Effects of data aggregation on the power of tests for a unit root

A simulation study

In Choi

The Ohio State University, Columbus OH, USA

Received 17 June 1992
Accepted 15 October 1992

This paper studies the effects of data aggregation on the powers of the Phillips–Perron and augmented Dickey–Fuller tests for a unit root by simulation. It is shown that using the data generated by aggregating subinterval data results in lower powers of the tests. In addition, especially for the aggregate data, the Philpp–Perron tests appear to be more powerful than the augmented Dickey–Fuller test in finite samples, according to our experimental format.

1. Introduction

There has been much research regarding the presence of a unit root in macroeconomic time series since the work by Nelson and Plosser (1982). Most articles examining the presence of a unit root have applied different methods, Bayesian or non-Bayesian, to the Nelson–Plosser data set which is a collection of annual data for fourteen U.S. macroeconomic time series, and have not been able to reject the null of a unit for most series. However, because it is commonly acknowledged that conventional unit root tests are not powerful in finite samples, classical test results using these annual data (typically with sample size less than 100) have not been given much credence. By contrast, Bayesian approaches have provided some results favoring trend-stationarity for the Nelson–Plosser data, but these results are known to be sensitive to the selected priors (see articles in volume 6 of Journal of Applied Econometrics). Given these empirical results, it appears that there is no consensus regarding the presence of a unit root for annual macroeconomic time series.

One of the prevalent perceptions in testing for a unit root is that using monthly or quarterly data does not provide more credible evidence regarding the presence of a unit root. Shiller and Perron (1985) have provided a rationale for this perception; that is, increasing sampling frequencies do not result in higher powers of unit root tests, and therefore test results using monthly or quarterly series other than annual series should be taken with caution as well. However, as discussed in Choi (1992), Shiller and Perron’s simulation results support such an argument only when time is measured in years. When time is measured either in months or in quarters, using annual data...
results in lower powers of tests, as shown in Shiller and Perron. Further, while sampling frequencies are relevant to stock data (e.g., monetary aggregates), they are not to flow data (e.g., GNP), because annual flow data are generated by aggregating subinterval data in most cases, not by increasing sampling frequencies. In fact, increasing sampling frequencies would not affect the flow data collected over a given period of time, except that it is likely to provide more accurate ones.

The main purpose of this paper is to study how conventional unit root tests perform when we use the data generated by aggregating subinterval data rather than the subinterval data themselves. Because the coefficient for the AR(1) representation of the subinterval data is closer to 1 than that of the aggregate data, one might conjecture that using the subinterval data would not increase the powers of unit root tests. However, contrary to this conjecture, our simulation results will show that using the subinterval data results in higher powers. We will use Phillips and Perron's (1988; PP hereafter) \( Z(\hat{\alpha}) \) and \( Z(t) \) tests and the augmented Dickey–Fuller (ADF hereafter) test for simulation, since these are most often used in applications. Using the simulation results, we will also compare the performance of the PP and ADF tests in finite samples. It will be shown that the PP tests are more powerful than the ADF test in finite samples for aggregate data.

This paper is planned as follows. Section 2 reports simulation results. Section 3 concludes with a summary and further remarks.

2. Simulation results

In this section, we report the simulation results for the finite sample powers of the PP and ADF tests for the null of a unit root with drift against the alternative of trend-stationarity. We generated data according to the model

\[
y_t = \alpha y_{t-1} + e_t \quad (t = 1, 2, \ldots, T),
\]

where \( e_t \sim \text{iid } \mathcal{N}(0, 1) \), \( T = 400 \) and \( y_0 = 0 \). We used the GAUSS subroutine \text{RNDNS} for random number generation. Aggregating \( y_t \), we generated \( x_s \) as follows:

\[
x_s = y_{4s} + y_{4s-1} + y_{4s-2} + y_{4s-3} \quad (s = 1, 2, \ldots, S).
\]

Note that \( S = T/4 = 100 \), and that \( x_s = \alpha^4 x_{s-1} + u_s \), where \( u_s \) is a moving average process. If we consider \( \{y_t\} \) to be a quarterly data, \( \{x_s\} \) can be regarded as an annual data generated by aggregation.

The PP \( Z(\hat{\alpha}) \) and \( Z(t) \) tests were calculated by using the regression model

\[
z_t = \hat{\mu} + \hat{\beta} t + \hat{\alpha} z_{t-1} + \hat{\epsilon}_t, \quad z_t = y_t \text{ or } x_s,
\]

and the ADF \( t \)-test by using the model

\[
z_t = \hat{\mu} + \hat{\beta} t + \hat{\alpha} z_{t-1} + \sum_{k=1}^{I} \hat{\gamma}_k \Delta z_{t-k} + \hat{\nu}_t, \quad z_t = y_t \text{ or } x_s,
\]

where \( \Delta z_t = z_t - z_{t-1} \). The reader is referred to Phillips and Perron (1988) and Said and Dickey (1984) for the definitions of these tests. For the PP tests, we used the quadratic spectral window for the long-run variance estimation, following suggestions in Andrews (1991). First few observations were used as initial variables for these regression models. Hence, for the PP tests \( T = 399 \) and
Table 1
Empirical power of unit root tests for aggregate and subinterval data. \(^{a,b,c}\)

<table>
<thead>
<tr>
<th></th>
<th>(\alpha = 0.90)</th>
<th></th>
<th>(\alpha = 0.95)</th>
<th></th>
<th>(\alpha = 0.99)</th>
<th></th>
<th>(\alpha = 1.0)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(l = 2)</td>
<td>(l = 3)</td>
<td>(l = 4)</td>
<td>(l = 5)</td>
<td>(l = 2)</td>
<td>(l = 3)</td>
<td>(l = 4)</td>
</tr>
<tr>
<td>ADF</td>
<td>agg. 0.77</td>
<td>0.67</td>
<td>0.55</td>
<td>0.46</td>
<td>0.35</td>
<td>0.33</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>sub. 0.99</td>
<td>0.98</td>
<td>0.97</td>
<td>0.95</td>
<td>0.56</td>
<td>0.52</td>
<td>0.49</td>
</tr>
<tr>
<td>Z((\alpha))</td>
<td>agg. 0.94</td>
<td>0.93</td>
<td>0.91</td>
<td>0.89</td>
<td>0.39</td>
<td>0.39</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>sub. 1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.68</td>
<td>0.68</td>
<td>0.69</td>
</tr>
<tr>
<td>Z((t))</td>
<td>agg. 0.94</td>
<td>0.94</td>
<td>0.92</td>
<td>0.91</td>
<td>0.39</td>
<td>0.40</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>sub. 1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.63</td>
<td>0.63</td>
<td>0.64</td>
</tr>
</tbody>
</table>

\(^{a}\) agg.: empirical power for aggregate data; sub.: empirical power for subinterval data.

\(^{b}\) DGP: \(y_t = \alpha y_{t-1} + \epsilon_t, \epsilon_t \sim \text{iid} \, N(0, 1), y_0 = 0 (t = 1, \ldots, 400)\), for subinterval data;

\(^{c}\) \(x_t = y_{s_t} + y_{s_t-1} + y_{s_t-2} + y_{s_t-3} (s = 1, \ldots, 100)\), for aggregate data.

\(^{c}\) Based on 2,000 iterations.
S = 99; and for the ADF test, T = 400 − l and S = 100 − l. This experimental format corresponds with empirical practices where we usually use first few observations as initial variables rather than estimating them by statistical methods.

Generating \( \{y_t\}_{t=1}^{400} \) and \( \{x_t\}_{t=1}^{100} \) 2,000 times, we calculated the empirical powers of the PP and ADF tests, which are reported in table 1. Nominal size was set at 0.05, and critical values were taken from Fuller (1976). Note that \( l \) for the PP tests denotes the lag length for the long-run variance estimation. When \( \alpha = 1 \), we find that empirical sizes are reasonably close to 0.05 for all the tests and for \( \{y_t\} \) and \( \{x_t\} \). Thus, the empirical power calculations reported here are deemed to be close to the actual powers. When \( |\alpha| < 1 \), the PP and ADF tests appear to be substantially more powerful for the subinterval data \( \{y_t\} \) than for the aggregate data \( \{x_t\} \). This is noticeable particularly when \( \alpha = 0.95 \). When \( \alpha = 0.90 \), the difference in the powers of the ADF test for \( \{y_t\} \) and \( \{x_t\} \) is substantial, while those for the PP tests do not seem to be so. These results imply that we need to use subinterval data rather than aggregate data in order to obtain higher powers of the PP and ADF tests. Needless to say, the stochastic nature of the aggregate data can be deduced from that of the subinterval data.

We can also compare the performance of the PP and ADF tests given our simulation results. For the aggregate data \( \{x_t\} \), the PP tests appear to be more powerful than the ADF test, particularly when \( \alpha = 0.90 \) (the empirical power for the ADF test is 0.77 with \( l = 2 \), while those for the PP test are 0.94 with \( l = 2 \)). For the subinterval data \( \{y_t\} \), the PP tests also appear to be more powerful than the ADF test, especially when \( \alpha = 0.95 \). But this power difference is not as noticeable as for the aggregate data. These results are consistent with Phillips and Perron’s which show that the PP tests are more powerful in finite samples than the ADF test when errors follow the MA(1) process.

3. Summary and further remarks

We have studied the finite sample performance of the PP and ADF tests for subinterval and aggregate data. The major findings are:
(i) Using aggregate data for the PP and ADF tests results in lower powers of the tests than using subinterval data. This implies, for example, that the test results based on monthly flow data are more reliable than those based on the aggregate annual flow data. Of course, monthly and quarterly data are subject to seasonal variations and tend to be less accurate, but analyzing the effect of these on the power of unit root tests was not the purpose of this paper.
(ii) The PP tests show better finite sample performance than the ADF test for aggregate data, according to our experimental format. However, we need more experiments with different initial variables and different data-generating schemes for the subinterval data \( \{y_t\} \) in order to fully compare the performance of the PP and ADF tests for aggregate data.

References

Choi, I., 1992, Residual based test for the null of stationarity with application to U.S. macroeconomic time series, Working paper (The Ohio State University, Columbus, OH).