UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Using Panel Data to Construct Simple and Efficient Unit Root Tests in the Presence of GARCH

Joakim Westerlund and Paresh Narayan

EFMD
EQUIS
ACCREDITED

Using Panel Data to Construct Simple and Efficient Unit Root Tests in the Presence of GARCH [*]

Joakim Westerlund[†]
University of Gothenburg
Sweden

Paresh Narayan
Deakin University
Australia

September 10, 2009

Abstract

In search for more efficient unit root tests in the presence of GARCH, some researchers have recently turned their attention to estimation by maximum likelihood. However, although theoretically appealing, the new test is difficult to implement, which has made it quite uncommon in the empirical literature. The current paper offers a panel data based solution to this problem.

Keywords: Panel Data; Unit Root Tests; GARCH.

JEL Classification: C23; G00.

1   Introduction

The prevalence of GARCH effects in financial time series has recently led many authors to consider its interaction with unit root testing, which constitutes a major cornerstone of all types of time series econometric research. The resulting body of research can be divided into two broad categories.

The first category consists of a number of studies aimed at analyzing the small-sample behavior of existing unit root tests when applied to time series contaminated by GARCH. Naturally, due to its wide-spread empirical use and its central role as a theoretical benchmark, the least squares based t-test of Dickey and Fuller (1979), henceforth denoted τ_{LS}, has

---

[†]Corresponding author: Department of Economics, University of Gothenburg, P. O. Box 640, SE-405 30 Gothenburg, Sweden. Telephone: +46 31 786 5251, Fax: +46 31 786 1043, E-mail address: joakim.westerlund@economics.gu.se.

been subject to more scrutiny than any other test.[1] For example, Kim and Schmidt (1993) use simulation experiments to examine the impact of GARCH on the size accuracy of the test. They find that although there is a tendency for the test to become oversized when the GARCH effect increases, the problem is usually not very serious, which is also what others have found. Of course, this is not totally unexpected since like most tests around, τ_{LS} is asymptotically invariant with respect to heteroskedasticity, and therefore also with respect to GARCH. In other words, although the additional information contained in the GARCH structure of the errors is basically ignored, this has no effect on the asymptotic distribution of τ_{LS}.

Thus, while the results regarding the issue of neglected GARCH have not been very controversial, the second body of research is potentially more fruitful in the sense that it centers around the development of new tests that make full use of the information contained in the GARCH errors. The idea of joint maximum likelihood estimation of both the autoregressive unit root and GARCH parameters was introduced by Seo (1999). He shows that in contrast to τ_{LS}, the asymptotic distribution of the corresponding test based on maximum likelihood, henceforth denoted τ_{ML}, is not nuisance parameter free, and that it depends on the strength of the GARCH effect. He also shows that as the GARCH effect increases, the asymptotic distribution of τ_{ML} moves away from the conventional Dickey and Fuller (1979) distribution and towards the standard normal.

The problem is that while theoretically appealing in the sense that it is efficient, τ_{ML} is also quite difficult to implement, which is probably the most important reason for why it is almost never used in empirical work. In particular, not only is the test statistic difficult to compute, but there is also the numerical optimization of the likelihood function, which is computationally very burdensome, especially when compared to simple least squares estimation. The dependence of the asymptotic distribution on the strength of the GARCH effect is yet another inconvenience, which makes it necessary to tabulate critical values for all possible values of this parameter.

The discussion so far suggests that the choice of test basically boils down to a trade-off between efficiency and ease of implementation. While τ_{LS} is computationally convenient, it is also inefficient, which makes τ_{ML} an interesting alternative.

In this paper we make an attempt to combine only the good aspects of both tests, while at

---

[1]See Li *et al.* (2002) for an overview of the literature.

the same time eliminating all the bad aspects. In particular, the idea is to apply least squares, but not just to one time series but to a panel of multiple series. This means that although the GARCH structure of the errors is not exploited, because of the added information that all series have a unit root under the null, efficiency is still expected to be high. This is verified in small samples using both real and simulated data.

The rest of the paper is organized as follows. Section 2 introduces our panel based approach, while Section 3 contains the simulation study. Section 4 then reports the empirical results, while Section 5 concludes.

2 Testing for a unit root with GARCH

In this section, we begin with a brief account of the conventional unit root testing approach when GARCH is suspected, and then we go on to discuss our panel proposal.

2.1 Tests based on single series

Consider for simplicity the case when the time series variable y_t, observable for $t = 1, ..., T$, is generated by the following autoregression

$$y_t = \rho y_{t-1} + u_t, \tag{1}$$

where u_t is assumed to be a mean zero and serially uncorrelated error term. Let us further assume that the conditional variance of this error can be written as

$$\sigma_t^2 = \alpha + \phi u_{t-1}^2 + \beta \sigma_{t-1}^2. \tag{2}$$

That is, we assume that u_t is generated according to a first order GARCH model. If α is positive, while ϕ and β are nonnegative such that their sum is less than one, then the unconditional variance of u_t exists and is in fact given by $\sigma^2 = E(\sigma_t^2) = \frac{\alpha}{1-\phi-\beta}$. This is the case considered here. Note also that although there are no deterministic components in (1) and that the GARCH model in (2) is the simplest possible, this is only for simplicity. Nonzero intercept and trend terms, and higher order GARCH effects can be readily accommodated as in Seo (1999). As we demonstrate in Section 5, allowing for serial correlation is just as simple.

The null hypothesis to be tested is formulated as that $H_0 : \rho = 1$, while the alternative hypothesis is that $H_1 : \rho < 1$. As mentioned earlier, the by far most common way of

performing this test is through a simple *t*-test, which is typically based on the least squares estimator $\widehat{\rho}_{LS}$ of ρ, yielding τ_{LS}. But it can of course also be based on other estimators, such as the maximum likelihood estimator $\widehat{\rho}_{ML}$, which yields τ_{ML}.

In discussing the asymptotic null distributions of these tests, it is convenient to let DF and Z denote the Dickey and Fuller (1979) and standard normal distributions, respectively.[2] The relevant distributional result for τ_{LS} has been worked out by Ling *et al.* (2002) among others. They show that

$$\tau_{LS} \Rightarrow DF \ \text{ as } \ T \to \infty,$$

where the symbol $\Rightarrow$ denotes convergence in distribution. Thus, since the asymptotic distribution of τ_{LS} with GARCH is the same as the one with no GARCH, τ_{LS} is said to be asymptotically invariant in this respect.

The corresponding result for τ_{ML} can be found in Seo (1999). It reads

$$\tau_{ML} \Rightarrow \frac{\lambda_Z}{\lambda} \left(\lambda \, DF + \sqrt{1 - \lambda^2} \, Z \right) \ \text{ as } \ T \to \infty,$$

where $\lambda = \frac{1}{\sqrt{\sigma^2 H}}$ with H given by

$$H = \left(E(u_t^4) - 1 \right) \phi^2 \sum_{s=1}^{\infty} \beta^{2(s-1)} E\left(\frac{u_{t-s}^2}{\sigma_t^4} \right) + E\left(\frac{1}{\sigma_t^2} \right),$$

and where λ_Z is λ with $E(u_t^4)$ in H set equal to three, as when u_t is normally distributed.

Some remarks are in order. Firstly, in contrast to τ_{LS}, the asymptotic distribution of τ_{ML} is a mixture of the Dickey and Fuller (1979) test distribution and the standard normal. If there is no GARCH so that ϕ and β are zero, then σ_t^2 reduces to α, and so does σ^2. This means that both λ_Z and λ equals unity, which in turn implies that τ_{ML} goes to DF. At the other end of the scale we have the infinite variance case, in which $\lambda \to 0$ and so τ_{ML} goes to the normal distribution. In the intermediate case when either ϕ or β, or both, are positive, then the asymptotic distribution of τ_{ML} is a mixture of DF and Z.

Secondly, by determining the relative contribution of the DF distribution, λ also determines the relative efficiency of τ_{ML}. If λ is one, τ_{LS} and τ_{ML} are equally efficient, while if λ approaches zero, then τ_{ML} is infinitely more efficient than τ_{LS}, see Ling and Li (1998) for a more detailed discussion. When λ is between zero and one, then it measures the relative efficiency gain of τ_{ML}.

---

[2]These distributions are assumed to be known, and are introduced here without any further discussion. For more details, we make reference to Ling *et al.* (2002).

Thirdly, as pointed out by Seo (1999), there are at least three different versions of τ_{ML}, depending on the choice of variance estimator, which in turn determines its asymptotic distribution. We may use the information matrix, the inverse of the outer product of the gradient or the robust variance estimator. However, since the distributions of these tests are qualitatively very similar, and since their small-sample performance were almost identical, in this paper we focus on the first test based on the information matrix.

Finally, note that in contrast to τ_{LS}, τ_{ML} involves some severe complexity of expression, which in practice translate into a relatively difficult estimation problem. In particular, not only does it require numerical optimization of the likelihood function, but τ_{ML} also calls for consistent estimation of λ and λ_Z.[3] Then there is also the problem that even if λ and λ_Z were known, their presence makes it difficult to obtain exact critical values. As an illustration of this, consider the article of Seo (1999), in which the conventional approach of simulating critical values for a few equally spaced values of λ is adopted. This entails at least two problems. The first is that discretizing λ in this way is generally not appropriate, in which case critical values for a possible continuum of values may be required. The second problem is that, even if λ is discrete and equally spaced, one faces the complicating factor of having to consult a table for the critical values for every choice of λ, which may be a very tedious undertaking in itself.

The above discussion leaves us with an intricate dilemma. On the one hand, when exploiting the additional information contained in the GARCH structure of the errors, we end up with a test that is computationally very costly and borderline impractical. On the other hand, by pretending that there are no GARCH effects, although much simpler, we run the risk of obtaining an inefficient test.

2.2 A test based on panel data

While certainly appealing in many respects, as we have seen τ_{LS} and τ_{ML} also have their limitations, and this section therefore offers an alternative testing approach. The idea is basically to device a test that maintains the invariance and simplicity of τ_{LS}, while at the same time not sacrificing the efficiency of τ_{ML}. We start with the least squares estimator, which is simple. But instead of considering just one time series we consider a panel of multiple series, henceforth indexed $i = 1, ..., N$. Thus, although this means loosing the information

---

[3]See Sections 4 and 5 for more details regarding the implementation of τ_{ML}.

5

contained in the GARCH errors, by effectively increasing the total number of observations from T to NT, this is nevertheless expected to produce a more efficient test.

The panel model that we consider is given by

$$y_{it} = \rho y_{it-1} + u_{it}, \tag{3}$$

where the conditional variance of u_{it} is given by

$$\sigma_{it}^2 = \alpha_i + \phi_i u_{it-1}^2 + \beta_i \sigma_{it-1}^2. \tag{4}$$

Thus, we basically assume here that each of the cross-sectional units i evolves according to (1) and (2). Note that no restrictions are placed on σ_{it}^2, which is permitted to be completely heterogeneous. In fact, the only apparent assumption is that the autoregressive coefficient ρ is equal across i, which is very common when dealing with panel data. The reason for why we have it here is of course that if we are interested in getting an efficient estimator of ρ under the null, then nothing is lost by assuming it to be equal. We also assume that the error u_{it} is uncorrelated across i. But this is only for easy of presentation, and is by no means restrictive, as the test we consider can be easily generalized along the lines of for example Moon and Perron (2004) to allow for cross-sectionally correlated errors, see Section 5 for a more thorough discussion.

Let τ_{pLS} denote the t-test based on $\widehat{\rho}_{pLS}$, the pooled least squares estimator of ρ. The idea here is to propose this test as an effective and simple alternative to τ_{ML} and τ_{LS} when testing H_0 against H_1.[4] The essential insight behind the efficiency of τ_{pLS} is that while $\widehat{\rho}_{ML}$ is more efficient than $\widehat{\rho}_{LS}$ when applied to a single series, its rate of consistency under the null is nevertheless the same, T. By contrast, $\widehat{\rho}_{pLS}$ converges to ρ at rate $\sqrt{NT}$, and is therefore infinitely more efficient than both $\widehat{\rho}_{ML}$ and $\widehat{\rho}_{LS}$. Formally, we have the following relationship

$$\frac{\text{var}\left(\widehat{\rho}_{pLS} - 1\right)}{\text{var}\left(\widehat{\rho}_{ML} - 1\right)} = O_p\left(\frac{1}{NT^2}\right)O_p(T^2) = O_p\left(\frac{1}{N}\right)$$

which of course goes to zero as $N \rightarrow \infty$, reflecting the fact that the relative efficiency of $\widehat{\rho}_{pLS}$ increases with N.[5]

---

[4]Except for the fact that we have chosen not to consider serial correlation, τ_{pLS} is exactly the same test considered by Levin *et al.* (2002), see Section 5 for a more thorough discussion.

[5]For any real r, $y_T = O_p(T^r)$ indicates that y_T is at most of order T^r in probability, which simply means that $\frac{y_T}{T^r}$ converges in distribution as T grows.

3 Simulation results

In this section, we compare the small-sample performance of τ_{pLS} with that of τ_{ML} and τ_{LS} by means of simulations. The simulation design used for this purpose is taken from Seo (1999), and consists of creating $1,000$ panels using (3) and (4) to generate the data, where u_{it} is drawn from a normal distribution with mean zero and variance σ_{it}^2. The first 100 time series observations for each i are then discarded to avoid possible initial value effect.[6] As in Seo (1999), we set $\alpha_i = 1$ for all i, and instead consider varying the values of the GARCH parameters ϕ_i and β_i, which are both assumed to be equal across i. The homogenous values of these parameters are henceforth denoted ϕ and β, respectively. For the size simulations, $\rho = 1$ while for the power simulations, $\rho < 1$.

As indicated before, τ_{ML} is rather difficult to compute, not only because it requires an estimator of λ, the strength of the GARCH effect, but also because it is based on numerical optimization of the likelihood function.[7] In this paper, we use the GAUSS optimization library OC, which makes it possible to optimize while at the same time satisfying the various inequality restrictions needed for σ_{it}^2 to be well-defined. We use the Newton–Raphson algorithm with numerical derivatives, and the standard error of $\widehat{\rho}_{ML}$ is computed as described earlier, using the information matrix. To start up the estimation, we used the true parameter values, which means that the results obtained for τ_{ML} are probably somewhat better than the ones actually achievable in practice. In other words, the results reported here probably overstate the performance of τ_{ML}.

The results from the performance under the null when $\rho = 1$ are summarized in Table 1, which reports the size at the 5% level for each test, as well as the mean bias and root mean squared error, or MSE, of each estimator. The first thing to notice is that the performance of the pooled estimator $\widehat{\rho}_{pLS}$ is uniformly better than that of $\widehat{\rho}_{ML}$ and $\widehat{\rho}_{LS}$, both in terms of bias and MSE. But the performance is not only better, it is vastly superior. In fact, as seen from the table, it is not unusual for the performance of $\widehat{\rho}_{pLS}$ to be many hundred times better than that of the other two. The results in terms of bias are particularly impressive, and suggest that $\widehat{\rho}_{pLS}$ is essentially unbiased, while $\widehat{\rho}_{ML}$ and $\widehat{\rho}_{LS}$ are downward biased.

Another interesting result is that unless ϕ is positive, there is basically no difference in performance between $\widehat{\rho}_{ML}$ and $\widehat{\rho}_{LS}$. Thus, according to this, it is the ARCH parameter ϕ that

---

[6]Without loss of generality, we set the initial value of σ_{it}^2 to one, while those of y_{it} and u_{it} are set equal to zero.
[7]The parameter λ is estimated as in Seo (1999) and the critical values are taken from Table 1 of the same paper.

has the most adverse effect on $\widehat{\rho}_{LS}$.

The overall best performance in terms of size accuracy is obtained by using τ_{LS}, which is a little surprising in view of the fact that it is based on the worst preforming estimator. However, the difference is usually not very large, with the other two tests performing only slightly worse. Thus, in terms of size accuracy, there is really not much difference between the tests.

Consider next the results for the performance under the stationarity alternative, which are reported in Table 2 for the case when $\rho = 0.99$, and in Table 3 for the case when $\rho = 0.95$.[8] It is seen that although the performance of $\widehat{\rho}_{pLS}$ is generally not as good as under the null, the results are still very encouraging. In particular, $\widehat{\rho}_{pLS}$ still outperforms its competitors. The fact that both the bias and the MSE tend to be somewhat higher in Tables 2 and 3 than in Table 1 is well expected, as the rate of consistency under the alternative is $\sqrt{T}$ slower than under the null.

In agreement with the good performance of $\widehat{\rho}_{pLS}$, we also see that τ_{pLS} generally turns out to be the most powerful test, and that the relative power advantage increases as ρ gets closer to its hypothesized value under the null.[9] We also see that the power of τ_{pLS} seems to be largely unaffected by the parametrization of the GARCH model. This leads us to the conclusion that our panel proposal should be well suited for financial applications, which are likely to involve data that are heteroskedastic and nonstationary under the null, and highly persistent under the alternative. The fact that the power advantage appears to be larger when N and T are small, a typical scenario in applied work, reinforces this conclusion. The lowest power is always obtained by using τ_{LS}, with the power of τ_{ML} lying somewhere in between.

In summary, we find that τ_{pLS} show higher power than the other tests considered and, at the same time, maintain the nominal size well in small samples. Since the power advantage is particularly striking in small and highly autoregressive panels, this leads us to the

---

[8]Since the size distortions were so marginal, we only present the results for the raw power. Some results based on the size-adjusted power are available upon request from the corresponding author.

[9]Although the assumption of a homogenous ρ is irrelevant under the null that all N units have a unit root, as Moon *et al.* (2007) shows, this is not the case under the alternative of stationarity. In particular, if ρ is homogenous, then τ_{pLS} uniformly most powerful, while if it is heterogeneous, then the power will depend on the extent of the heterogeneity. However, as long as the average autoregressive parameter is larger than one, then τ_{pLS} should be more powerful than the other tests, which is also what we find in our simulations. In other words, although the magnitude of the power gain will generally depend on the heterogeneity of ρ, the relative ordering of the test should still be the same. Because of space constraints, we therefore only report the results for the case when ρ is homogeneous.

conclusion that the new tests should be well suited for financial applications.

4   Empirical results

In this section, we illustrate empirically the issues previously discussed using three common examples. Namely, the purchasing power parity (PPP) hypothesis, the Fisher hypothesis and the efficient market hypothesis. We begin with a brief description of the data, and then we go on to discuss the implementation of the tests and the results obtained.

4.1   Data

Data on stock prices, consumer price indices and real exchange rates are obtained from Bloomberg. All data are monthly.

The stock price data covers 20 OECD countries across the period January 1988 to April 2003. This means that there is a total of 3,680 observations available for testing the efficient market hypothesis, stating that prices should be nonstationary and hence unpredictable. The consumer price index data cover 23 OECD countries and stretch the period January 1991 to August 2003. Thus, the number of observations available for examining the Fisher hypothesis is 3,496. More precisely, by testing if prices are stationary, we can determine the appropriateness of the conventional way in which this hypothesis is tested, namely as a cointegration test between inflation and nominal interest rates. If prices, and hence inflation, are stationary, then this test is no longer valid. The exchange rate panel covers 20 OECD countries between January 1981 to May 2003, which means that there is no less than 5,380 observations available for testing the PPP hypothesis of a stationary real exchange rate. All three variables are expressed in logs.

To foreshadow the more formal treatment in the next subsections, we begin with a graphical inspection of each of the three variables, which are plotted in Figures 1 to 3. The first thing to note is the periodically high volatility, which makes it is very difficult to say whether the series are in fact stationary or not. For example, stock prices seem to be more volatile in the first half of the sample than in the second. This is not very surprising given that the post-1996 period has been a turbulent one, with the market exposed to the Asian financial crisis and various terrorist attacks around the world. While much less volatile, the consumer price indices plotted in Figure 2 seem to follow the same upwards trend as the stock prices. The exchange rate series plotted in Figure 3 are generally more volatile in the beginning and

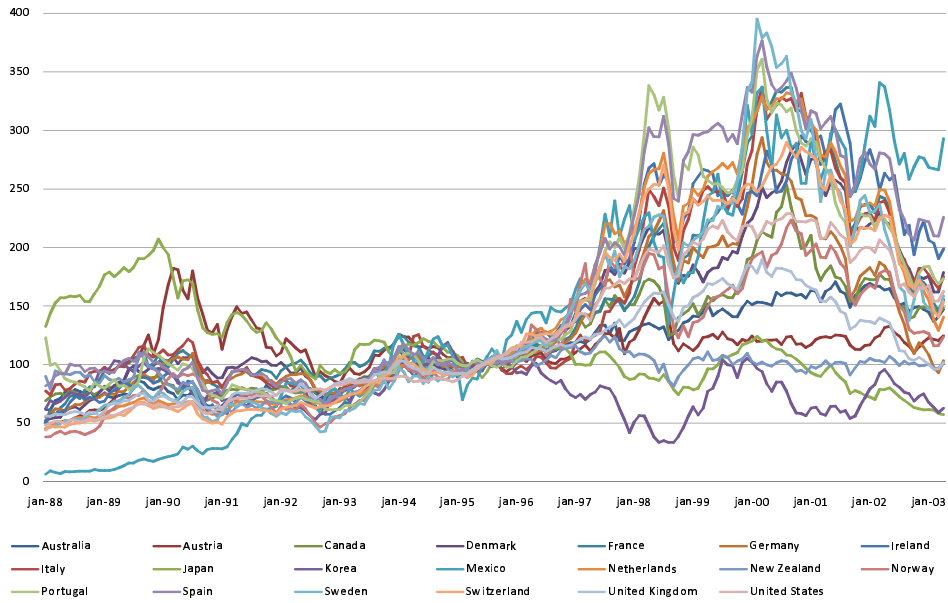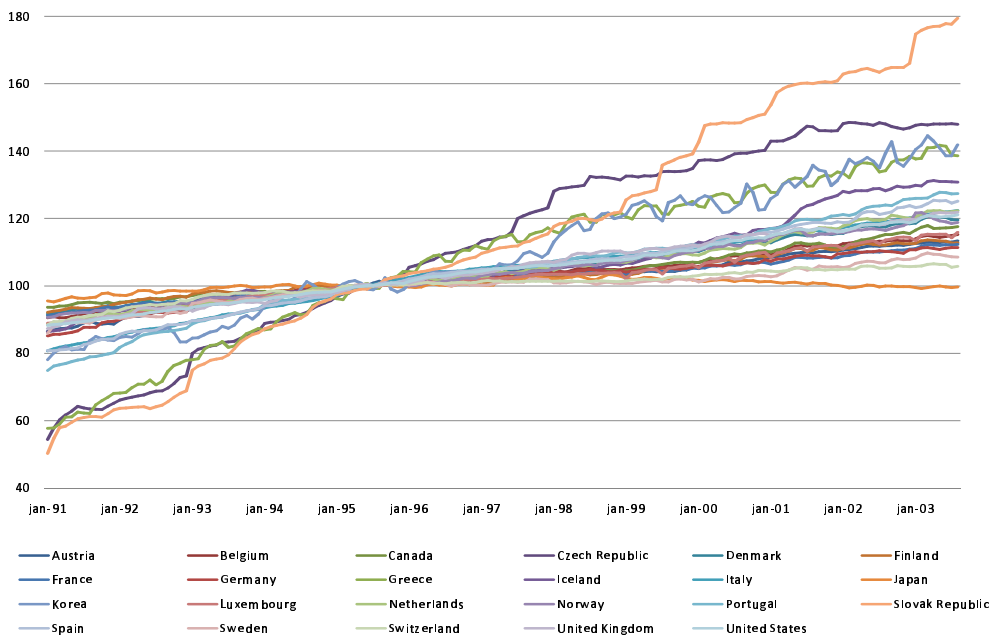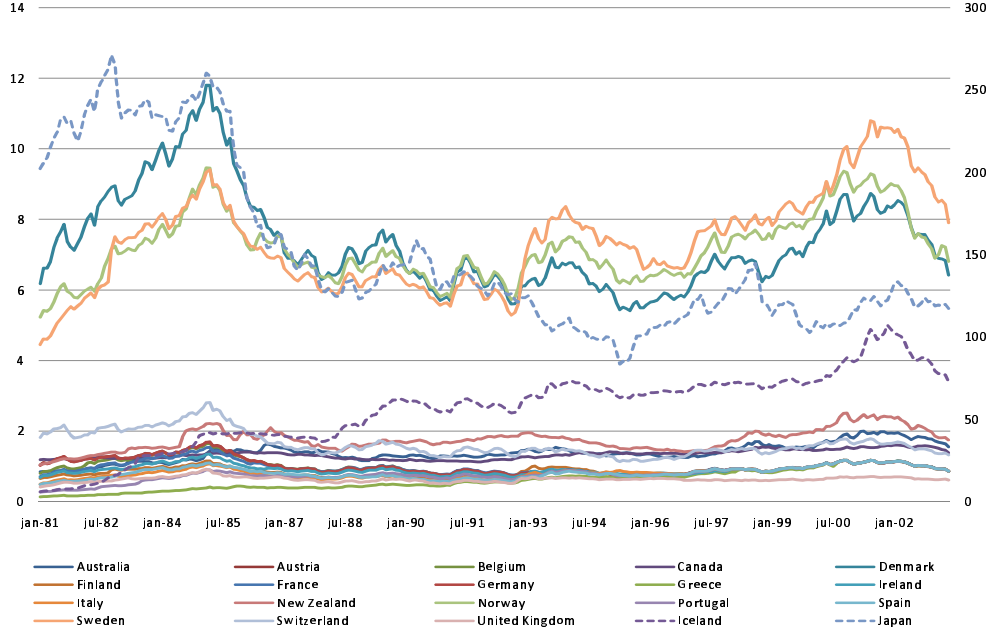Figure 1: Share prices.

Figure 2: Prices.

end of the sample, while it is relatively stable in the middle.[10]

_____
[10]We use dotted lines to indicate that the exchange rate of Iceland and Japan are measured along the secondary right-hand side axis.

Figure 3: Exchange rates.

This evidence suggests that the presence of GARCH cannot be excluded, and that in turn makes it difficult to determine order of integration of the data. We also see that there is a strong tendency for the series to move together, both across time as well as across countries. Dependence is therefore an important aspect to consider now when we proceed to discuss the implementation of the tests.

4.2 Implementation

So far we have been focusing exclusively on GARCH, and we have said nothing about the treatment of other empirically important features like deterministic terms, and serial and cross-sectional dependence.

Accounting for the presence of nonzero intercept and trend terms, and serial correlation is particularly easy, and involves replacing (3) with its augmented version

$$y_{it}^{\circ} = \rho y_{it-1}^{\circ} + \sum_{s=1}^{p_i} \gamma_{is} \Delta y_{it-s}^{\circ} + u_{it}, \tag{5}$$

where y_{it}° denotes the least squares residual from regressing y_{it} onto an appropriate vector of deterministic components, typically representing a constant and trend. Given that the error u_{it} is uncorrelated across both i and t, estimation of (5) can proceed exactly as explained

11

before, using maximum likelihood or individual or pooled least squares.[11] In order to do so, however, we first need to determine p_i, the lag order of each unit, which in this paper is done by minimizing the Schwarz Bayesian criterion. Thus, in this setup τ_{LS} represents the usual augmented Dicky and Fuller (1979) test, while τ_{pLS} represents the t_δ^* test of Levin *et al.* (2002), which can be seen as a panel generalization of the former. While not necessary for τ_{LS}, in addition to p_i, in order to fully eliminate the effects of serial correlation, t_δ^* also requires a choice of bandwidth, which in this paper is accomplished using the automatic rule of Newey and West (1994).

Although replacing (3) with (5) takes care of any correlation across time, there is still the problem that u_{it} might be correlated across units. This has two effects. The first one is not so serious in the sense that it just amounts to a loss of information. The second effect, however, is more problematic, and amounts to an asymptotic bias in the limiting distribution of the tests. Although usually disregarded in the time series literature, in the panel literature this bias has led to the development of several new tests that are robust not only with respect to serial correlation but also with respect to cross-sectional correlation.[12]

To be able to understand the basic idea behind these tests, suppose that u_{it} can be decomposed in the following way

$$u_{it} = \lambda_i' f_t + e_{it}, \tag{6}$$

where f_t is a vector of unobserved common factors, which could represent oil-price shocks or any other feature affecting y_{it} that is common across i, while e_{it} is assumed to be completely idiosyncratic. The reason for having f_t here is to model the cross-sectional dependence in u_{it}, whose extent is determined by λ_i, a vector of loading parameters that measure the effect of the common factors. This is easily seen by writing

$$E(u_{it} u_{jt}) = \lambda_i' \text{var}(f_t) \lambda_j \text{ for } i \neq j.$$

Thus, if λ_i is zero, then there is no correlation, whereas if λ_i is nonzero, then u_{it} is cross-sectionally correlated. The tests that we employ here, denoted t_a and t_b, are taken from Moon and Perron (2004), and are based on first using the method of principal components to estimate the factors and their loadings, and then to run (5) based on the de-factored series,

---

[11]Thus, as in Seo (1999), τ_{ML} is based on using least squares to obtain y_{it}°, and then to estimate the remaining parameters using maximum likelihood.

[12]See Breitung and Pesaran (2008) for a recent review of this literature.

which should be asymptotically cross-section uncorrelated. The resulting tests are implemented in the same way as t_δ^*, except for the added difficulty that now we also need to determine the number of factors to use. This is done using the IC_1 information criterion recommended by Bai and Ng (2004).

Thus, if we want to test the unit root null while permitting for both serial and cross-sectional correlation, τ_{pLS} is either t_a or t_b, or both if we want to refer to them jointly.

4.3 Results

We begin by looking at the country-by-country results obtained from the τ_{LS} and τ_{ML} tests. The graphical evidence reported earlier suggests that a constant alone might not be enough to capture the deterministic behavior of the variables, and that there is a need allow for a linear trend. In interest of comparison, however, the results for the model with no trend are also reported. The information contained in Table 4 may be summarized in the following.[13]

Turning first to the results for the model with no trend, we see that the null is rarely rejected. In fact, even if we look at the liberal 10% level, the number of rejections is never larger than seven. At the 5% level, we count at most three rejections for the stock price, one rejection for the consumer price index, and four rejections for the exchange rate. The evidence against the null is therefore weak, at best. For the model that includes a tend, the evidence against the null is even weaker. Indeed, if we look at the 1% level then there is not a single rejection, and there is at most four rejections at the 10% level.

Although the results so far indicate that the evidence against the null is weak, there are at least two problems with this conclusion. Firstly, both τ_{LS} and τ_{ML} are constructed under the assumption that the data are cross-sectionally uncorrelated, which, as noted in the previous section, is unlikely to hold. These tests are therefore not strictly valid. Then there is also the more conventional problem of using repeated testing, which make our results difficult to interpret. As an example, imagine that we are interested in knowing whether we should reject the unit root null for the whole panel at level α. Clearly, it would be a mistake to conclude that we should reject simply because the null is rejected for one of the countries. Doing so would ignore the fact that for a panel of N members, we would expect to reject in αN times even if the null is true.

---

[13]Due to space constraints, we only report the results from the number of rejections of the null hypothesis at conventional levels of significance. The complete set of results can be obtained from the corresponding author upon request.

Fortunately, once the individual significance levels have been computed, it is also possible to use these to produce a more conservative test that is also invariant to the presence of cross-sectional dependency. Indeed, by the Bonferroni inequality we have that α is at most N times the significance level of the individual tests. In other words, we should only reject the null for the full panel if any one country rejects at level of $\frac{\alpha}{N}$ or stronger.

In our case with N close to 20, if we set α to 20%, this means that the null should be rejected if any of the individual tests end up in a rejection at the 1% level. Thus, in view of Table 4, the null of a unit root in the full panel is rejected for the model with a constant but not for the model with a trend.

Consider next the results reported in Table 5 for the three panel tests. If we look at the model without trend, we see that while t_δ^* results in a rejection of the null for all three variables, this is not the case for the other two tests, at least not at the 1% level. In other words, if we disregard t_δ^*, which is likely to be biased due to the cross-sectional correlations, then there is not much evidence against the null. The same picture appears if we look at the model with the trend included, in which case the null cannot be rejected at the 5% level, except for consumer prices.

In sum, if we concentrate on the most appropriate model with a trend, then the results obtained from both the individual and panel tests lead to the same conclusion, namely that the null cannot be rejected, except possibly for the consumer price index. This in turn implies that while there is no evidence of PPP, the efficient market hypothesis is accepted. It also implies that the Fisher hypothesis cannot be inferred by a simple cointegration test between inflation and nominal interest rates.

5   Concluding remarks

The greater usage of financial time series data and the potential impact of GARCH have motivated researchers to be mindful of such effects when testing for unit roots. This means that researchers are faced with a difficult dilemma, either they use the fully efficient but computationally very demanding maximum likelihood based test or else they use the simple but inefficient least squares counterpart.

In this paper, we attempt to ease this tension by offering a simple panel based solution. The idea is that by using panel data, we can increase efficiency without sacrificing the simplicity of least squares. This is verified in a series of simulation experiments, where we show

that the use of panels results in tests that are vastly more efficient and powerful than existing tests based on pure time series. We also show that this holds regardless of whether there are GARCH effects present or not. When we apply our proposal to test for a unit root in consumer price indices, stock prices and real exchange rates, we only find evidence against the unit root null for the first variable.

References

Bai, J., Ng, S. (2004). A PANIC attack on unit roots and cointegration. *Econometrica* **72**, 1127–1177.

Breitung, J., Pesaran, M. H. (2008). Unit roots and cointegration in panels. Forthcoming in Matyas, L., Sevestre, P. (Eds.) *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Kluwer Academic Publishers, Boston.

Dickey, D. A., Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**, 427–431.

Kim, K., Schmidt, P. (1997). Unit root tests with conditional heteroskedasticity. *Journal of Econometrics* **59**, 287–300.

Levin, A., Lin, C. F., Chu, C.-S. J. (2002). Unit root tests in panel data: Asymptotic and finite sample properties. *Journal of Econometrics* **108**, 1–22.

Li, W. K., Ling, S., McAleer, M. (2002). Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys* **16**, 245–269.

Moon, H. R., Perron, B. (2004). Testing for a unit root in panels with dynamic factors. *Journal of Econometrics* **122**, 81–126.

Moon, H. R., Perron, B. (2007). Asymptotic local power of Pooled *t*-ratio tests for unit roots in panels with fixed effects. *Econometrics Journal* **11**, 80–104.

Newey, W., West, K. (1994). Autocovariance lag selection in covariance matrix estimation. *Review of Economic Studies* **61**, 613–653.

Seo, B. (1999). Distribution theory for unit root tests with conditional heteroskedasticity. *Journal of Econometrics* **91**, 113–144.

Table 1: Simulation results under the null hypothesis when $\rho = 1$.

Case	T	N	Size			Bias			Root MSE		
			τ_{ML}	τ_{LS}	τ_{pLS}	$\widehat{\rho}_{ML}$	$\widehat{\rho}_{LS}$	$\widehat{\rho}_{pLS}$	$\widehat{\rho}_{ML}$	$\widehat{\rho}_{LS}$	$\widehat{\rho}_{pLS}$
1	50	10	6.8	5.9	6.5	-1.83	-1.83	-0.09	4.52	4.49	0.50
		20	6.5	5.5	5.7	-1.80	-1.80	-0.04	4.44	4.39	0.32
		40	6.7	5.8	5.6	-1.78	-1.79	-0.02	4.42	4.39	0.21
	100	10	6.1	5.6	6.1	-1.15	-1.15	-0.06	2.58	2.56	0.29
		20	5.8	5.4	7.4	-1.12	-1.12	-0.03	2.54	2.54	0.20
		40	5.8	5.3	6.4	-1.13	-1.13	-0.02	2.54	2.53	0.14
2	50	10	7.8	6.6	7.2	-0.82	-1.98	-0.11	3.08	5.62	0.83
		20	7.7	6.2	9.0	-0.82	-1.95	-0.11	3.78	5.52	1.33
		40	7.6	6.2	10.8	-0.84	-1.96	-0.12	3.77	5.51	1.94
	100	10	6.3	6.8	11.7	-0.42	-1.30	-0.13	1.84	3.44	0.55
		20	6.2	6.5	10.0	-0.41	-1.28	-0.06	2.54	3.32	0.37
		40	6.4	6.4	13.0	-0.41	-1.28	-0.05	1.70	3.38	0.32
3	50	10	8.6	5.7	6.0	-1.92	-1.92	-0.08	4.69	4.57	0.52
		20	9.1	6.0	7.2	-1.91	-1.93	-0.04	4.72	4.64	0.35
		40	8.7	5.8	6.3	-1.91	-1.93	-0.02	4.72	4.67	0.24
	100	10	6.7	5.3	6.8	-1.17	-1.17	-0.07	2.65	2.63	0.33
		20	6.6	5.1	7.3	-1.15	-1.16	-0.03	2.59	2.58	0.22
		40	6.9	5.3	6.1	-1.17	-1.18	-0.02	2.62	2.62	0.14
4	50	10	8.9	6.3	8.4	-1.33	-1.92	-0.12	3.86	4.88	0.60
		20	8.5	6.0	7.9	-1.35	-1.92	-0.07	3.88	4.86	0.38
		40	8.8	6.1	8.7	-1.39	-1.95	-0.03	4.11	5.03	0.36
	100	10	7.1	6.1	8.4	-0.74	-1.24	-0.08	2.05	2.97	0.45
		20	7.3	5.9	6.2	-0.75	-1.22	-0.05	2.08	2.96	0.28
		40	7.3	5.9	9.5	-0.73	-1.23	-0.03	2.03	2.93	0.18

Notes: In Case 1, ϕ and β are set to zero, while in Case 4, they are set to 0.45. In Case 2, $\phi = 0.95$ and $\beta = 0$, while in Case 3, $\phi = 0$ and $\beta = 0.95$. The abbreviations *LS*, *ML* and *pLS* refer to the least squares, maximum likelihood and pooled least squares estimator, respectively. The size results are for a nominal 5% level test, and all bias and root mean squared error, or MSE, results are upscaled by 100.

Table 2: Simulation results under the alternative hypothesis when $\rho = 0.99$.

Case	T	N	Power			Bias			Root MSE		
			τ_{ML}	τ_{LS}	τ_{pLS}	$\widehat{\rho}_{ML}$	$\widehat{\rho}_{LS}$	$\widehat{\rho}_{pLS}$	$\widehat{\rho}_{ML}$	$\widehat{\rho}_{LS}$	$\widehat{\rho}_{pLS}$
1	50	10	10.0	8.8	48.0	−2.36	−2.37	−0.14	5.42	5.39	0.74
		20	10.2	8.9	75.7	−2.39	−2.40	−0.09	5.36	5.33	0.52
		40	10.4	9.1	95.7	−2.38	−2.39	−0.05	5.36	5.34	0.34
	100	10	11.1	10.3	78.0	−1.41	−1.41	−0.12	3.10	3.09	0.54
		20	11.2	10.4	97.0	−1.44	−1.44	−0.05	3.14	3.13	0.34
		40	11.0	10.2	100.0	−1.43	−1.43	−0.05	3.11	3.09	0.25
2	50	10	23.6	12.1	55.3	−1.10	−2.67	−0.40	3.82	6.75	1.78
		20	24.1	12.5	76.5	−1.14	−2.72	−0.24	4.54	6.88	1.95
		40	24.5	12.5	90.5	−1.15	−2.69	−0.17	3.86	6.80	1.44
	100	10	34.9	14.5	76.8	−0.52	−1.72	−0.28	2.00	4.08	1.20
		20	34.9	14.7	90.3	−0.56	−1.77	−0.23	3.04	4.42	1.06
		40	35.1	14.6	96.2	−0.54	−1.76	−0.21	2.16	4.49	1.68
3	50	10	12.7	8.7	45.2	−2.45	−2.47	−0.18	5.56	5.50	0.79
		20	12.8	8.5	70.8	−2.35	−2.39	−0.06	5.42	5.30	0.52
		40	12.6	8.6	92.3	−2.33	−2.38	−0.02	5.43	5.32	0.35
	100	10	12.6	10.2	80.0	−1.40	−1.43	−0.14	3.07	3.07	0.53
		20	13.0	10.6	97.7	−1.46	−1.48	−0.08	3.21	3.19	0.35
		40	12.3	10.0	100.0	−1.41	−1.44	−0.04	3.12	3.11	0.24
4	50	10	17.4	10.1	52.4	−1.79	−2.51	−0.27	4.81	5.94	0.98
		20	17.2	10.7	72.3	−1.81	−2.52	−0.09	4.89	5.89	0.75
		40	17.5	10.8	91.8	−1.86	−2.59	−0.08	5.23	6.05	0.50
	100	10	21.3	12.5	79.1	−0.94	−1.56	−0.15	2.54	3.55	0.66
		20	21.2	13.0	94.2	−0.97	−1.62	−0.11	2.57	3.64	0.53
		40	22.2	13.2	98.2	−1.02	−1.67	−0.08	2.61	3.72	0.42

Notes: The power results are for a nominal 5% level test, and have not been adjusted for size.
See Table 1 for an explanation of the remaining features.

Table 3: Simulation results under the alternative hypothesis when $\rho = 0.95$.

Case	T	N	Power			Bias			Root MSE		
			τ_{ML}	τ_{LS}	τ_{pLS}	$\widehat{\rho}_{ML}$	$\widehat{\rho}_{LS}$	$\widehat{\rho}_{pLS}$	$\widehat{\rho}_{ML}$	$\widehat{\rho}_{LS}$	$\widehat{\rho}_{pLS}$
1	50	10	21.3	19.3	99.9	−3.00	−3.00	−0.27	7.14	7.03	1.56
		20	21.3	19.4	100.0	−3.14	−3.14	−0.15	7.30	7.22	1.07
		40	21.4	19.3	100.0	−3.06	−3.07	−0.03	7.15	7.10	0.69
	100	10	39.8	38.6	100.0	−1.75	−1.76	−0.23	4.38	4.36	1.08
		20	39.2	38.3	100.0	−1.73	−1.74	−0.12	4.40	4.38	0.73
		40	39.0	37.9	100.0	−1.71	−1.71	−0.05	4.36	4.34	0.49
2	50	10	55.8	27.5	95.7	−1.78	−4.13	−0.91	5.78	9.79	3.69
		20	55.3	26.5	98.6	−1.77	−4.24	−0.80	6.27	10.09	3.14
		40	55.4	26.8	99.4	−1.77	−4.19	−0.80	5.75	9.98	3.31
	100	10	83.2	45.0	99.2	−0.91	−2.88	−0.99	4.58	7.19	3.63
		20	83.0	45.6	99.5	−0.81	−2.88	−0.81	3.11	7.13	2.94
		40	82.7	44.9	100.0	−0.79	−2.85	−0.62	3.20	7.21	2.42
3	50	10	24.7	19.6	99.9	−3.09	−3.15	−0.28	7.43	7.24	1.51
		20	24.4	19.4	100.0	−3.00	−3.08	−0.12	7.35	7.16	1.00
		40	24.9	19.5	100.0	−3.09	−3.16	−0.13	7.33	7.15	0.74
	100	10	40.0	37.8	100.0	−1.64	−1.68	−0.15	4.32	4.29	1.04
		20	41.1	37.9	100.0	−1.69	−1.71	−0.08	4.41	4.35	0.73
		40	40.9	38.1	100.0	−1.70	−1.73	−0.04	4.42	4.36	0.50
4	50	10	37.5	24.9	97.3	−2.79	−3.88	−0.65	7.12	8.86	2.57
		20	37.1	23.9	99.8	−2.64	−3.66	−0.45	6.90	8.42	2.09
		40	37.8	24.3	99.7	−2.75	−3.75	−0.27	7.11	8.60	1.42
	100	10	63.0	43.9	99.2	−1.38	−2.30	−0.38	3.92	5.57	1.80
		20	62.9	43.4	100.0	−1.41	−2.34	−0.30	4.06	5.63	1.38
		40	62.9	43.2	100.0	−1.38	−2.32	−0.18	4.07	5.61	1.01

Notes: See Tables 1 and 2 for an explanation of the various features of the table.

Table 4: Empirical rejection counts for the individual tests.

Variable	Test	Constant			Trend		
		1%	5%	10%	1%	5%	10%
Stock price	τ_{ML}	1	3	4	0	2	4
	τ_{LS}	0	2	2	0	0	0
Consumer price	τ_{ML}	0	0	0	0	0	0
	τ_{LS}	0	1	1	0	2	4
Exchange rate	τ_{ML}	1	1	3	0	0	0
	τ_{LS}	3	4	7	0	1	1

Notes: The table reports the number of rejections of the null hypothesis of a unit root at the 1%, 5% and 10% significance levels.

Table 5: Empirical results for the panel tests.

Variable	Value	Constant			Trend		
		t_a	t_b	t_δ^*	t_a	t_b	t_δ^*
Stock price	Test	-0.071	-0.260	-4.440	-0.915	-0.832	3.944
	p-value	0.472	0.397	0.000	0.180	0.203	1.000
Consumer price	Test	0.244	15.601	-10.597	-3.414	-4.163	-12.215
	p-value	0.596	1.000	0.000	0.000	0.000	0.000
Exchange rate	Test	-0.288	-1.866	-7.246	-1.580	-1.582	-4.384
	p-value	0.387	0.031	0.000	0.057	0.057	0.000

Notes: The tests of Moon and Perron (2004) are denoted t_a and t_b, while the Levin *et al.* (2002) test is denoted t_δ^*. All three tests take nonstationarity as the null hypothesis. The *p*-values are based on the normal distribution.