

Reviewer report for

The Balassa-Samuelson Hypothesis in Developed Countries and Emerging Market Economies: Different Outcomes Explained

by

Jose Garcia Solanes and Fernando Torrejon Flores

My comments are grouped in three parts. The first part is concerned with the economic analysis, the second with the econometric analysis and the third relates (by means of just very few examples) to matters of presentation or writing.

Let me start by stating that I truly believe that the biggest challenge in economics is to perform high standard empirical analysis since in order to do so one needs to master several fields: First, one needs to know the relevant economic theory that enables one to formulate an interesting question, second one needs to be an (almost) expert in econometrics (in order to assess the potential usefulness of methods as well as to ensure proper implementation) and last but not least one needs to be well versed in understanding economic data. My comments have to be seen against this background.¹

I am confident that my comments will be interpreted in the way intended: as potential help for strengthening the analysis in the paper. I believe that you address an interesting question – which deserves to be studied with utmost care in particular also with respect to the econometric methods applied.

1 Economics

The description of the BS model is relatively standard, where I would just like to mention that what you call first step of the hypothesis is originally known in the literature as Baumol-Bowen effect, see Baumol and Bowen (1966), using the instructive example of the relative productivities in theater performances and automobile production and the corresponding evolution of relative prices (cars becoming relatively cheaper, since the number of actors for Hamlet as well as the time they need stays constant essentially).

A detailed yet simple derivation and discussion is contained e.g. in Wagner and Hlouskova (2004). In that paper also the impact of extensions like relaxing wage

¹ As a general remark let me add that I believe that empirical economics often is not as esteemed as it deserves to, because there are too many papers written that do not adhere to the above (or any similar) list of important characteristics.

homogeneity across sectors (unexplored in this paper) as well as the absence of PPP in the tradable sector (what you refer to as second step) are explored.

I have the impression that a streamlining of the model setup and description is possible which would allow the reader, at least this reviewer, to follow the arguments more easily. E.g. to a certain extent I wonder about the necessity of Section 2.2, which is nowhere again referred to, since you do reject the null of no cointegration between the nominal exchange rate and the price differential in the tradable sector (i.e. something like stationarity of the RER of tradables).

I will abstain from a detailed discussion concerning the economic implications of your work, since I see some potentially severe problems with the econometric analysis that are detailed in the next section of this report. I would only like to ask how you come to the assessment/potential explanation that trade costs (not only measuring physical transport costs but all costs associated with cross-border trade) should be lower in Latin America than across the OECD? Many (most) of the OECD countries are members of the European Union, which has the single market as a major objective and on top of that I would believe that also physical transport infrastructure in Europe is developed to a larger extent than in Latin America (which is also geographically larger than Europe)? In case you have quantitative evidence for lower trade costs in Latin America than in the OECD or the EU it would be very interesting to include this in the paper to substantiate your hypothesis.

2 Econometric Analysis

As I have tried to state at the beginning of the report I believe that a researcher applying econometric methods has to exert greatest care when deciding which econometric methods to use. I believe that it is the user who bears the responsibility for using econometric methods properly. To use a potentially *strange* metaphor: It is a bit like that the buyer of a gun (user of econometric method) has to make sure that it is stored safely and only used properly so that not everybody can take it and shoot around. And to stick to my metaphor: I do not believe that if something happens one should sue the gun producer (developer of econometric method) if some legal gun owner shoots somebody.

Getting back to the paper, the econometric strategy you follow consists essentially of the following three steps, which you perform twice - once for the relationship between (using the notation of the paper) $dp+da_N$ and da_T and then for the relationship the nominal exchange rate and the price differential in the tradable sector.

The three main steps are given by:

1. Application of one of the tests of Banerjee and Carrion-i-Silvestre (2006), BCiS henceforth, to test the null hypothesis of no cointegration, whilst allowing for cross-sectional dependence via common factors, potentially also for breaks (but maybe not given that $T=13$, this has not become clear to me when reading the paper). You reject the null hypothesis of no cointegration.

Then you estimate the cointegrating relationship for every country writing on page 12 after the table “Given that there is a cointegration relationship between the series, we estimate...” Strictly speaking rejection of the null does not imply acceptance of the alternative, in the panel context in particular not necessarily the acceptance of the alternative for every country. In particular you consider two estimation approaches:

2. One being country specific estimation using the FM-OLS estimator developed in Phillips and Hansen (1990).
3. The other is to use the pooled FM-OLS estimator developed in Kao and Chiang (2000).

Let me now outline some potential problems related to each of these steps:

Step 1:

It is very important to address cross-sectional dependence when considering RER and related panels, compare the discussions in Wagner (2005, 2007) and with respect to the Balassa-Samuelson model in particular the detailed analysis in Wagner and Hlouskova (2004). In this context it is very important to use tests that allows for cross-sectional dependence via common factors, as e.g. those of BCiS (to which I will come back below).

I guess, however, that it would be important to start with panel unit root analysis prior to cointegration analysis in order to make sure that the series involved in the potential cointegrating relationship are adequately characterized as being $I(1)$ processes. Clearly, such tests also need to allow for cross-sectional dependence in a similar way as the cointegration test. Maybe it might also be worthwhile to consider an unrestricted version of the step1 relationship and test for the respective coefficient being equal to -1 .

The small sample that you have, is however, a major limitation for the methods you use (and also for similar methods for unit root testing). If you look at the BCiS paper you can see that in their simulation study they use panel dimensions $T=100,250$ and $N=20,40$, compared with your $T=13$ and $N=16$. It is entirely unclear how the method you

use performs for the sample size you have at hand. Some potential upper bound for the performance is given by the simulation results reported in Hlouskova and Wagner (2005) and Wagner and Hlouskova (2007), who study (procedurally simpler) first generation panel unit root and cointegration tests and estimators for cross-sectionally independent panels. Their smallest time dimension is $T=10$ for which depending upon experiment considered **quite strong** distortions occur. This is as expected to a certain extent, since for all methods at some point or another some time series unit root or cointegration test or estimation step is performed. Now, given that you in addition have common factors (which need to be estimated and extracted) the performance of your approach cannot be better than for the corresponding methods in the cross-sectionally independent case, i.e. for Peter Pedroni's tests. Furthermore, also the determination of the number of factors using the methods of Jushan Bai is a rather difficult issue for small values of N , since consistent determination of the space spanned by the factors and the number of factors requires a sufficiently large cross-sectional dimension to work in practice.

I have the impression that before strong conclusions are drawn from the BCiS test, a much better understanding of its small sample performance has to be reached first. In a sense that probably is not too much additional work for you, since you perform bootstrap inference (for the FM-OLS coefficients), which means you must have code available that would allow you to perform a simulation study, potentially upon slight modification.² Extending this to also study the performance of the factor model related parts of the method (i.e. the performance of the Bai information criteria for determining the number of factors) would be very interesting.

Furthermore, operationally it would be interesting to see e.g. how many common factors are extracted in the BCiS procedure and as said above, appropriate panel unit root results.³ Another question is, why you report the results for exactly that one out of the seven tests?

Step 2:

The Phillips and Hansen estimator is based on non-parametric estimates of long-run variances respectively one-sided long-run variances. These are given by kernel estimates

² In the inferential part of your paper you resort to bootstrap inference. I believe that consistency of bootstrap procedures for panels with permanent cross-sectional dependencies are still missing in the literature.

³ I think, compare Wagner and Hlouskova (2004) that one possible outcome of all the additional experiments is that methods for nonstationary panels are not sensibly useable for very small panels with cross-sectional dependence. I know that this result is not per se "publishable" but potentially still valuable in terms of knowledge creation.

based on empirical variances and covariances. The short sample with 13 observations is bound to lead to rather poor performance of non-parametric kernel estimation; but basically for any estimator of cointegration.

Given that you are not afraid to use a time series cointegration estimator when $T=13$, I do not really understand why not also test for cointegration country-specifically? Why only test with a panel set-up? If $T=13$ is good enough for estimation, why is not good enough for testing?

Let me note as a side remark that for the country specific estimation procedure it might be possible to use or develop consistent bootstrap procedures based on the work of Park (2003) and Swensen (2006), both in slightly different contexts which nevertheless provide important inputs for bootstrap analysis in unit root contexts.

Step 3:

Next to the FM-OLS estimation you also report estimates based on the Kao and Chiang (2000) pooled version of FM-OLS estimation. However, their estimation method is based on the assumption of cross-sectional independence. I do not see how this is now consistent with the strong case for cross-sectional dependence in RER panels, which you also agree with in the discussion and by using a test that allows for cross-sectional dependence. Are these two decisions not contrasting (if not even contradicting) each other? I do believe, compare again Wagner and Hlouskova (2004) or Wagner (2005, 2007) that panel cointegration estimates of the BS effect need to be studied with estimation methods allowing for cross-sectional dependence. This is a major drawback of the empirical BS literature that should be overcome.

You also report bootstrap inference for the FM-OLS coefficients, noting that (under the Kao and Chiang assumptions) the t-statistics will be asymptotically, with first T tending to infinity followed by N tending to infinity) normally distributed, but that this approximation may be poor in small samples. Yes, clearly the approximation will be very bad, due to in particular the short time dimension. But the question is, how do you obtain consistent bootstrap t-values? I guess under the assumption of cross-sectional independence it might be feasible (maybe again based on Park and Swensen), but you allow (in the testing step) for common factors? Hence, it is unclear to me how these bootstrap t-values behave?

If you have, however, consistent bootstrap procedures for cross-sectionally dependent panels, then these could also be used to perform bootstrap inference in the

testing step. I believe that this is a very important open issue. If you do indeed know of and use bootstrap methods that are able to cope with all these issues, then detailed references, a detailed list of assumptions and a thorough description would be very useful for the literature (or for this reviewer at least).

Note that as in the single country case the short time dimension of your sample will – even if all other necessary assumptions were to hold – impinge the performance, as country specific long-run covariance estimates are required and estimated from 13 observations.

All the same econometric problems occur in the analysis of what you call second step of the BS hypothesis.

I believe the above discussion has made clear that I see several important unresolved issues in your econometric analysis. These resort to both the usage of mutually not necessarily consistent methods (tests that allow for cross-sectional dependence but estimators that do not) and also with respect to the performance of the methods used for the sample at hand as well as the theoretical properties of bootstrap algorithms in the context of nonstationary panels with common factors. Before these issues are resolved it is for me essentially impossible to judge the validity and hence relevance of the empirical analysis.

3 Minor Issues

I believe that in you will go through the paper carefully again, since e.g. some of the references are not precise. Matthew Canzonery is actually Matthew Canzoneri (page 4, second paragraph), or the ECB working paper of Anindya Banerjee and Josep Lluís Carrion-i-Silvestre is actually dated 2006 not 2007. There are also some other typos left in the paper. But, clearly, all these things will be detected by a careful proof reading so that I do not have to come up with a full list as a reviewer.

References:

Banerjee, A. and J.L. Carrioni-i-Silvestre (2006). *Cointegration in Panel Data with Breaks and Cross-Section Dependence*. ECB Working Paper No. 591.

Baumol, W. and W. Bowen (1966). *Performing Arts: The Economic Dilemma*. New York, 20th Century Fund.

Hlouskova, J. and M. Wagner (2006). *The Performance of Panel Unit Root and Stationarity Tests: Results from a Large Scale Simulation Study*. Econometric Reviews **25**, 85 – 116.

Kao, C. and M.-H. Chiang (2000). *On the Estimation and Inference of a Cointegrated Regression in Panel Data*. In Baltagi, B.H. (Ed.) *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, Elsevier, Amsterdam.

Park, J.Y. (2003). *Bootstrap Unit Root Tests*. Econometrica **71**, 1845 – 1895.

Phillips, P.C.B. and B. Hansen (1990). *Statistical Inference in Instrumental Variables Regression with I(1) Processes*. Review of Economic Studies **57**, 99--125.

Swensen, A.R. (2006). *Bootstrap Algorithms for Testing and Determining the Cointegration Rank in VAR Models*. Econometrica **74**, 1699 –1714.

Wagner, M. (2005). *The Balassa-Samuelson Effect in 'East & West': Differences and Similarities*, Jahrbuch für Wirtschaftswissenschaften/Review of Economics **56**, 230 – 248.

Wagner, M. (2007). *On PPP, Unit Roots and Panels*. Forthcoming in Empirical Economics.

Wagner, M. and J. Hlouskova (2004). *What's Really the Story with this Balassa-Samuelson Effect in the CEECs?* Working Paper 04-16, Department of Economics, University of Bern.

Wagner, M. and J. Hlouskova (2007). *The Performance of Panel Cointegration Methods: Results from a Large Scale Simulation Study*. Forthcoming in Econometric Reviews.