).

Levin, Andrew and Chien-fu Lin, "Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties," University of California-San Diego Discussion Paper 92-23 (1992).

Table 1. Results from the SURADF Test

	Panel A.				Panel B.				Panel C.			
	SURADF	Critical Values			SURADF	Critical Values			SURADF	Critical Values		
Country		0.01	0.05	0.10		0.01	0.05	0.10		0.01	0.05	0.10
Norway	-5.515***	-6.176	-5.449	-5.099	-4.789	-6.257	-5.424	-5.084	-3.872	-5.979	-5.226	-4.824
Netherlands	-5.138	-6.528	-5.861	-5.506	-4.665	-6.486	-5.741	-5.400	-3.901	-6.313	-5.584	-5.183
Denmark	-5.860***	-6.629	-5.936	-5.569	-5.134	-6.592	-5.830	-5.497	-4.345	-6.389	-5.594	-5.257
Spain	-3.459	-5.649	-4.946	-4.556	…	…	…	…	…	…	…	…
Germany	-5.056	-6.580	-5.826	-5.487	-4.503	-6.511	-5.794	-5.420	-4.242	-6.315	-5.588	-5.251
Belgium	-5.490	-6.731	-6.014	-5.642	-4.873	-6.715	-5.970	-5.641	-4.366	-6.283	-5.583	-5.180
Greece	-4.806**	-5.217	-4.462	-4.082	-4.981**	-5.172	-4.413	-4.028	-3.784	-5.218	-4.386	-4.023
France	-4.444	-6.075	-5.351	-4.972	-4.054	-6.075	-5.268	-4.916	…	…	…	…
Finland	-3.967	-5.509	-4.760	-4.371	-3.503	-5.527	-4.730	-4.361	…	…	…	…
Luxembourg	-5.420	-6.602	-5.916	-5.552	-4.895	-6.623	-5.897	-5.540	…	…	…	…
Iceland	…	…	…	…	-2.683	-4.989	-4.198	-3.828	-2.986	-4.942	-4.184	-3.774
Japan	…	…	…	…	…		…	…	-0.681	-5.086	-4.340	-3.956
Canada	…	…	…	…	…		…	…	-2.452	-4.941	-4.128	-3.735
UK	…	…	…	…	…		…	…	0.389	-5.187	-4.407	-3.994

** Statistical significance at the 5% level.
*** Statistical significance at the 10% level.