Testing in Nonstationary and Dependent Panels with Applications to Purchasing Power Parity

Inauguraldissertation zur Erlangung
des akademischen Grades
Doctor rerum politicarum
der Universität Dortmund

Christoph Hanck

Februar 2007

Contents

List of Figures

List of Tables

Preface

The following chapters present a synthesis of my research under the supervision of Walter Krämer and Kornelius Kraft during the past 18 months. It has been a pleasure to work under their fine guidance.

Many additional people have been a source of encouragement, providing helpful conversation, thoughtful advice and valuable feedback. I sincerely appreciated all of these, and have tried to follow most. Thanks go to my fellow doctoral students at the Ruhr Graduate School, to my colleagues and to Heide Aßhoff at the Institut für Wirtschafts- und Sozialstatistik, Universität Dortmund. Special thanks are given at the beginning of the respective chapters to which they contributed in some way or another. It has been a pleasure to co-author chapters with Walter Krämer and Guglielmo Maria Caporale.

I am grateful for the efforts of my teachers at the doctoral program of the Ruhr Graduate School which both shaped my research and helped me become a more versatile economist. It is a pleasure to acknowledge the generosity of the Ruhr Graduate School through a research grant and funding of several presentations at international conferences. Both were instrumental in producing this thesis.

I am indebted to my parents for many things, but for the present purposes, I want to say that I would not have written this thesis had they not reared me as an (I think) inquiring mind. Finally, I thank Steffi for being and making me cheerful throughout. This would not have been possible without her.

Chapter 1

Introduction

The present dissertation aims to make a contribution to improved econometric practice in testing statistical hypotheses on nonstationary and (cross-sectionally) dependent panel datasets. I have chosen to illustrate the challenges involved in testing in nonstationary and dependent panels by focusing on the Purchasing Power Parity (henceforth, PPP) Condition, or Law of One Price, one of the most intensively discussed theories in international economics. However, nonstationarity and cross-sectional dependence are common features of many macroeconomic panels. Hence, I conjecture that many of the issues discussed and methods suggested here may also be fruitfully applied to other macroeconometric questions, such as the Fisher relation or savings and investment correlations.

This Introduction briefly outlines the main results of and the common thread of the different chapters. These chapters are mainly taken from eight self-contained essays I have written (three of them together with Walter Krämer and Guglielmo Maria Caporale). I therefore have to beg the reader's pardon for some redundancies, which, however, were inevitable to make each of the papers as coherent as possible.

Chapter 2, written together with Guglielmo Maria Caporale, adopts a more traditional time series based framework to highlight drawbacks in testing for PPP by means of Engle and Granger (1987) or Johansen (1988) type cointegration tests on the exchange rate and home and foreign price levels. It has long been known that these tests suffer from problems such as low power (Haug, 1996), size distortion (Gonzalo, 1994) or excessive sensitivity to nuisance parameters (Kremers, Ericsson, and Dolado, 1992). Similarly, it is well-known that the outcome of tests for PPP are quite sensitive to, e.g., how one constructs the price variables (Coakley, Kellard, and Snaith, 2005). Chapter 2 adds to this literature by demonstrating that existing popular time series based cointegration tests seem to be uninformative about PPP (or "erratic") in that the conclusion of the test is highly sensitive to the sample period an investigator uses.

One of the remedies suggested in the literature to overcome these problems is to employ panel data—i.e., to pool observations on several units (here, countries) over time—to conduct estimation and inference (see, e.g., Baltagi, 2001). Several panel unit root and cointegration tests and estimators (see Breitung and Pesaran, 2007, for a recent survey) have been developed to this end. The contribution of Chapter 3 is to provide some new tests for panel cointegration. I extend the meta analytic panel unit root tests of Maddala and Wu (1999) and Choi (2001) to the cointegration setting. Compared with other tests in the literature, the meta analytic approach has the advantage of being highly flexible, relatively easy to implement and quite intuitive. I conduct a simulation study to demonstrate that several variants of the new tests compare quite favorably with existing ones in terms of both type I and type II error rates.

The tests put forward in Chapter 3 belong to the class of so called "First-Generation Tests" in that they ignore the potential presence of cross-sectional dependence among the countries in the panel. This is of course easily seen to be a very restrictive assumption (made to simplify the derivation of the asymptotic distribution of the tests). To give an example, it is tantamount to assuming that the exchange rate behaviour of the British Pound to the U.S. dollar is independent of the exchange rate behaviour of the Euro to the Dollar.

Recent work—so called "Second-Generation Tests" (e.g., Bai and Ng, 2004; Phillips and Sul, 2003)—has therefore focused on relaxing this assumption. Chapter 4 builds on Chapter 3 in that the most promising variants of the meta analytic panel tests are used to provide additional cross-sectional correlation robust tests for panel cointegration. I employ a sieve bootstrap approach to account for nonstationarity and dependence under the maintained null hypothesis of no panel cointegration. This semi-parametric bootstrap scheme allows for more general forms of cross-sectional dependence than other robust tests previously suggested in the literature. Again, a simulation study reveals that the tests are capable of handling dependence of a quite general form. An empirical application to the PPP condition shows that properly accounting for the loss of information incurred by combining correlated rather than independent data weakens the evidence in favor of PPP.

Chapter 5, written together with Guglielmo Maria Caporale, revisits the question of erratic behaviour of PPP test statistics. We use cross-sectional correlation robust panel unit root tests to test the PPP hypothesis over different sample periods. It turns out that the behaviour of the panel test statistics shows no evidence of erraticism. We therefore conjecture that panel tests are capable of not only alleviating the above-mentioned well-known problems of, e.g., low power but also the issue of erraticism discussed in Chapter 2 and in Caporale, Pittis, and Sakellis (2003).

The previous chapters contribute evidence that panel tests—if properly designed—hold much promise to become a useful standard tool in macroeconometric practice. The following three chapters, on the other hand, highlight some open issues in the use of panel tests.

Chapter 6 provides an analytic framework to understand the puzzling finding of many simulation studies that the performance of meta-analytic panel tests deteriorates if one increases the number of units in the panel. (We measure performance by control of the type I error rate. That is, a test performs well if its rejection rate—when applied to repeated drawings of finite samples generated under the null—is close to the nominal one.) This is, a priori, rather counterintuitive because one would normally expect the performance to improve if more information becomes available. I demonstrate that this puzzle can be explained by seemingly negligible size distortions of the time series tests used to construct the meta-analytic panel test statistics. These distortions then "add up" when combining ever more units to yield an increasingly size-distorted panel test. In practice, it is therefore strongly recommended to use time series tests which carefully control the type I error rate. I propose to solve or at least alleviate this problem by using adjusted critical values which give a better approximation to the finite sample distribution than the fixed (first-order) asymptotic critical values. In practice, adjusted critical values may be obtained via the response surface regression approach pioneered in the cointegration literature by MacKinnon (1991).

Chapter 7 compares results from test outcomes of several popular panel cointegration tests when applied to artificial data. The use of artificial data allows to control the Data Generating Process, such as to know whether a particular test decision is correct or not. It turns out that panel cointegration tests produce "mixed signals"—just as time series cointegration tests do (Gregory, Haug, and Lomuto, 2004). That is, it frequently occurs that one tests produces a rather emphatic rejection of the null hypothesis while another one confirms the null—even though both are applied to the same sample. Put differently, it is likely that it will be possible to find at least one test that, as desired, confirms or rejects any given hypothesis tested on some dataset. The likely explanation for this is that panel cointegration tests (unlike classical Wald, Likelihood Ratio or Lagrange Multiplier tests, for which Berndt and Savin (1977) show that the "mixed signals" problem vanishes asymptotically) have different implicit alternatives. E.g., Johansen (1988) type tests look at the eigenvalues of some residual moment matrix, whereas Engle and Granger (1987) type tests rely on the mean reversion of a residual series. While the alternatives of both tests is taken to be "cointegration," the mathematical characterization obviously differs substantially.

The recommendation for empirical practice therefore is to report results on preferably several panel cointegration tests to increase the confidence one can put into the rejection or non-rejection of some particular economic hypothesis, provided the test results agree.

Chapter 8, on a more general level, takes issue with popular approaches to testing for PPP and suggests a novel one. One popular, traditional strategy is to gather price and exchange rate data on several countries (relative to some reference country) and test for PPP in each by, e.g., testing the null of a unit root in the real exchange rate. It is then argued that PPP holds for those countries for which the null of a unit root is rejected at, say, the 5% level, because the real exchange rate is then statistically significantly mean-reverting. While this intuitive procedure is by construction adequate when one only considers a single country, it is questionable from a statistical point of view when applied to several countries simultaneously as it ignores the problem of multiplicity. That is, I demonstrate that one is almost bound to spuriously find some evidence in favor of PPP even if it is not present in any country because testing at, e.g., the 5% level in several countries does not control the type I error at the 5% level for the entire panel of countries.

Similarly, the panel approaches discussed in some of the previous chapters are to be used with caution, too, as many of the available tests' alternatives might lead one to mistakenly believe PPP to hold for all countries in case of a rejection even if in fact it only holds for some. Moreover, existing panel tests do not allow to identify the countries for which PPP holds.

I therefore modify a recent proposal of Romano and Wolf (2005) to be applicable to the nonstationary testing problem. The main idea is to control for multiplicity with a multiple testing technique which, unlike traditional Bonferroni type multiple tests, guarantees high power of the procedure by exploiting the dependence structure among the countries with a bootstrap procedure. Indeed, my empirical results show that the modified Romano and Wolf (2005) approach seems to be more powerful than traditional multiple testing techniques in that it identifies PPP to hold for more countries. Conversely, the results suggest that the simple testing approach ignoring multiplicity as well as panel tests apparently produce spurious rejections which do not result if one properly accounts for multiplicity.

Finally, Chapter 9 adopts an alternative approach to model cross-sectional dependence. This chapter contributes to the spatial econometrics literature, where one proceeds by assuming a known form of cross-sectional dependence. This assumption can be a plausible one if one can correctly capture the dependence structure among panels by observables as, e.g., trade shares. Consistent estimation is then possible even if only one panel wave is available.

The chapter first shows that the standard estimator of the disturbance variance, i.e., the degrees of freedom adjusted mean of the sum of squared Ordinary Least Squares (OLS) residuals, is applicable by demonstrating that the variance covariance matrix of the reduced form errors of a spatial model is homoskedastic under many popular specifications of the dependence structure. This clarifies a recurrent misunderstanding in the literature. Then, the chapter derives the exact finite sample bias of this estimator and proves its consistency. Hence, it is shown that a simple and readily computed estimator can be used as an ingredient in test statistics without affecting the asymptotic validity of the tests.

Chapter 2

Cointegration Tests Of PPP: Do They Exhibit Erratic Behaviour?

Abstract

We analyse whether tests of PPP exhibit erratic behaviour (as previously reported by Caporale, Pittis, and Sakellis, 2003) even when (possibly unwarranted) homogeneity and proportionality restrictions are not imposed, and trivariate cointegration (stage-three) tests between the nominal exchange rate, domestic and foreign price levels are carried out (instead of stationarity tests on the real exchange rate, as in stage-two tests). We examine the US dollar real exchange rate vis-à-vis 21 other currencies over a period of more than a century, and find that stage-three tests produce similar results to those for stage-two tests, namely the former also behave erratically. This confirms that neither of these traditional approaches to testing for PPP can solve the issue of PPP.[1]

Keywords: Purchasing Power Parity (PPP), Real Exchange Rate, Cointegration, Stationarity, Parameter Instability

[1]This chapter has been written jointly with Guglielmo Maria Caporale.

2.1 Introduction

Purchasing Power Parity (PPP) is one of the most popular theories for explaining the long-run behaviour of exchange rates, and has therefore been extensively investigated. Froot and Rogoff (1995) distinguish three stages in the time series literature on PPP. Stage-one tests were flawed by their failure to take into account possible non-stationarities in the series of interest. Stage-two tests focused on the null that the real exchange rate follows a random walk, the alternative being that PPP holds in the long run. However, such unit root tests were found to have very low power, and not to be able to distinguish between random-walk behaviour and very slow mean-reversion in the PPP-consistent level of the real exchange rate (see, e.g., Frankel, 1986, and Lothian and Taylor, 1997), unless very long spans of data were used (see, e.g., Lothian and Taylor, 1996, and Cheung and Lai, 1994).[2] Stage-three tests have used cointegration tests, but essentially suffer from the same problem of low power, and consequently have not significantly improved our understanding of real exchange rate behaviour (see Rogoff, 1996).

Caporale, Pittis, and Sakellis (2003) aimed to find an explanation for the contradictory evidence on PPP, even when long runs of data are used to increase the power of test statistics. They focused on stage-two tests and argued that the reason is that the type of stationarity exhibited by the real exchange rate cannot be accommodated by the fixed-parameter autoregressive homoscedastic models normally employed in the literature. Using a dataset including 39 countries and spanning a period of up to two centuries, they analysed the behaviour of both WPI- and CPI-based measures of the real exchange rate. In particular, they computed a recursive t-statistic (see below for a precise definition), and showed that it has an erratic behaviour, suggesting the presence of instability,

[2] Indeed, Krämer and Marmol (2004) show that the divergence of Dickey-Fuller type unit root tests is slower against slowly mean-reverting $I(d)$ alternatives.

and of a type of non-stationarity more complex than the unit root one usually assumed.

In the present study we explore this issue further by analysing whether erratic behaviour also characterises stage-three tests. The advantage of such tests is that they do not impose the homogeneity and proportionality restrictions entailed by stage-two tests, which might not hold in practice. Therefore, by carrying out cointegration tests of PPP we check whether there might be a relation between the presence of erratic behaviour and the imposition of overly strong restrictions. The layout of the chapter is as follows. Section 2 reviews the PPP condition in its different forms. Section 3 describes the data and presents some empirical evidence based on two different cointegration methods. Section 4 summarises the main findings and offers some concluding remarks.

2.2 The PPP Condition

In its absolute form, the PPP condition states that the nominal exchange rate should be proportional to the ratio of the domestic to the foreign price level, i.e.:

$$s_t = \alpha + \beta_0 p_t + \beta_1 p_t^*, \tag{2.1}$$

where s_t is the nominal exchange rate, p_t the domestic price level, and p_t^* the foreign price level, all in logs.[3] This is known as a trivariate relationship. Imposing the "symmetry" restriction $\beta_0 = -\beta_1 = \beta$ on the price coefficients, one obtains the following bivariate relationship:

$$s_t = \alpha + \beta(p_t - p_t^*). \tag{2.2}$$

Finally, the "proportionality" restriction $\alpha = 0,\ \beta = 1$ implies

$$q_t = s_t - p_t + p_t^*, \tag{2.3}$$

[3]Relative PPP implies that the percentage change in the exchange rate between two currencies equals the inflation differential, i.e. $\Delta s_t = \beta_0 \Delta p_t - \beta_1 \Delta p_t^*$.

where q_t is the real exchange rate.

Most of the literature in the 1980s tested PPP by means of (stage-two) unit root tests (DF or ADF—see Dickey and Fuller, 1979) on the real exchange rate, which, under PPP, should revert to its long-run equilibrium value given by PPP after being hit by shocks. The null hypothesis is that it follows a random walk (it has a unit root), since market efficiency implies that its changes should be unpredictable, whilst the alternative is that PPP holds. The maintained (joint) hypothesis is that the symmetry/proportionality restrictions both hold, which might not be true in practice. Consequently, the evidence presented by Caporale, Pittis, and Sakellis (2003) on the erratic behaviour of unit root tests might reflect unwarranted restrictions.

By contrast, a trivariate cointegration test of PPP entails running the following cointegrating regression (which does not impose any such restrictions):

$$s_t = \alpha + \beta_0 p_t - \beta_1 p_t^* + u_t \tag{2.4}$$

where the variables are defined as before, and u_t stands for the regression errors. PPP is then tested by means of DF and ADF tests on the estimated residuals. In the present chapter, by implementing cointegration tests of this type, we aim to establish whether or not evidence of erratic behaviour can still be found, even without the abovementioned restrictions, and consequently whether or not the findings of Caporale, Pittis, and Sakellis (2003) are robust or instead are due the imposition of unwarranted restrictions.

2.3 Cointegration Tests of PPP

Data sources and definitions

We revisit the dataset employed by Taylor (2002), which includes annual data for the nominal exchange rate, CPI and the GDP deflator. This dataset is

TABLE 2.1: ADF UNIT ROOT TESTS

country	p_{CPI}	p_{GDP}	e
Argentina	1.836	3.976	1.319
Australia	-0.671	-0.906	0.578
Belgium	-1.666	-2.510	-0.771
Brazil	0.681	5.162	1.204
Canada	-1.279	-2.079	0.875
Chile	0.111	0.061	0.229
Denmark	-1.941	-2.381	0.291
Finland	-1.158	-0.973	-0.763
France	-0.956	-0.849	-0.245
Germany	-2.190	-2.123	-1.525
Italy	-0.843	-0.528	-0.527
Japan	-0.282	-1.189	-1.401
Mexico	1.277	–[b])	1.751
Netherlands	-1.875	-1.484	0.175
New Zealand	-0.953	–[b])	-0.313
Norway	-1.931	-2.188	0.017
Portugal	-1.069	-1.089	-0.914
Spain	-0.314	-0.406	0.950
Sweden	-1.487	-2.226	0.185
Switzerland	0.096	-0.526	0.151
UK	-0.472	-0.564	0.793
United States	0.741	0.792	–[a])

N.A.: [a]) reference country
[b]) series unavailable/too short
The number of lagged differences is chosen according to the MAIC (Ng and Perron, 2001). Yearly data from 1892 to 1996. p_{CPI} is the log CPI price level, p_{GDP} is the log GDP deflated price level and e is the log nominal exchange rate.

particularly useful for our purposes because it covers a long period, ranging from 1892 through to 1996. The countries contained in our panel are given in Table 2.1. We use the United States as the reference country throughout. See Taylor (2002) for further details on data sources and definitions.

Empirical analysis

As a first step, we carried out standard augmented Dickey-Fuller (Said and Dickey, 1984) unit root tests to establish whether the series are all $I(1)$, and it is therefore legitimate to test for cointegration. The results indicate that this hypothesis can indeed not be rejected (see Table 2.1). We then proceeded to the estimation of cointegrating regressions using the Engle and Granger (1987) methodology. That is, we estimated (2.4) by OLS, and used the residuals to test the null hypothesis that they are nonstationary (i.e., that PPP does not hold) by means of DF and ADF tests. In order to investigate possible parameter instability, we created a new time series "t-stat" which is the computed t-statistic from the successive estimation of the coefficients of the following model whose order is selected using the Modified AIC (MAIC) of Ng and Perron (2001):

$$\Delta\hat{u}_t = \alpha_0 + \alpha_1\hat{u}_{t-1} + \sum_{j=1}^{p}\gamma_j\Delta\hat{u}_{t-j} + \varepsilon_t \tag{2.5}$$

Here, $\hat{u}_t$ are the residuals from OLS estimation of the cointegrating regression (2.4), ε_t is a white noise error term, and t-stat is defined as $\hat{\alpha}_1/est.s.e.(\hat{\alpha}_1)$. Equation (2.4) is estimated using the first k observations to produce the first residual series, from which we compute the unit root test statistic $\hat{\alpha}_1/est.s.e.(\hat{\alpha}_1)$. We then add an extra observation to compute the second estimate based on $k+1$ data points, and repeat the process until all T available observations have been used to yield $T-k+1$ values of the test statistics. We let $k \approx 20-25$ to discard estimates which are heavily affected by small-sample size-distortion. One can then plot the t-statistics based on the successive estimates to see more clearly whether it changes substantially as more data points are added, which would be a strong indication of instability in the parameter. Big jumps in either the rejection or the acceptance region, or from one to the other, are a strong sign of a structural break in the DGP. The results are summarised

TABLE 2.2: SUCCESSIVE T-STAT RESULTS

country	Min	Max	Accept	Reject	Obs
Argentina	-5.635	-1.259	0.728	0.272	80
Australia	-4.788	-0.536	0.975	0.025	80
Belgium	-3.734	0.220	1	0	80
Brazil	-2.947	-0.395	1	0	60
Canada	-4.253	-0.437	0.988	0.012	80
Chile	-4.427	-1.229	0.650	0.350	59
Denmark	-3.782	-1.509	1	0	80
Finland	-4.797	-0.471	0.481	0.519	80
France	-5.311	0.204	0.951	0.049	80
Germany	-3.867	-1.104	1	0	80
Italy	-3.666	-1.336	1	0	80
Japan	-6.253	-4.109	0	1	24
Mexico	-5.481	0.289	0.383	0.617	80
Netherlands	-4.092	0.774	0.988	0.012	80
New Zealand	-5.372	-2.259	0.560	0.440	24
Norway	-4.289	-0.496	0.988	0.012	80
Portugal	-5.923	-1.710	0.852	0.148	80
Spain	-3.242	-0.018	1	0	80
Sweden	-4.219	-1.773	0.852	0.148	80
Switzerland	-3.234	-0.279	1	0	80
UK	-6.642	-1.551	0.802	0.198	80

Minimum and maximum t-test statistics, acceptance and rejection percentages and number of available observations for each country, using CPI price series

in Table 2.2. Columns 4 and 5 show that the test decision on whether PPP holds or not is not constant over the sample in the vast majority of countries. Frequent switches from the rejection to the non-rejection regions are found to occur, the successive t-statistic exhibiting erratic behaviour very similarly to the case of stage-two tests. For some graphical illustration, consider the cases of Argentina (Figure 2.1), Finland (Figure 2.2), Mexico (Figure 2.3), or Chile (Figure 2.4).[4] The instability found clearly does not concern specific points in time, such that it could be dealt with using procedures for cointegration testing in the presence of structural breaks (see, e.g., Hansen, 1992, or Gre-

[4] The two lines at the bottom are the 10% and 5% critical values calculated as in MacKinnon (1991).

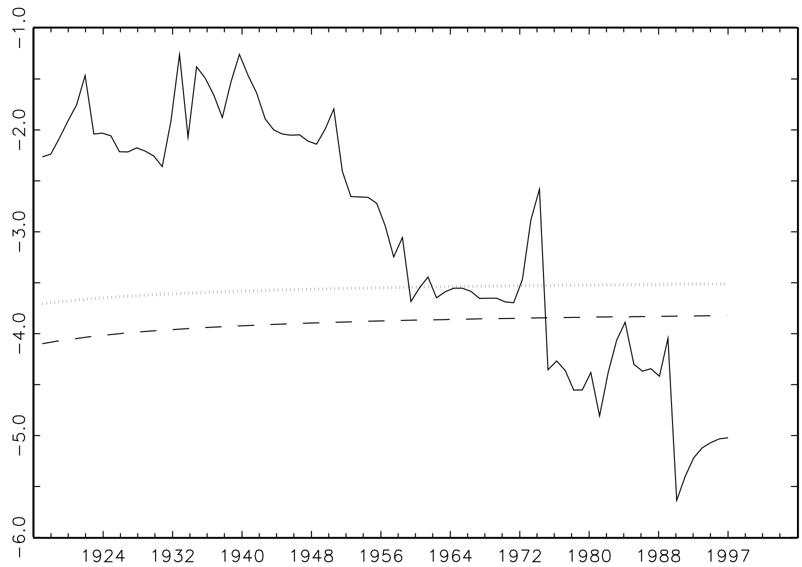


Figure 2.1: CPI-based Argentine t-stat series using a data-dependent rule (Ng and Perron, 2001) for the choice of lags in the ADF regression

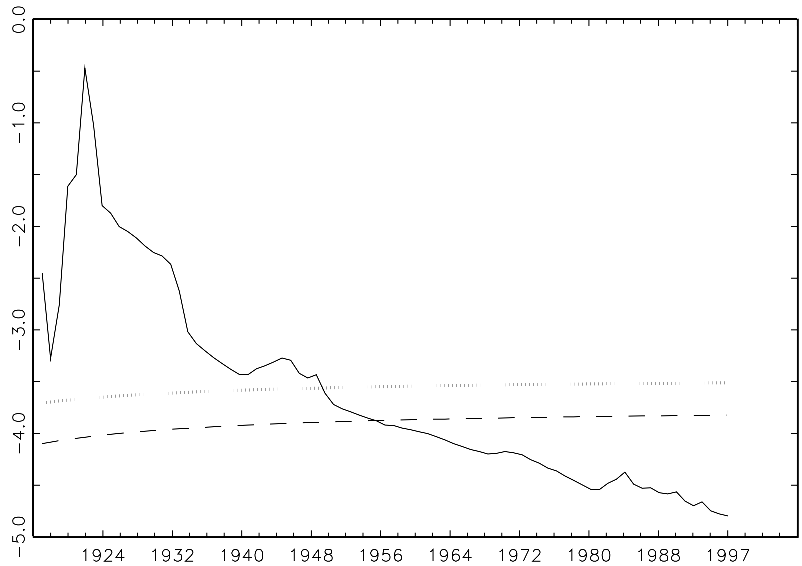


Figure 2.2: CPI-based Finnish t-stat series using a data-dependent rule (Ng and Perron, 2001) for the choice of lags in the ADF regression

gory and Hansen, 1996), but appears instead to be of an endemic type. As a counterexample where no switches occur at the finite sample 5% level, see Denmark (Figure 2.5).

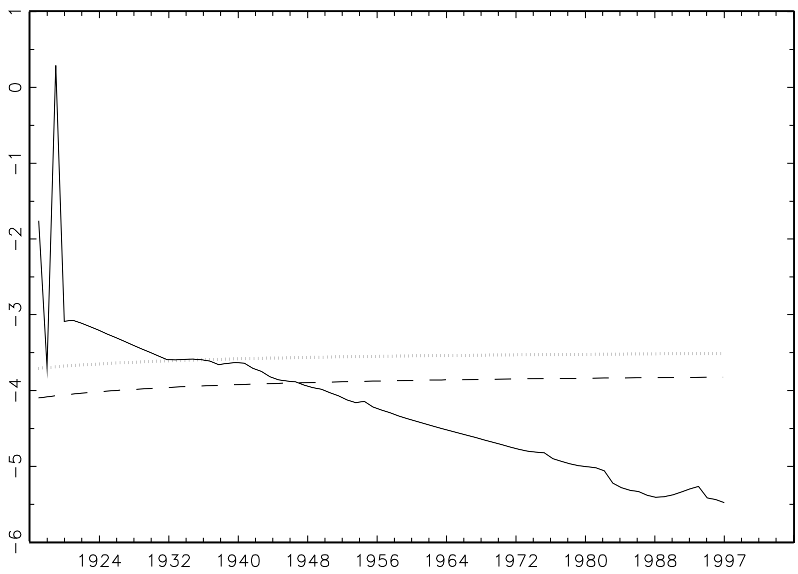


FIGURE 2.3: CPI-BASED MEXICAN T-STAT SERIES USING A DATA-DEPENDENT RULE (NG AND PERRON, 2001) FOR THE CHOICE OF LAGS IN THE ADF REGRESSION

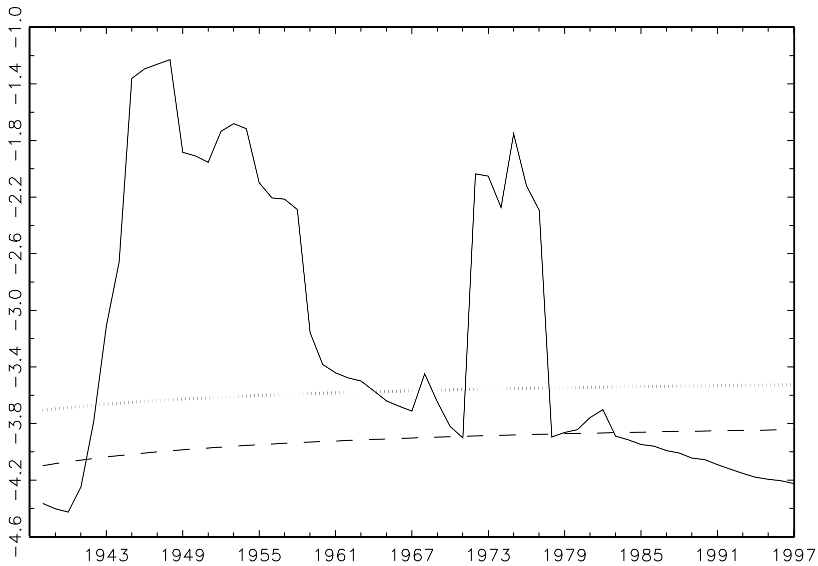


FIGURE 2.4: CPI-BASED CHILEAN T-STAT SERIES USING A DATA-DEPENDENT RULE (NG AND PERRON, 2001) FOR THE CHOICE OF LAGS IN THE ADF REGRESSION

We conducted the same type of analysis using the GDP deflator this time to construct the real exchange rate, obtaining a very similar picture, namely erratic behaviour in the majority of cases. For instance, compare Figure 2.7 with

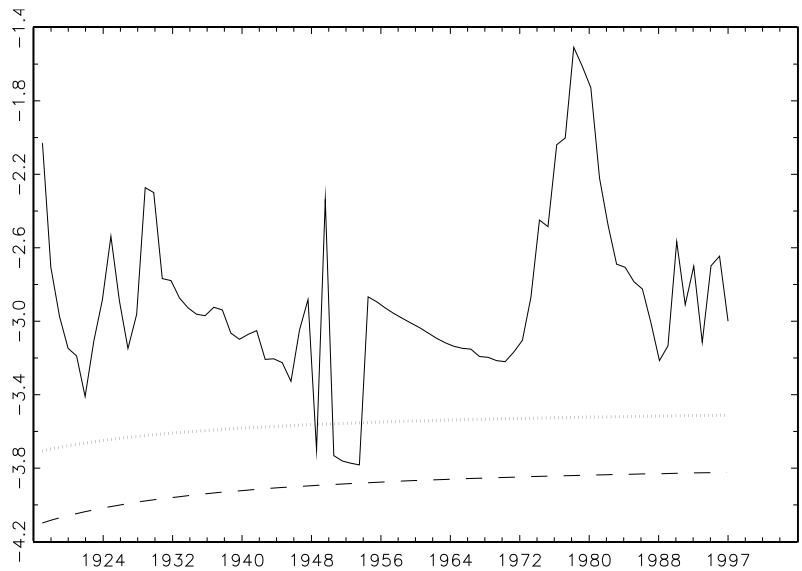


Figure 2.5: CPI-based Danish t-stat series using a data-dependent rule (Ng and Perron, 2001) for the choice of lags in the ADF regression

the corresponding CPI based Figure 2.1. There are only a few exceptions, such as Denmark, where no rejections occur (Figure 2.8).[5] Further, as a robustness check, we tried different number of lags in the ADF regressions (2.5). Overall, a qualitatively similar pattern emerges throughout, although we find that higher number of lags are associated with fewer rejections (see Figure 2.1 and Figures 2.9 to 2.12). This is what one would expect, the estimation of too many parameters resulting in lower power (Phillips and Perron, 1988).

To explore more in depth the issue of possible structural breaks, we also used fixed-size windows.[6] That is, we select a fixed sample size T^* and create the nth entry of the series t-stat as before but now based on observations $t = n, \ldots, T^* + n$, where $n = k, \ldots, T - T^*$. One would expect using fixed windows to reduce the likelihood of structural breaks occurring within the chosen sample, and hence to result in more frequent rejections of the null hypothesis that PPP

[5]Results for other countries are available upon request.

[6]A variety of other methods could also be used to shed additional light on whether structural breaks are present. See, e.g., Ploberger, Krämer, and Kontrus (1989).

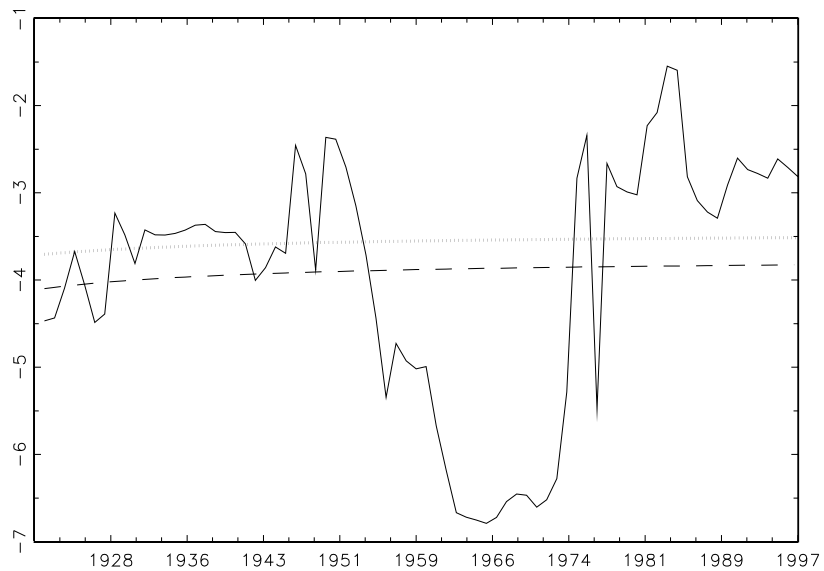


FIGURE 2.6: CPI-BASED DANISH T-STAT SERIES WITH A MOVING WINDOW OF $T^* = 30$

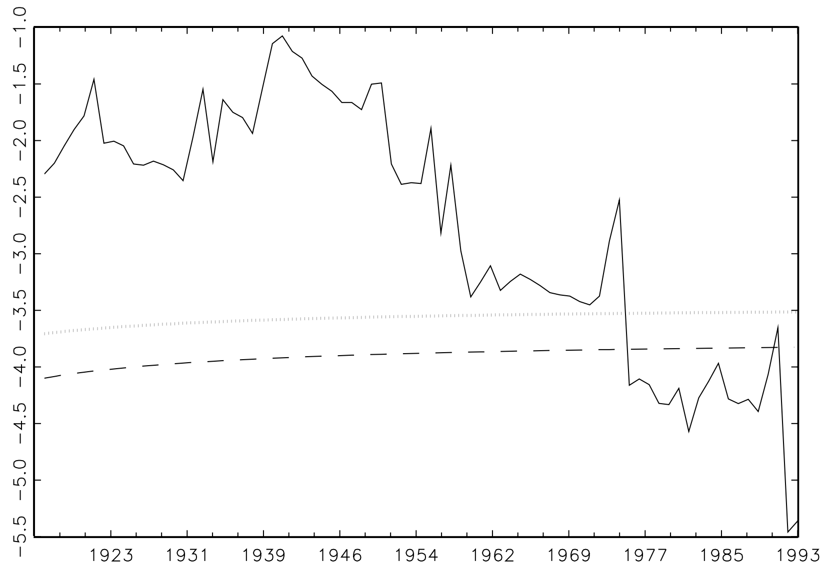


FIGURE 2.7: WPI-BASED ARGENTINE T-STAT SERIES USING A DATA-DEPENDENT RULE (NG AND PERRON, 2001) FOR THE CHOICE OF LAGS IN THE ADF REGRESSION

does not hold. However, it turns out that the behaviour of the t-stat series is, if anything, even more erratic than for increasing window sizes. It appears that the answer to whether or not PPP holds is highly dependent on the chosen sample. For instance, using Danish data ending in the 1960s and early 70s an

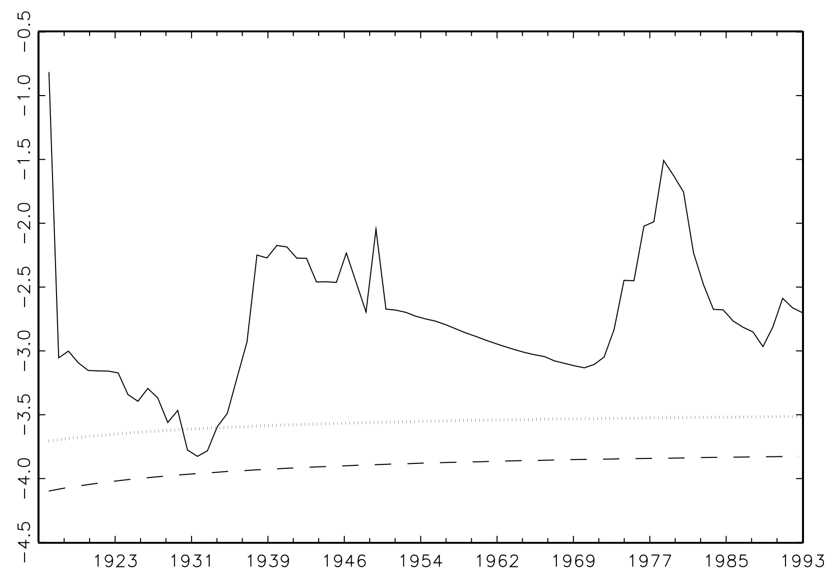


FIGURE 2.8: WPI-BASED DANISH T-STAT SERIES USING A DATA-DEPENDENT RULE (NG AND PERRON, 2001) FOR THE CHOICE OF LAGS IN THE ADF REGRESSION

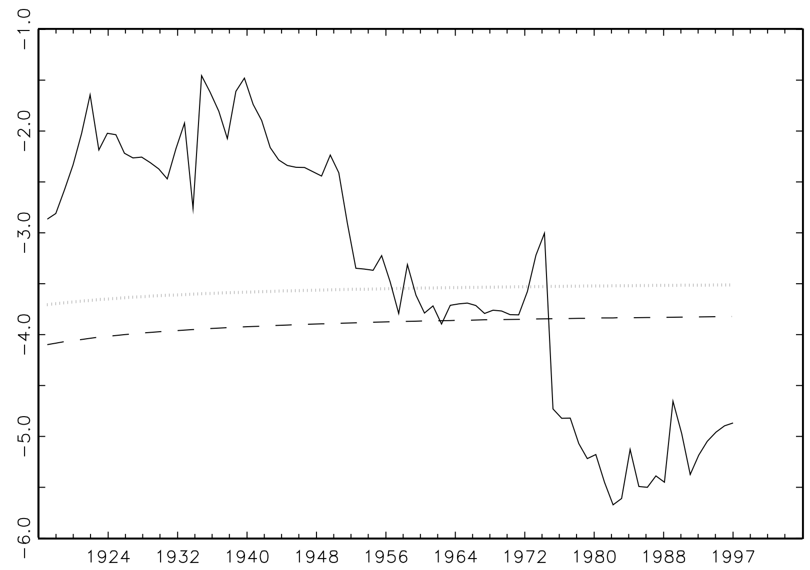


FIGURE 2.9: CPI-BASED ARGENTINE T-STAT SERIES USING 1 LAG IN ADF REGRESSION

investigator using $T^* = 30$ years of data would strongly reject the null of PPP not holding (see Figure 2.6).

Finally, we carried out alternative cointegration tests in all cases. Specifically,

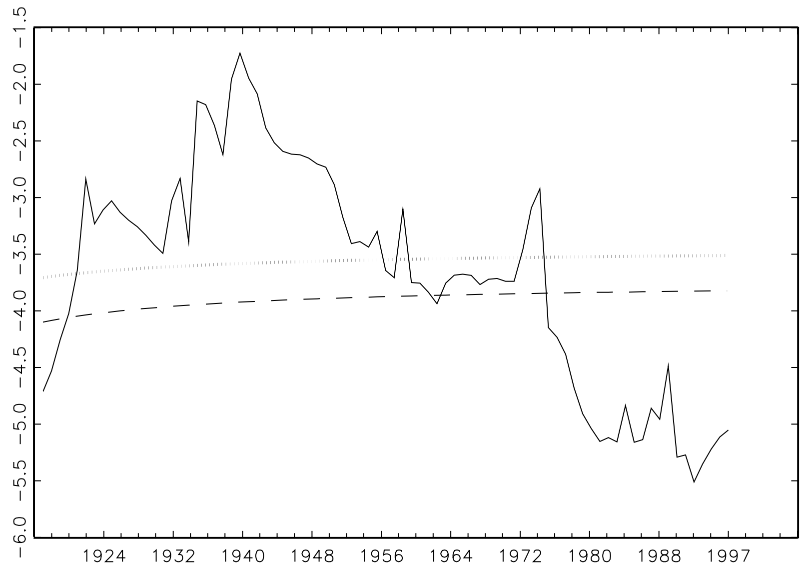


FIGURE 2.10: CPI-BASED ARGENTINE T-STAT SERIES USING 2 LAGS IN ADF REGRESSION

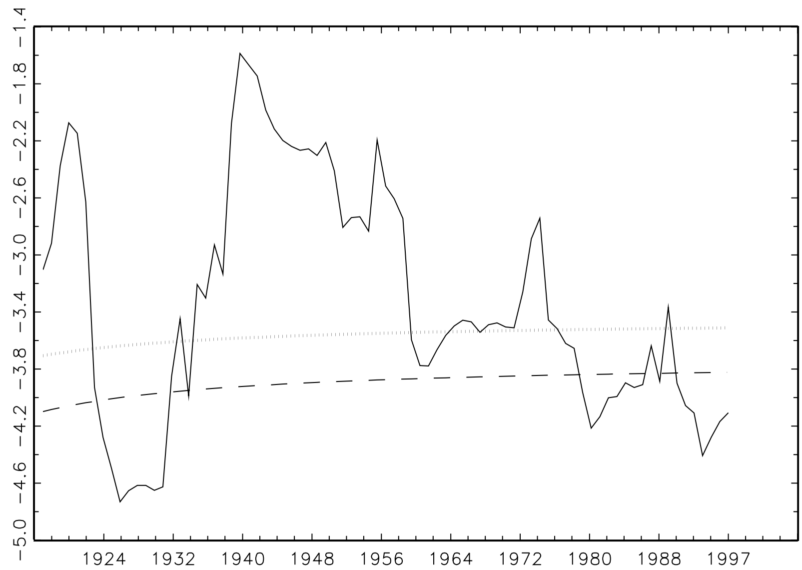


FIGURE 2.11: CPI-BASED ARGENTINE T-STAT SERIES USING 3 LAGS IN ADF REGRESSION

we used the λ-trace test (Johansen, 1988, 1991). Here the critical values were obtained by modifying the asymptotic ones from Osterwald-Lenum (1992) using the response surface regression results of Cheung and Lai (1993). Some

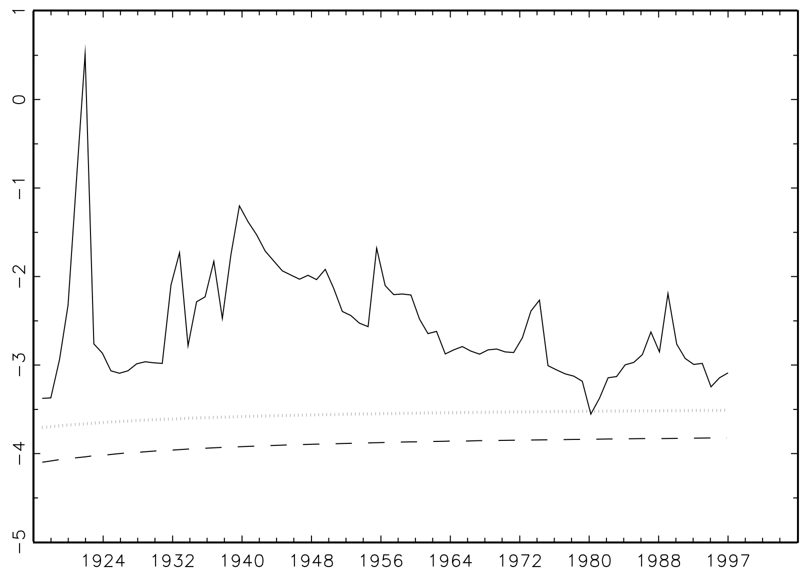


FIGURE 2.12: CPI-BASED ARGENTINE T-STAT SERIES USING 4 LAGS IN ADF REGRESSION

results are reported in Table 2.3.[7] Since this test statistic's null distribution is related to the χ^2 distribution, unlike in the previous cases, the rejection region is now above the critical value lines. As can be seen, we find further evidence of erratic behaviour (Figure 2.13), suggesting that this is not due to the type of cointegration test used, but it is a more fundamental issue pertaining to the stochastic properties of the PPP relationship. Interestingly, switches from the rejection to the non-rejection region occur around the same time in a number of cases—compare, e.g., Figures 2.1 and 2.13.[8]

2.4 Conclusion

In this chapter we have analysed whether tests of PPP exhibit erratic behaviour (as previously reported by Caporale, Pittis, and Sakellis, 2003) even

[7] Again, using WPI data or a different number of lagged differences in the Johansen procedure does not make a qualitative difference. Detailed results are available upon request.

[8] Similar patterns emerge for Australia, Brazil, Canada, Denmark, Finland, France, Mexico, the Netherlands, Norway, Portugal, Sweden and the UK, that is 14 out of 19 countries for which the sample size is sufficiently large to make statistically meaningful statements.

Table 2.3: Successive $\hat{\lambda}$-test results

country	Min	Max	Accept	Reject	Obs
Argentina	14.961	60.399	0.450	0.550	80
Australia	12.710	54.296	0.900	0.100	80
Belgium	28.806	81.093	0.025	0.975	80
Brazil	16.048	35.891	0.883	0.117	60
Canada	22.723	64.069	0.375	0.625	80
Chile	18.919	40.487	0.322	0.678	59
Denmark	28.834	75.214	0.025	0.975	80
Finland	49.730	75.629	0	1	80
France	16.919	66.961	0.688	0.313	80
Germany	19.392	77.090	0.025	0.975	80
Italy	27.789	85.518	0.300	0.700	80
Japan	34.481	70.718	0	1	24
Mexico	45.355	96.397	0	1	80
Netherlands	16.440	89.232	0.775	0.225	80
New Zealand	21.102	48.917	0.458	0.542	24
Norway	25.484	81.281	0.175	0.825	80
Portugal	12.222	95.157	0.325	0.675	80
Spain	16.991	34.764	0.925	0.075	80
Sweden	34.655	111.531	0	1	80
Switzerland	16.630	36.658	0.813	0.188	80
UK	28.242	78.313	0.063	0.938	80

Minimum and maximum $\hat{\lambda}$-test statistics, acceptance and rejection percentages and number of available observations for each country, using CPI price series

when (possibly unwarranted) homogeneity and proportionality restrictions are not imposed, and trivariate cointegration (stage-three) tests between the nominal exchange rate, domestic and foreign price levels are carried out (instead of stationarity tests on the real exchange rate, as in stage-two tests). We examine the US dollar real exchange rate vis-à-vis 21 other currencies over a period of more than a century, and find that stage-three tests produce similar results to those for stage-two tests, namely the former also behave erratically. This corroborates the findings of Caporale, Pittis, and Sakellis (2003), in the sense that these do not appear to be the consequence of arbitrarily imposed (symmetry/proportionality) restrictions.

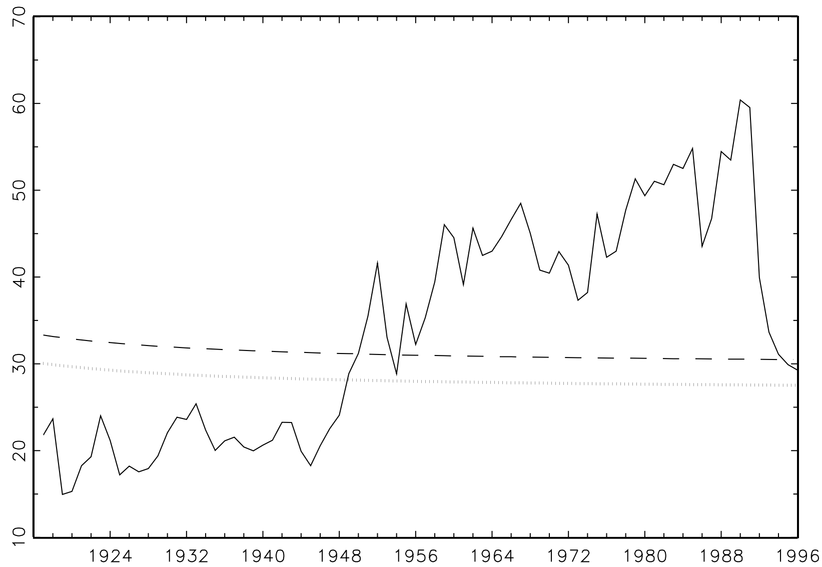


Figure 2.13: CPI-based Argentine $\hat{\lambda}$-trace stat series

Our results confirm that neither of the two traditional approaches to testing for PPP (stage-two and stage-three tests) can solve the issue of PPP. Consistently with Caporale, Pittis, and Sakellis (2003), the reported evidence again points to some form of non-stationarity in the data which is unlike the standard unit-root type normally assumed, or even the "separable" type discussed in Caporale and Pittis (2002), but rather one where all the unconditional moments are unknown functions of time. Future research should aim to determine its exact dynamic features.

Chapter 3

A Meta Analytic Approach to Testing for Panel Cointegration

Abstract

We propose new tests for panel cointegration by extending the panel unit root tests of Choi (2001) and Maddala and Wu (1999) to the panel cointegration case. The tests are flexible, intuitively appealing and relatively easy to compute. We investigate the finite sample behavior in a simulation study. Several variants of the tests compare favorably in terms of both size and power with other widely used panel cointegration tests.

Keywords: Panel cointegration tests, Monte Carlo study, Meta Analysis

3.1 Introduction

There is wide consensus in economics that cointegration is an important statistical concept which is implied by many economic models. In practice, however, evidence of cointegration or non-cointegration is often weak because of the rather small sample sizes typically available in macroeconometrics. To overcome this problem, the cointegration methodology has recently been extended to panel data. This allows the researcher to work with larger samples, thereby improving the power of tests and efficiency of estimators.

Pedroni (2004) and Kao (1999) generalize the residual-based tests of Engle and Granger (1987) and Phillips and Ouliaris (1990), Larsson, Lyhagen, and Löthgren (2001) extend the Johansen (1988) tests to panel data while McCoskey and Kao (1998) propose a test for the null of panel cointegration in the spirit of Shin (1994).

The present chapter introduces some new tests for panel cointegration, extending the p-value combination panel unit root tests of Maddala and Wu (1999) and Choi (2001) to the cointegration setting. In this framework, it is straightforward to account for unbalanced panels and arbitrary heterogeneity in the serial correlation structure of the series. Moreover, the tests are simple to implement and intuitively appealing. We explore the finite sample performance of the new tests in a simulation study. Certain variants of the tests compare favorably with many previously proposed panel cointegration tests.

3.2 P-Value Combination Tests for Panel Cointegration

The present section develops the new tests for panel cointegration. The following notation is used throughout. $\mathbf{x}_{ik}$ is a $(T_i \times 1)$ column vector collecting the observations on the kth variable of unit i of the panel, where $i = 1, \ldots, N$ and $k = 1, \ldots, K$. Additionally, we allow for time polynomials of order up to 2, i.e. constants, trend and squared trend terms. The number of observations T_i per unit may depend on i, i.e. the panel may be unbalanced. Denote by p_i the marginal significance level, or p-value, of a time series cointegration test applied to the ith unit of the panel. Let θ_{i,T_i} be a time series cointegration test statistic on unit i for a sample size of T_i. F_{T_i} denotes the exact, finite T_i null distribution function of θ_{i,T_i}. Since the tests considered here are one-sided, $p_i = F_{T_i}(\theta_{i,T_i})$ if the test rejects for small values of θ_{i,T_i} and $p_i = 1 - F_{T_i}(\theta_{i,T_i})$ if the test rejects for large values of θ_{i,T_i}. We only consider time series tests with the null of no cointegration.

We are interested in testing the following null hypothesis

$$H_0: \text{ There is no (within-unit) cointegration for any } i, \; i = 1, \ldots, N, \tag{3.1}$$

against the alternative

$$H_1: \text{ There is (within-unit) cointegration for at least one } i, \; i = 1, \ldots, N.$$

The alternative H_1 states that a rejection is evidence of 1 to N cointegrated units in the panel. That is, a rejection neither allows to conclude that the entire panel is cointegrated nor does it provide information about the number of units of the panel that exhibit cointegrating relationships.

The main idea of the suggested testing principle has been used in meta analytic studies for a long time (cf. Fisher, 1970; Hedges and Olkin, 1985). Consider the testing problem on the panel as consisting of N testing problems for each unit of the panel. That is, conduct N separate time series cointegration tests and obtain the corresponding p-values of the test statistics.[1] We make the

[1]Both Maddala and Wu (1999) and Choi (2001) suggest extending their panel unit root tests to the cointegration case. However, to the best of our knowledge, they do not provide an actual implementation nor do they investigate the performance of the tests. Furthermore, our approach is more general and likely to be more accurate in some respects to be discussed below.

following assumptions (see Pedroni, 2004).

Assumption 1 (*Continuity*)
Under H_0, θ_{i,T_i} has a continuous distribution function for all $i = 1, \ldots, N$.

Assumption 2 (*Cross-Sectional Uncorrelatedness*)
$x_{ik,t} = x_{ik,t-1} + \xi_{ik,t},\ t = 1, \ldots, T_i,\ i = 1, \ldots, N,\ k = 1, \ldots, K$. Let $\boldsymbol{\xi}_{i,t} \equiv (\xi_{i1,t}, \ldots, \xi_{iK,t})'$. We require $E[\boldsymbol{\xi}_{i,t}\boldsymbol{\xi}'_{j,s}] = \mathbf{0}\ \forall\ s, t = 1, \ldots, T_i$ and $i \neq j$. The error process $\boldsymbol{\xi}_{i,t}$ is generated as a linear vector process $\boldsymbol{\xi}_{i,t} = C_i(L)\boldsymbol{\eta}_{i,t}$, where L is the lag operator and C_i are coefficient matrices. $\boldsymbol{\eta}_{i,t}$ is vector white noise.

Remarks

- Assumption 1 is a regularity condition that asymptotically ensures a uniform p-value distribution of the time series test statistics under H_0 on the unit interval: $p_i \sim \mathcal{U}[0, 1]$ $(i \in \mathbb{N}_N)$ (see, e.g., Bickel and Doksum, 2001, Sec. 4.1). It is satisfied by the tests considered in this chapter.

- The second assumption is strong (see, e.g., Banerjee, Marcellino, and Osbat, 2005). It implies that the different units of a panel must not be linked to each other beyond relatively simple forms of correlation such as common time effects which can be eliminated by demeaning across the cross sectional dimension. This assumption is likely to be violated in many typical macroeconomic panel data sets. We will return to this issue in Chapter 4.

We now present the new tests. Combine the N p-values of the individual time series cointegration tests, $p_i,\ i = 1, \ldots, N$, as follows to obtain three test statistics for panel cointegration:

$$P_{\chi^2} = -2\sum_{i=1}^{N} \ln(p_i) \tag{3.2a}$$

$$P_{\Phi^{-1}} = N^{-\frac{1}{2}}\sum_{i=1}^{N} \Phi^{-1}(p_i) \tag{3.2b}$$

$$P_t = \sqrt{\frac{3(5N+4)}{\pi^2 N(5N+2)}}\sum_{i=1}^{N} \ln\left(\frac{p_i}{1-p_i}\right) \tag{3.2c}$$

Here, Φ is the cdf of the standard normal distribution. When considered together we refer to Eqs. (3.2a) to (3.2c) as P tests from now on. The P tests, via pooling p-values, provide convenient tests for panel cointegration by imposing minimal homogeneity restrictions on the panel. For instance, the different units of the panel can be unbalanced. Furthermore, the evidence for (non-)cointegration is first investigated for each unit of the panel and then compactly expressed with the p-value of the time series cointegration test. Hence, the coefficients describing the relationship between the different variables for each unit of the panel can be heterogeneous across i. Thus, the availability of large-T time series allows for pooling the data into a panel without having to impose strong homogeneity restrictions on the slope coefficients as in traditional panel data analysis.[2] Under Assumptions 1 and 2, as $T_i \to \infty$ for all i, the test statistics are asymptotically distributed as

$$P_{\chi^2} \to_d \chi^2_{2N}$$

$$P_{\Phi^{-1}} \to_d \mathcal{N}(0,1)$$

$$P_t \overset{\text{approx.}}{\to_d} \mathcal{T}_{5N+4},$$

where χ^2 is a chi-squared distributed random variable and $\mathcal{T}$ denotes Student's t distribution. The subscripts give the degrees of freedom. Using consistent time series cointegration tests, $p_i \to_p 0$ under the alternative of cointegration. Hence, quite intuitively, the smaller p_i, the more it acts towards rejecting

[2]For an overview of panel data models relying on $N \to \infty$ asymptotics see Hsiao (2003).

the null of no panel cointegration. The decision rule therefore is to reject the null of no panel cointegration when P_{χ^2} exceeds the critical value from a χ^2_{2N} distribution at the desired significance level. For (3.2b) and (3.2c) one would reject for large negative values of the panel test statistics $P_{\Phi^{-1}}$ and P_t, respectively.

We now discuss how to obtain the p-values required for computation of the P test statistics. (In Chapter 5, we will show that using accurate p-values for meta analytic panel tests is crucial to achieve a precise control of the type I error rate.) The null distributions of both residual and system-based time series cointegration tests converge to functionals of Brownian motion. Hence, analytic expressions of the distribution functions are hard to obtain, and p-values of the tests cannot simply be obtained by evaluating the corresponding cdf.

A remedy frequently adopted in the literature is to derive the critical values (and, consequently, the p-values) by Monte Carlo simulation. However, this approach is unsatisfactory for (at least) the following reason. These simulations are typically only performed for one sample size which is meant to provide an approximation to the asymptotic distribution. This sample size need neither be large enough to be useful as an asymptotic approximation nor does it generally yield accurate critical values for other sample sizes. MacKinnon, Haug, and Michelis (1999) show for cases where analytic expressions of the distribution functions are available that this approach may deliver fairly inaccurate critical values. In the time series case, it is now fairly standard practice to report p-values of unit root and cointegration tests using the results of the response surface regressions introduced by MacKinnon (1991). We follow this approach here.

The null hypothesis (3.1) formulates no precise econometric characterization of (non-) cointegration. This is to allow for generality in testing the long-run equilibrium properties of the series, enabling the researcher to use whichever time series tests seem suitable to test for time series (non-)cointegration in the different units of the panel. We use p-values of the Augmented Dickey-Fuller (ADF) cointegration tests (Engle and Granger, 1987) as provided by MacKinnon (1996).[3] That is, the p-values are obtained from the t-statistic of $\gamma_i - 1$ in the OLS regression

$$\Delta \hat{u}_{i,t} = (\gamma_i - 1)\hat{u}_{i,t-1} + \sum_{p=1}^{P} \nu_p \Delta \hat{u}_{i,t-p} + \epsilon_{i,t}.$$

Here, $\hat{u}_{i,t}$ is the usual residual from a first stage OLS regression of one of the $\mathbf{x}_{ik}$ on the remaining $\mathbf{x}_{i,-k}$ and, possibly, deterministic terms. Alternatively, one could capture serial correlation by the semiparametric approach of Phillips and Ouliaris (1990). Finally, we obtain the p-values for the Johansen (1988) λ_{trace} and λ_{max} tests provided in MacKinnon, Haug, and Michelis (1999). That is, we test for the presence of h cointegrating relationships by estimating the number of significantly non-zero eigenvalues of the matrix $\hat{\Pi}_i$ estimated from the Vector Error Correction Model

$$\Delta \mathbf{x}_{i,t} = -\Pi_i \mathbf{x}_{i,t-P} + \sum_{p=1}^{P-1} \Gamma_{i,p} \Delta \mathbf{x}_{i,t-p} + \boldsymbol{\epsilon}_{i,t}$$

by the λ_{trace}-test

$$\lambda_{trace,i}(h) = -T \sum_{k=h+1}^{K} \ln(1 - \hat{\pi}_{k,i}) \qquad (3.3)$$

and the λ_{max}-test

$$\lambda_{max,i}(h|h+1) = -T \ln(1 - \hat{\pi}_{h+1,i}). \qquad (3.4)$$

Here, $\hat{\pi}_{k,i}$ denotes the kth largest eigenvalue of $\hat{\Pi}_i$. In (3.3), the alternative is a general one, while one tests against $h+1$ cointegration relationships in (3.4).

[3] MacKinnon improves upon his prior work by using a heteroskedasticity and serial correlation robust technique to approximate between the estimated quantiles of the response surface regressions. Our application is based on a translation of James MacKinnon's Fortran code into a GAUSS procedure which is available upon request. The procedure implements all panel data tests developed in this section.

Hence, we obtain the p-values required for performing the P tests from the most widely used time series cointegration tests.

3.3 Finite Sample Performance

We now present a Monte Carlo study of the finite sample performance of the tests proposed in the previous section. The Data Generating Process (DGP) is similar to the one used by Engle and Granger (1987). The extension to the panel data setting is discussed in Kao (1999). For simplicity, only consider the bivariate case, i.e. $K = 2$:

DGP A

$$x_{i,1t} - \alpha_i - \beta x_{i,2t} = z_{i,t}, \qquad a_1 x_{i,1t} - a_2 x_{i,2t} = w_{i,t}$$

where

$$z_{i,t} = \rho z_{i,t-1} + e_{zi,t}, \qquad \Delta w_{i,t} = e_{wi,t}$$

and

$$\begin{pmatrix} e_{zi,t} \\ e_{wi,t} \end{pmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \psi\sigma \\ \psi\sigma & \sigma^2 \end{bmatrix} \right)$$

Remarks

- When $|\rho| < 1$ the equilibrium error in the first equation is stationary such that $x_{i1,t}$ and $x_{i2,t}$ are cointegrated with $\boldsymbol{\beta}_i = (1 \quad -\alpha_i \quad -\beta)'$.

- When writing the above DGP as an error correction model (see, e.g., Gonzalo, 1994) it is immediate that $x_{i2,t}$ is weakly exogenous when $a_1 = 0$.

We investigate all combinations of the following values for the parameters of the model: $\beta = 2,\ a_1 \in \{0, 1\},\ a_2 = -1,\ \sigma \in \{0.5, 1\},\ \rho \in \{0.9, 0.99, 1\}$ and $\psi \in \{-0.5, 0, 0.5\}$. The fraction of cointegrated series in the panel is increased from 0 to 1 in steps of 0.1, i.e. $\delta \in \{0, 0.1, \ldots, 1\}$. The dimensions of the panel are $N \in \{10, 20, 50, 100, 150\}$ and, after having discarded 150 initial observations,

$T \in \{10, 30, 50, 100, 250, 500\}$, for a total of $2 \times 1 \times 2 \times 3 \times 3 \times 11 \times 5 \times 6 = 11,880$ experiments. For a given cross-sectional dimension, the unit specific intercepts are drawn as $\alpha_i \sim \mathcal{U}[0, 10]$ and kept fixed for all T. Each experiment involves $M = 5,000$ replications.[4] We choose a common β for all i in order to be able to compare the performance of our tests with results for other panel cointegration tests as reported by Gutierrez (2003). The p-values are from the Engle and Granger (1987) ADF test, holding the number of lagged differences fixed at 1. We further test for cointegration using the λ_{trace}-test for $h = 0$ vs. an unrestricted number of cointegrating relationships.

For brevity, we only give the results for $\psi = 0$, $a_1 = 0$ and $\sigma = 1$.[5] Table 3.1 shows the empirical size of the tests ($\rho = 1$) at the nominal 5% level using the ADF- and λ_{trace}-tests as the underlying time series tests. Two conclusions are obvious. First, the Engle/Granger-based tests are undersized. This issue is particularly severe in short panels but vanishes with increasing T. Oddly, all tests become *more* undersized as N increases. Chapter 5 provides an analysis of this behavior. The P_{χ^2} test seems to have slightly better size than the other two. We also investigate whether using MacKinnon's (1996) p-values improves the behavior of the tests relative to obtaining quantiles by generating only one set of replicates. For smaller panels, the latter approach (with 50,000 replications) exhibits non-negligible upward size distortions even when using quantiles specifically generated for the sample sizes considered. Interestingly, however, there does not seem to be a trend towards lower size with increasing N. For medium- and large-dimensional panels neither approach has a clear advantage over the other.

[4]Uniform random numbers are generated using the KM algorithm from which Normal variates are created with the fast acceptance-rejection algorithm, both implemented in GAUSS. Part of the calculations are performed with COINT 2.0 by Peter Phillips and Sam Ouliaris.

[5]The full set of results of the finite sample study are available upon request. Broadly speaking, a lower σ seems to have little, if any, systematic effect. Correlation in the error processes ($\psi \neq 0$) has a slightly negative effect on power.

Table 3.1: Empirical Size of the P Tests

T	N	*ADF* 10	20	50	100	150	λ_{trace} 10	20	50	100	150
(i) P_{χ^2}											
10		.038	.040	.024	.018	.011	.956	.999	1.00	1.00	1.00
30		.035	.031	.021	.014	.009	.184	.260	.467	.702	.845
50		.041	.033	.027	.022	.021	.102	.137	.216	.344	.437
100		.047	.042	.036	.034	.029	.074	.077	.109	.143	.171
250		.046	.044	.046	.045	.036	.052	.056	.063	.076	.086
500		.049	.048	.049	.048	.047	.056	.052	.054	.068	.068
(ii) $P_{\Phi^{-1}}$											
10		.027	.019	.006	.002	.001	.954	.998	1.00	1.00	1.00
30		.034	.022	.016	.009	.005	.180	.265	.468	.711	.848
50		.038	.030	.026	.018	.016	.102	.136	.218	.355	.447
100		.046	.038	.032	.033	.025	.072	.081	.111	.139	.177
250		.043	.045	.044	.041	.032	.051	.061	.061	.079	.086
500		.049	.047	.045	.044	.041	.056	.053	.055	.070	.065
(iii) P_t											
10		.030	.019	.006	.002	.001	.957	.998	1.00	1.00	1.00
30		.035	.023	.016	.010	.005	.183	.264	.473	.716	.852
50		.039	.030	.027	.018	.014	.102	.139	.217	.36	.447
100		.046	.038	.033	.032	.024	.074	.079	.109	.141	.176
250		.046	.045	.044	.041	.031	.053	.061	.063	.082	.086
500		.050	.048	.046	.045	.041	.056	.057	.059	.070	.067

Note: $\rho = 1,\ \psi = 0,\ \sigma = 1$ and $a_1 = 0$. $M = 5,000$ replications.
5% nominal level. *ADF* and λ_{trace} are the underlying time series tests.

Second, the Johansen-based tests are grossly oversized in panels of small and medium dimensions. Two reasons may be put forward for this disappointing performance. First, the underlying λ_{trace}-test overrejects in short time series when using asymptotic critical values (see also Cheung and Lai, 1993). This flaw then inevitably carries over to the panel tests via erroneously small p-values. Second, MacKinnon, Haug, and Michelis (1999) emphasize that the p-values estimated for the Johansen (1988) tests, unlike those estimated in MacKinnon (1996) for the Engle and Granger (1987) test, are only valid asymptotically. It may thus not be appropriate to use these for shorter time series. We therefore waive to report the essentially meaningless empirical power for shorter panels.

Table 3.2 shows the raw[6] power of the tests at $\rho = 0.9$. The major findings are as follows. First, after having discarded the severely size-distorted panels, both the Engle/Granger- and Johansen-based tests behave consistently in that power for all variants grows with both dimensions. The use of panel data is therefore justified. Second, the $P_{\Phi^{-1}}$ and the P_t tests outperform the P_{χ^2} test at least for the ADF variant. This finding is in line with the results reported by Choi (2001) for his panel unit root tests. Whether to choose the $P_{\Phi^{-1}}$ or the P_t in any application would be a matter of taste. Third, in each of the cases, power grows faster along the time series dimension. More specifically, the power of the tests rises quickly between $T = 50$ and $T = 100$. The simulation evidence therefore suggests that the P tests are particularly useful in relatively long panels. Figure 3.1 plots the power of the Engle/Granger-based tests for $N = 100$ as the fraction of cointegrated variables in the system, δ, increases. Panels (a) and (b) depict the cases $T = 50$ and $T = 100$, respectively. It can be seen that the power of the P tests rises to one substantially quicker when the underlying time series are longer.

We now relate our results to those of Gutierrez (2003). We first give the key statistics of the various tests that are considered. For more details we refer to the original contributions. Furthermore, Banerjee (1999), Baltagi and Kao (2000) or Breitung and Pesaran [forthcoming] provide surveys of the literature.

Pedroni (2004)

Pedroni (2004) derives seven different tests for panel cointegration. These may be categorized according to what information on the different units of the panel is pooled. The "Group-Mean" Statistics are essentially means of the conven-

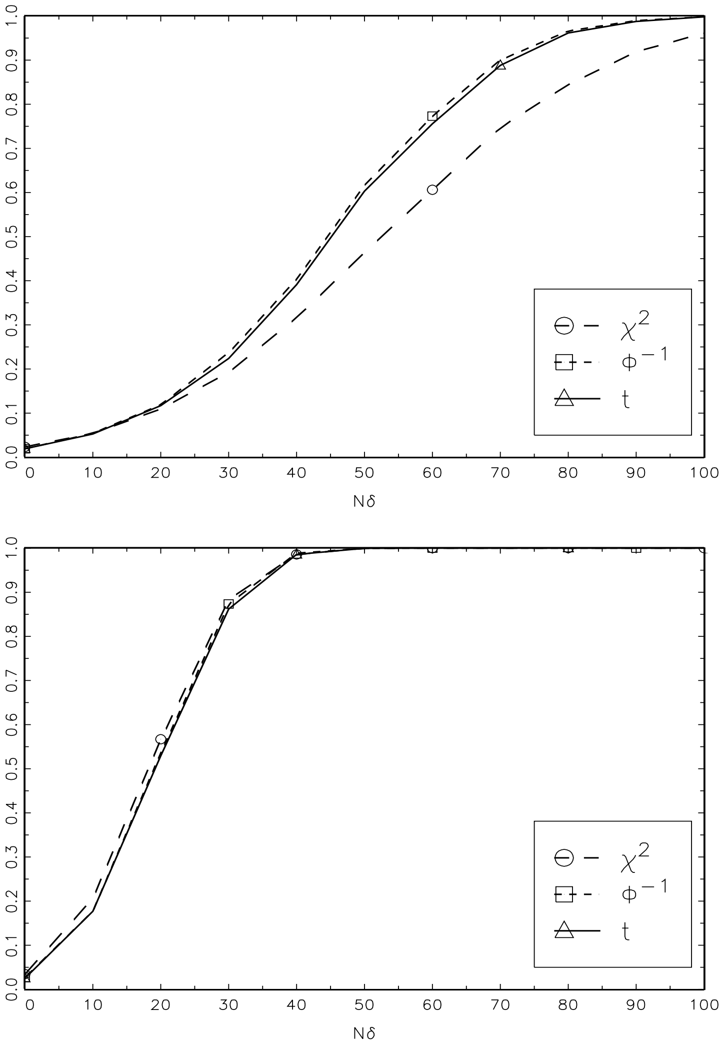
[6] Horowitz and Savin (2000) emphasize that size-adjusted critical values are of little use in empirical work. We therefore do not calculate size-adjusted power.

Table 3.2: Empirical Power of the P Tests

		ADF					λ_{trace}				
T	N	10	20	50	100	150	10	20	50	100	150
(*i*) P_{χ^2}											
10		.061	.052	.040	.037	.033					
30		.062	.070	.088	.120	.135					
50		.115	.158	.287	.465	.609	.082	.131	.138	.152	.176
100		.403	.659	.955	1.00	1.00	.110	.217	.246	.254	.309
250		.999	1.00	1.00	1.00	1.00	.564	.966	.997	.999	1.00
500		1.00	1.00	1.00	1.00	1.00	.998	1.00	1.00	1.00	1.00
(*ii*) $P_{\Phi^{-1}}$											
10		.039	.031	.015	.006	.005					
30		.063	.078	.108	.151	.188					
50		.136	.201	.376	.617	.771	.076	.130	.128	.129	.143
100		.426	.700	.968	1.00	1.00	.106	.223	.238	.238	.271
250		.990	1.00	1.00	1.00	1.00	.443	.892	.975	.996	1.00
500		1.00	1.00	1.00	1.00	1.00	.941	1.00	1.00	1.00	1.00
(*iii*) P_t											
10		.041	.030	.015	.005	.004					
30		.063	.076	.105	.150	.182					
50		.132	.196	.359	.603	.751	.078	.135	.131	.134	.148
100		.423	.685	.961	1.00	1.00	.108	.222	.240	.241	.277
250		.994	1.00	1.00	1.00	1.00	.480	.915	.982	.997	1.00
500		1.00	1.00	1.00	1.00	1.00	.979	1.00	1.00	1.00	1.00

Note: $\rho = 0.9,\ \psi = 0,\ \sigma = 1,\ \delta = 0.5$ and $a_1 = 0$. $M = 5,000$ replications. 5% nominal level. ADF and λ_{trace} are the underlying time series tests.

tional time series tests (see Phillips and Ouliaris, 1990). The "Within" Statistics separately sum the numerator and denominator terms of the corresponding time series statistics. Let $A_i = \sum_{t=1}^{T} \tilde{e}_{i,t}\tilde{e}'_{i,t}$, where $\tilde{e}_{i,t} = (\Delta\hat{e}_{i,t}, \hat{e}_{i,t-1})'$. The $\hat{e}_{i,t}$ are obtained from heterogenous Engle/Granger-type first stage OLS regressions of an $\mathbf{x}_{ik}$ on the remaining $\mathbf{x}_{i,-k}$, and possibly some deterministic regressors. We consider the "Group-ρ", "Panel-ρ" and (nonparametric)



(A) $T = 50,\ N = 100.$ (B) $T = N = 100.$
$\rho = 0.9,\ \psi = 0,\ \sigma = 1$ AND $a_1 = 0$

FIGURE 3.1: POWER OF THE P PANEL COINTEGRATION TESTS

"Panel-t"-test statistics which are given by, respectively,

$$\begin{aligned}
\tilde{Z}_{\hat{\rho}_{NT}-1} &= \sum_{i=1}^{N} A_{22i}^{-1}(A_{21i} - T\hat{\lambda}_i), \\
Z_{\hat{\rho}_{NT}-1} &= \left(\sum_{i=1}^{N} A_{22i}\right)^{-1} \sum_{i=1}^{N}(A_{21i} - T\hat{\lambda}_i) \quad \text{and} \\
Z_{\hat{t}_{NT}} &= \left(\tilde{\sigma}_{NT}^2 \sum_{i=1}^{N} A_{22i}\right)^{-1/2} \sum_{i=1}^{N}(A_{21i} - T\hat{\lambda}_i).
\end{aligned}$$

The expressions $\hat{\lambda}_i$ and $\tilde{\sigma}^2_{NT}$ estimate nuisance parameters from the long-run conditional variances. After proper standardization, all statistics have a standard normal limiting distribution. The decision rule is to reject the null hypothesis of no panel cointegration for large negative values.

Kao (1999)

Kao (1999) proposes five different panel extensions of the time series $(A)DF$-type tests. We focus on those that do not require strict exogeneity of the regressors. More specifically,

$$DF^*_\rho = \frac{\sqrt{N}T(\hat{\rho}-1) + \dfrac{3\sqrt{N}\hat{\sigma}^2_\nu}{\hat{\sigma}^2_{0\nu}}}{\sqrt{3 + \dfrac{36\hat{\sigma}^4_\nu}{5\hat{\sigma}^4_{0\nu}}}} \quad \text{and}$$

$$DF^*_t = \frac{t_\rho + \dfrac{\sqrt{6N}\hat{\sigma}^2_\nu}{2\hat{\sigma}_{0\nu}}}{\sqrt{\dfrac{\hat{\sigma}^2_{0\nu}}{2\hat{\sigma}^2_\nu} + \dfrac{3\hat{\sigma}^2_\nu}{10\hat{\sigma}^2_\nu}}}.$$

Here, $\hat{\rho}$ is the estimate of the AR(1) coefficient of the residuals from a fixed effects panel regression and t_ρ is the associated t-statistic. The remaining terms play a role similar to the nuisance parameter estimates in the Pedroni (2004) tests. Again, both tests are standard normal under the null of no panel cointegration and reject for large negative values.

Larsson, Lyhagen, and Löthgren (2001)

The panel cointegration test of Larsson, Lyhagen, and Löthgren (2001) applies a Central Limit Theorem to (3.3). Defining $\overline{\lambda}_{trace} = N^{-1}\sum_{i=1}^{N}\lambda_{trace,i}$, their panel cointegration test statistic is given by

$$\Upsilon_{LR} = \sqrt{N}\left(\frac{\overline{\lambda}_{trace} - \mathsf{E}[\overline{\lambda}_{trace}]}{\sqrt{\mathsf{Var}[\overline{\lambda}_{trace}]}}\right).$$

Under some conditions, including $\sqrt{N}T^{-1} \to 0$, Larsson, Lyhagen, and Löthgren (2001) can show that $\Upsilon_{LR} \xrightarrow{T,N} \mathcal{N}(0,1)$. The moments are obtained by stochastic simulation and are tabulated in the chapter. The null hypothesis of no cointegration at a level α is rejected if the test statistic exceeds the $(1-\alpha)$-quantile of the standard normal distribution, i.e. for large values.

Now, let us compare the results in Figure 3.1 with those obtained by Gutierrez (2003).[7] The P tests are somewhat less powerful than the residual-based panel tests $\tilde{Z}_{\hat{\rho}_{NT}-1}$, $Z_{\hat{\rho}_{NT}-1}$, DF^*_ρ and DF^*_t for shorter panels. However, power for longer panels is similar. Furthermore, the P tests always outperform the system-based Υ_{LR} test by Larsson, Lyhagen, and Löthgren (2001). Note, though, that these results are not based on size-adjusted critical values as in Gutierrez (2003). Given that the P tests seem to be undersized (see Table 3.1), their power would be higher if it were reported on the basis of exact rather than nominal critical values.

We think that DGP A is restrictive. Apart from the unit specific intercepts, no heterogeneity is allowed for. But, in many practical applications, the units of a panel, say, countries, differ in their short-run dynamic adjustment behavior. We therefore elicit how the performance of the tests changes when we introduce heterogeneity in the serial correlation properties. Since, to the best of our knowledge, no comparison of the different panel cointegration tests under these circumstances is available in the literature, we also include some of the tests presented above.

Consider the following modification of DGP A to introduce higher order serial correlation in the equilibrium error $z_{i,t}$. We draw, for each cointegrated series in the panel, the order of the AR-process according to $\tilde{\zeta}_i = [\zeta_i]$, where $\zeta_i \sim \mathcal{U}[1,6]$, $i = 1,\ldots,\delta N$ and $[y]$ is the integer part of y. We then generate the AR-coefficients from $\varphi_{i,p} \sim \mathcal{U}[0,0.99]$, $i = 1,\ldots,\delta N$; $p_i = 1,\ldots,\tilde{\zeta}_i$, discarding

[7] Figure 3.1 corresponds to the middle and lower right panel in Fig. 1 in Gutierrez (2003).

TABLE 3.3: POWER WITH AR(p) ERRORS

T	N	10	30	50	10	30	50
		$P_{\chi^2 DF}$			$P_{\Phi^{-1}DF}$		
20		.899	.999	1.00	.792	.992	.999
30		.990	1.00	1.00	.961	1.00	1.00
50		.999	1.00	1.00	.997	1.00	1.00
		P_{tDF}			$Z_{\hat{\rho}_{NT}-1}$		
20		.832	.995	.999	.806	.996	1.00
30		.973	1.00	1.00	.899	.999	1.00
50		.998	1.00	1.00	.980	1.00	1.00
		$Z_{\hat{t}_{NT}}$			DF^*_ρ		
20		.968	1.00	1.00	.030	.019	.015
30		.969	1.00	1.00	.101	.109	.130
50		.988	1.00	1.00	.276	.355	.424

Note: $\sigma = 1,\ \delta = 0.5,\ \psi = a_1 = 0$.
$M = 5,000$ replications. 5% nominal level.

all processes with eigenvalues outside the unit circle.

DGP B

$$
\begin{aligned}
x_{i1,t} - \alpha_i - \beta x_{i2,t} &= z_{i,t}, & a_1 x_{i1,t} - a_2 x_{i2,t} &= w_{i,t},\\
z_{i,t} &= \sum_{p_i=1}^{\tilde{\zeta}_i} \varphi_{i,p_i} z_{i,t-p_i} + e_{zi,t}, & \Delta w_{i,t} &= e_{wi,t},\\
\begin{pmatrix} e_{zi,t} \\ e_{wi,t} \end{pmatrix} &\overset{iid}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \psi\sigma \\ \psi\sigma & \sigma^2 \end{bmatrix}\right)
\end{aligned}
$$

Table 3.3 gives results on the power of the tests for $\sigma = 1,\ \psi = 0,\ a_1 = 0$ and $\delta = 0.5$. The dimensions are $T \in \{20, 30, 50\},\ N \in \{10, 30, 50\}$. The second part of the subscript ('DF') indicates that Engle and Granger's (1987) ADF test is chosen as the underlying time series test for the P tests. The number of lagged differences for the ADF regression is chosen according to the automatic procedure suggested by Ng and Perron (2001). It is not possible to compare the power with the results from Table 3.2 because the alternative is now different. But, Table 3.3 shows that the first five tests clearly have higher power than the last one. This is intuitive as the P and Pedroni (2004) tests are designed to accommodate cross-sectional heterogeneity. The tests put forward in this chapter may therefore be useful in a fairly wide range of practical applications.

3.4 Conclusion

We introduce new tests for panel cointegration. As in Maddala and Wu (1999) and Choi (2001), we use a meta analytic p-value combination approach to develop tests for nonstationary panel data. The new tests are flexible, intuitively appealing and easy to implement. The tests employ highly accurate p-values obtained from response surface regressions (MacKinnon, 1996; MacKinnon, Haug, and Michelis, 1999). A finite sample study reveals that the Engle and Granger (1987)-based variant of the suggested tests is somewhat undersized in very short and wide panels. However, the empirical size of the tests is very close to the nominal one for panel dimensions often encountered in applied macroeconometric work. In terms of power, their performance is intermediate between other widely used panel cointegration tests.

As most tests in this literature, the ones suggested here rely on the assumption of cross-sectional uncorrelatedness (see Assumption 2). This assumption is likely to be overly strong for many macroeconomic panels and may lead, if violated, to erroneous conclusions (cf. O'Connell, 1998). We therefore suggest to extend the tests developed here to allow for cross-sectional correlation by, e.g., the bootstrap method. Maddala and Wu (1999) report encouraging results along these lines for their panel unit root test. There is a growing literature on bootstrapping cointegrating regressions (see Li and Maddala, 1997) that can be fruitfully applied to the present problem. Recent useful contributions include Chang and Park (2003) and Chang, Park, and Song (2006). We turn to this question in the next chapter.

Chapter 4

Cross-Sectional Correlation Robust Tests for Panel Cointegration

Abstract

We use meta analytic combination procedures to develop new tests for panel cointegration. The main idea consists in combining p-values from time series cointegration tests on the different units of the panel. The tests are robust to heterogeneity as well as to cross-sectional dependence between the different units of the panel. To achieve the latter, we employ a sieve bootstrap procedure with joint resampling of the residuals of the different units. A simulation study shows that the suggested bootstrap tests can have substantially smaller error-in-rejection probabilities than tests ignoring the presence of cross-sectional dependence while preserving high power. We apply the tests to a panel of Post-Bretton Woods data to test for weak Purchasing Power Parity (PPP).[1]

Keywords: panel cointegration tests, cross-sectional dependence, sieve bootstrap

[1]I would like to thank conference participants at the FEMES 2006, Beijing and the Statistische Woche 2006, Dresden, and Pavel Stoimenov for helpful comments.

4.1 Introduction

The application of unit root and cointegration tests in macroeconometric practice is often hampered by the lack of large sample sizes. The (first-order) asymptotic approximation used for deriving the distribution of the test statistics may then be rather inaccurate. One solution to improve the power and reduce the error-in-rejection probability (size distortion) of these tests is to pool several time series into a panel data set and develop panel unit root and cointegration tests.

Pedroni (2004) and Kao (1999) generalize the residual-based tests of Engle and Granger (1987) and Phillips and Ouliaris (1990), Larsson, Lyhagen, and Löthgren (2001) extend the Johansen (1988) tests to panel data while McCoskey and Kao (1998) propose a test for the null of panel cointegration in the spirit of Shin (1994). All these tests, however, rely on the assumption that the different cross-sectional units of the panel are independent or, at most, exhibit dependence of a rather simple form. This assumption, characterizing the so-called first generation tests, greatly simplifies the derivation of limiting distributions of the panel test statistics, but may not hold in practice. Phillips and Sul (2003), Moon and Perron (2004) and Bai and Ng (2004) put forward factor approaches to deal with cross-sectional correlation in panel unit root and cointegration testing. Their approach may, however, require the validity of the factor structure assumption modelling the correlation structure of the panel units. As argued by Breitung and Das (2005), size distortions may result if this assumption, which is hard to verify, is not met.

The main contribution of the present chapter therefore is to suggest new tests for panel cointegration that are robust to cross-sectional dependence (or, synonymously, cross-sectional correlation) of an arbitrary form. The main idea of the testing principle has been used in meta analytic studies for a long time (see

Fisher, 1970; Hedges and Olkin, 1985) and was introduced into the panel literature by Maddala and Wu (1999) and Choi (2001), who propose meta analytic panel unit root tests: Consider the testing problem on the panel as consisting of N testing problems for each unit of the panel. Conduct N separate time series tests and obtain the corresponding p-values of the test statistics. Then, combine the p-values of the N tests (in a sense to be made precise below) into a single panel test statistic. Chapter 3 extends their framework to the panel cointegration setting.

To robustify the panel cointegration tests against cross-sectional correlation of an arbitrary form, we use a bootstrap scheme that jointly resamples entire cross-sections of residuals to preserve the cross-sectional correlation structure in the panel. We provide some simulation evidence to demonstrate the effectiveness of the suggested procedure. In particular, the bootstrap tests can have dramatically smaller error-in-rejection probabilities than tests ignoring the presence of cross-sectional dependence. At the same time, the bootstrap tests preserve high power. We use the tests to investigate the weak PPP hypothesis for a panel of Post-Bretton Woods exchange rate data. Our main result is that using cross-sectional correlation corrected critical values may make an important difference in econometric practice.

The remainder of the chapter is organized as follows. The next section establishes notation and reviews the meta analytic p-value combination tests for panel cointegration under the assumption of cross-sectional independence. Section 3 discusses the bootstrap algorithm used to robustify the tests against general forms of cross-sectional dependence. Section 4 summarizes the simulation evidence on the effectiveness of the bootstrap tests under cross-sectional dependence. Section 5 illustrates the use of the tests. The final section concludes.

4.2 *P*-Value Combination Tests for Panel Cointegration

Consider the multivariate time series regression

$$y_{i,t} = \alpha_i + \kappa_i t + \vartheta_i t^2 + \boldsymbol{\beta}_i \mathbf{x}_{i,t} + u_{i,t} \quad (t \in \mathbb{N}_{T_i}) \tag{4.1}$$

for each of the N units of a possibly unbalanced panel. ($a \in \mathbb{N}_c^b$ is shorthand for $a = b, \ldots, c$, omitting b if $b = 1$.) The $(K \times 1)$ column vector $\mathbf{x}_{i,t} = (x_{i1t}, \ldots, x_{iKt})^\top$ $(i \in \mathbb{N}_N,\ t \in \mathbb{N}_{T_i})$ collects the observations on the K regressors for given i and t. $y_{i,t}$ and $\mathbf{x}_{i,t}$ are integrated of order one, $I(1)$ $(i \in \mathbb{N}_N)$. The row vector $\boldsymbol{\beta}_i$, α_i, κ_i and ϑ_i may vary across i, allowing for heterogeneous cointegrating vectors and time polynomials of order up to two, i.e. constants, trend and squared trend terms. We furthermore make the following Functional Central Limit Assumption on the Data Generating Process (DGP) of the variables.

Assumption 1 *(Invariance Principle).*
Let $z_{i,t} := (y_{i,t}, \mathbf{x}_{i,t}^\top)^\top$ and $\xi_{i,t} := (\xi_{iyt}, \xi_{i1t}, \ldots, \xi_{iKt})^\top$. The true process $z_{i,t}$ is generated as $z_{i,t} = z_{i,t-1} + \xi_{i,t}$, $(i \in \mathbb{N}_N, t \in \mathbb{N}_{T_i})$.
$\xi_{i,t}$ satisfies $T_i^{-\frac{1}{2}} \sum_{t=1}^{[T_i r]} \xi_{i,t} \Rightarrow B_{i,\Omega_i}(r)$ $(i \in \mathbb{N}_N)$ as $T_i \to \infty$, where $r \in [0,1]$ and $\Rightarrow$ denotes weak convergence. $[x]$ is the integer part of x and $B_{i,\Omega_i}(r)$ is vector Brownian motion with asymptotic covariance matrix Ω_i. Also, the $K \times K$ lower right submatrix of Ω_i, $\Omega_{\mathbf{xx},i}$, has full rank.

This assumption ensures, among other things, that there are no cointegrating relationships among the regressors in (4.1).[2] p_i denotes the p-value of a time series cointegration test applied to the ith unit of the panel. Let θ_{i,T_i} be a

[2] See Pedroni (2004) for further discussion.

time series cointegration test statistic on unit i for a sample size of T_i. Let F_{T_i} denote the null cumulative distribution function (cdf) of θ_{i,T_i}. Since the tests considered here are one-sided, $p_i = F_{T_i}(\theta_{i,T_i})$ if the test rejects for small values of θ_{i,T_i} and $p_i = 1 - F_{T_i}(\theta_{i,T_i})$ if the test rejects for large values of θ_{i,T_i}. We only consider time series tests with the null of no cointegration.

We test the following null hypothesis:

$$H_0: \text{ There is no cointegration for any } i, \qquad (i \in \mathbb{N}_N)$$

against the alternative

$$H_1: \text{ There is cointegration for at least one } i.$$

Under H_0, $\{u_{i,t}\}_t$ in (4.1) is an $I(1)$ stochastic process $(i \in \mathbb{N}_N)$. The alternative H_1 states that there are 1 to N cointegrated units in the panel. That is, a rejection neither allows to conclude that the entire panel is cointegrated nor does it provide information about the number of units of the panel that exhibit cointegrating relationships.

We make the following assumptions (see Pedroni, 2004):

Assumption 2 *(Continuity).*
Under H_0, θ_{i,T_i} has a continuous distribution function $(i \in \mathbb{N}_N)$.

Assumption 3 *(Cross-Sectional Uncorrelatedness).*
$\mathsf{E}[\xi_{i,t}\xi_{l,s}^\top] = \mathbf{0}$ $(s, t \in \mathbb{N}_{T_i}, i \neq l)$. The error process $\xi_{i,t}$ is generated as a linear vector process $\xi_{i,t} = C_i(L)\eta_{i,t}$, where L is the lag operator and $\eta_{i,t}$ is vector white noise.

Remarks

- Assumption 2 is a regularity condition which ensures a uniform distribution of the p-values of the time series test statistics under H_0: $p_i \sim$

$\mathcal{U}[0,1]$ ($i \in \mathbb{N}_N$) (see, e.g., Bickel and Doksum, 2001, Sec. 4.1). It is satisfied by the time series tests considered in this chapter.

- Assumption 3 is strong (see, e.g., Banerjee, Marcellino, and Osbat, 2005). It implies that the different units of a panel must not be linked to each other beyond relatively simple forms of correlation such as common time effects. These can be eliminated by demeaning across the cross-sectional dimension. This assumption is likely to be violated in many typical macroeconomic panel data sets. For instance, consider a panel data set consisting of exchange rates vis-à-vis the U.S. dollar. The exchange rates of, say, the Euro and the Mexican Peso generally do not react identically to a macroeconomic shock in the U.S., given the very different structure of financial and trade links with the U.S.

- We emphasize that Assumption 3 is sufficient, but not necessary. Section 3 presents an approach that allows to dispense with this assumption.

We now present the test statistics for panel cointegration put forward in this chapter. Combine the N p-values of the individual time series cointegration tests, p_i ($i \in \mathbb{N}_N$), as follows:

$$P_{\chi^2} = -2\sum_{i=1}^{N} \ln(p_i) \tag{4.2a}$$

$$P_{\Phi^{-1}} = N^{-\frac{1}{2}}\sum_{i=1}^{N} \Phi^{-1}(p_i), \tag{4.2b}$$

where Φ^{-1} denotes the inverse of the cdf of the standard normal distribution.[3] When considered as a group we refer to Eqs. (4.2a) and (4.2b) as P tests. Furthermore, we refer to P tests relying on Assumption 3 as "simple" P tests. The P tests, via pooling p-values, provide convenient tests for panel cointegration by imposing minimal homogeneity restrictions on the panel. For instance, the

[3]See also Maddala and Wu (1999) and Choi (2001).

different units of the panel can be unbalanced. Furthermore, the evidence for (non-)cointegration is first investigated for each unit of the panel and then compactly expressed with the p-value of the time series cointegration test. Hence, the coefficients describing the relationship between the different variables for each unit of the panel can be heterogeneous across i. Thus, the availability of large-T time series allows for pooling the data into a panel without having to impose strong homogeneity restrictions on $\boldsymbol{\beta}$ as in traditional panel data analysis.[4]

We now turn to the asymptotic distributions of the tests.

THEOREM 1.

Under the null of no panel cointegration and Assumptions 1, 2 and 3, as $T_i \to \infty$ $(i \in \mathbb{N}_N)$, the P tests are asymptotically distributed as

$$(i) \;\; P_{\chi^2} \to_d \chi^2_{2N}$$

$$(ii) \;\; P_{\Phi^{-1}} \to_d \mathcal{N}(0,1).$$

PROOF. (i) The proof is an application of the transformation theorem for absolutely continuous random variables (r.v.s) (see, e.g., Bierens, 2005, Thm. 4.2). Under H_0, $p_i \sim \mathcal{U}[0,1]$. Let $y = g(p_i) := -2\ln(p_i)$. Then, $p_i = g^{-1}(y) = e^{-\frac{1}{2}y}$ and the density of $-2\ln(p_i)$ is given by

$$f_{-2\ln(p_i)}(y) = f_{p_i}(g^{-1}(y)) \left| \frac{\partial}{\partial y} g^{-1}(y) \right|.$$

Hence, $\frac{\partial}{\partial y} g^{-1}(y) = -\frac{1}{2}e^{-\frac{1}{2}y}$ and $|\frac{\partial}{\partial y} g^{-1}(y)| = \frac{1}{2}e^{-\frac{1}{2}y}$. We have $f_{p_i}(g^{-1}(y)) = 1 \; \forall \; g^{-1}(y) \in [0,1]$. This implies $f_{-2\ln(p_i)}(y) = \frac{1}{2}e^{-\frac{1}{2}y}$. The density of a χ^2_R r.v. is $f_{\chi^2_R}(y) = \frac{1}{2^{R/2}\Gamma(R/2)} y^{\frac{R}{2}-1} e^{-\frac{y}{2}}$. With $R = 2$, we get $f_{\chi^2_2}(y) = \frac{1}{2\Gamma(1)} e^{-\frac{y}{2}}$. Recall that $\Gamma(1) = \int_0^\infty t^{1-1}e^{-t}\,dt = 1$. So,

$$f_{\chi^2_2}(y) = \frac{1}{2}e^{-\frac{y}{2}}.$$

We have shown that $f_{-2\ln(p_i)}(y) = f_{\chi^2_2}(y)$. The proof is complete since the sum of N *independent* χ^2_R r.v.s is distributed as χ^2_{NR}.

(ii) Since, under H_0, $\mathsf{P}(\Phi^{-1}(p_i) \leqslant x) = \mathsf{P}(p_i \leqslant \Phi(x)) = \Phi(x)$, we have $\Phi^{-1}(p_i) \sim \mathcal{N}(0,1)$. By the convolution theorem (see, e.g., Spanos, 1986, pp. 99), the sum of N *independent* $\mathcal{N}(0,1)$ r.v.s has a $\mathcal{N}(0,N)$ distribution. Hence, $P_{\Phi^{-1}}$ is also normal with

$$\mathsf{E}[P_{\Phi^{-1}}] = 0 \;\text{ and }\; \mathsf{Var}[P_{\Phi^{-1}}] = \mathsf{Var}\left[N^{-\frac{1}{2}} \sum\nolimits_{i=1}^{N} \Phi^{-1}(p_i)\right] = \frac{1}{N}\mathsf{Var}\left[\sum\nolimits_{i=1}^{N} \Phi^{-1}(p_i)\right] = 1.$$

$\square$

[4] For an overview of panel data models relying on $N \to \infty$ asymptotics see Hsiao (2003).

Using consistent (as $T_i \to \infty$) time series cointegration tests, $p_i \to_p 0$ under the alternative of cointegration. Hence, quite intuitively, the smaller p_i, the more it contributes towards rejecting the null of no panel cointegration. The decision rule therefore is to reject the null of no panel cointegration when the realized test statistic P_{χ^2} exceeds the critical value from a χ^2_{2N} distribution at the desired significance level. For (4.2b) one would reject for large negative values of the panel test statistic $P_{\Phi^{-1}}$. We also see from the proof that the tests have a well-defined asymptotic distribution for any finite N. This feature is attractive because in many applications of panel cointegration analysis like the above example, the assumption of N, the number of units in the panel, going to infinity may not be a natural one. To rationalize the alternative hypothesis H_1, note that a small fraction of the units of the panel exhibiting strong evidence of time series cointegration, thus yielding low p-values, might lead to a rejection of the null hypothesis. Thus, as stated above, rejection of H_0 should not be taken as evidence of the entire panel being cointegrated.

We now discuss how to obtain the p-values required for computation of the P test statistics. The null distributions of residual based cointegration tests generally converge to functionals of Brownian motion. Hence, analytic expressions of the distribution functions are not available and p-values of the tests cannot simply be obtained by evaluating the corresponding cdf. In the time series case, it is now fairly standard practice to obtain p-values of unit root and cointegration tests using response surface regressions. We use p-values of the Augmented Dickey-Fuller (ADF) cointegration tests (Engle and Granger, 1987) as provided by MacKinnon (1996).[5] Suppressing deterministic trend terms for brevity, the p-values are derived from the t-statistic of $\varrho_i - 1$ in the

[5]MacKinnon improves upon his prior work by using a heteroskedasticity and serial correlation robust technique to approximate between the estimated quantiles of the response surface regressions.

OLS regression

$$\Delta\hat{u}_{i,t} = (\varrho_i - 1)\hat{u}_{i,t-1} + \sum_{j=1}^{J_i} \nu_j \Delta\hat{u}_{i,t-j} + \epsilon_{J_i,i,t}, \tag{4.3}$$

where $\Delta := 1 - L$ and J_i is the number of lagged differences required to render $\epsilon_{J_i,i,t}$ white noise. Here, $\hat{u}_{i,t}$ is the usual OLS residual from the first stage Engle and Granger (1987) regression (4.1). However, as should be clear from the above discussion, the P tests are general enough to accommodate any time series cointegration test for which p-values are available. Alternatively, one could, for instance, capture serial correlation by the semiparametric approach of Phillips and Ouliaris (1990).[6] We focus on the Engle and Granger (1987) ADF cointegration test because of its popularity and widespread availability.

4.3 Allowing for cross-sectional error dependence

We now relax Assumption 3. As can be seen from the proof of Theorem 1, this assumption guarantees the correct null distributions of the test statistics. Theorem 1 no longer holds under general forms of cross-sectional dependence. We suggest a bootstrap approach to capture the dependence structure in the panel with the aim to construct panel cointegration tests robust to the presence of cross-sectional correlation. We employ the sieve bootstrap.[7] The sieve bootstrap approximates $u_{i,t}$ with a finite order autoregressive process, where the order increases with sample size, and resamples from the residuals. Under the following assumption, the sieve bootstrap yields an accurate approximation

[6] In Chapter 3, we also employed the p-values for the Johansen (1988) λ_{trace} and λ_{max} tests. The corresponding P tests perform poorly, however, as the Johansen (1988) time series cointegration tests severely overreject for the time series lengths usually available in macroeconometric practice. Accordingly, the corresponding p-values are not distributed as $\mathcal{U}[0,1]$ under H_0 and hence not suitable for the P tests. Chapter 3 also discusses other p-value combination tests and provides more extensive Monte Carlo evidence on the simple P tests.

[7] For related approaches, see Maddala and Wu (1999) and Swensen (2003).

(Chang and Park, 2003):

> Assumption 4 *(Linearity).*
>
> The first differences of the equilibrium errors are generated as (possibly heterogeneous) linear processes, $\Delta u_{i,t} = \phi_i(L)\epsilon_{i,t}$, where $\phi_i(z) := \sum_{\ell=0}^{\infty} \phi_{i,\ell} z^\ell$.

More precisely, the bootstrap algorithm proceeds as follows.

1. Compute the P test statistic(s) according to (4.2a) and (4.2b). Denote the realizations by $\tilde{P}_{\chi^2}$ and $\tilde{P}_{\Phi^{-1}}$.
2. Suppressing deterministic trend terms and denoting estimates by a $\hat{\ }$, estimate equation (4.1) by OLS:

$$y_{i,t} = \hat{\alpha}_i + \hat{\boldsymbol{\beta}}_i \mathbf{x}_{i,t} + \hat{u}_{i,t}. \qquad (i \in \mathbb{N}_N,\ t \in \mathbb{N}_{T_i})$$

3. Fit an autoregressive process to $\Delta\hat{u}_{i,t}$ $(i \in \mathbb{N}_N, t \in \mathbb{N}_{T_i}^2)$. It is natural to use the Yule-Walker procedure because it always yields an invertible representation (Brockwell and Davis, 1991, Secs. 8.1–2). Letting $\overline{\Delta\hat{u}_i} := (T_i - 1)^{-1} \sum_{t=2}^{T_i} \Delta\hat{u}_{i,t}$, compute

$$\hat{\gamma}_i(j) := \frac{1}{T_i - 1 - j} \sum_{t=2}^{T_i - j} (\Delta\hat{u}_{i,t} - \overline{\Delta\hat{u}_i})(\Delta\hat{u}_{i,t+j} - \overline{\Delta\hat{u}_i}), \qquad (i \in \mathbb{N}_N,\ j \in \mathbb{N}_q)$$

the empirical autocovariances of $\Delta\hat{u}_{i,t}$ up to order q. Defining

$$\hat{\Gamma}_{i,q} := \begin{pmatrix} \hat{\gamma}_i(0) & \cdots & \hat{\gamma}_i(q-1) \\ \vdots & \ddots & \vdots \\ \hat{\gamma}_i(q-1) & \cdots & \hat{\gamma}_i(0) \end{pmatrix}$$

and $\hat{\gamma}_i := (\hat{\gamma}_i(1), \ldots, \hat{\gamma}_i(q))^\top$, obtain the AR coefficient vector as

$$(\hat{\phi}_{q,i,1}, \ldots, \hat{\phi}_{q,i,q})^\top := \hat{\Gamma}_{i,q}^{-1} \hat{\gamma}_i. \qquad (i \in \mathbb{N}_N)$$

4. The residuals are, as usual, given by

$$\hat{\epsilon}_{q,i,t} := \Delta\hat{u}_{i,t} - \sum_{\ell=1}^{q} \hat{\phi}_{q,i,\ell} \Delta\hat{u}_{i,t-\ell}. \quad (i \in \mathbb{N}_N, t \in \mathbb{N}_{T_i}^{q+2})$$

Center $\hat{\epsilon}_{q,i,t}$ to obtain

$$\tilde{\epsilon}_{q,i,t} := \hat{\epsilon}_{q,i,t} - \frac{1}{T_i - q - 1} \sum_{g=q+2}^{T_i} \hat{\epsilon}_{q,i,g}. \quad (i \in \mathbb{N}_N, t \in \mathbb{N}_{T_i}^{q+2})$$

5. Resample nonparametrically from $\tilde{\epsilon}_{q,i,t}$ to get $\epsilon^*_{q,i,t}$. To preserve the empirical cross-sectional dependence structure, jointly resample residual vectors

$$\tilde{\epsilon}_{q,\bullet,t} := (\tilde{\epsilon}_{q,1,t}, \ldots, \tilde{\epsilon}_{q,N,t}). \qquad (t \in \mathbb{N}_{T_i}^{q+2})$$

6. Recursively construct the bootstrap samples as[8]

$$\Delta u^*_{q,i,t} = \sum_{\ell=1}^{q} \hat{\phi}_{q,i,\ell} \Delta u^*_{q,i,t-\ell} + \epsilon^*_{q,i,t}. \quad (i \in \mathbb{N}_N, t \in \mathbb{N}_{T_i}^{q+2})$$

7. It is necessary to impose the null of a unit root when generating the artificial data in bootstrap unit root tests to achieve consistency (Basawa, Mallik, McCormick, Reeves, and Taylor, 1991). Accordingly, impose the null of non-cointegration by integrating $\Delta u^*_{i,t}$ to obtain $u^*_{i,t}$ and form

$$y^*_{q,i,t} = \hat{\alpha}_i + \hat{\boldsymbol{\beta}}_i \mathbf{x}_{i,t} + u^*_{q,i,t}. \qquad (i \in \mathbb{N}_N, t \in \mathbb{N}_{T_i})$$

8. Perform the P tests using the artificial data set $(y^*_{q,i,t}, \mathbf{x}_{i,t}^\top)^\top$. Denote the realizations of the test statistics by, e.g., $P^{b*}_{\chi^2}$.
9. Repeat steps 4 to 8 many, say B, times.
10. Denote the indicator function by $\mathbf{1}\{\ \}$ and choose a significance level α. Then, reject H_0 of the P_{χ^2} or the $P_{\Phi^{-1}}$ test if

$$\frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\{P^{b*}_{\chi^2} > \tilde{P}_{\chi^2}\} < \alpha \qquad \text{or} \qquad \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\{P^{b*}_{\Phi^{-1}} < \tilde{P}_{\Phi^{-1}}\} < \alpha, \quad (4.4)$$

respectively.

[8]We run the recursion for 30 initial observations before using the $\Delta u^*_{q,i,t}$ to mitigate the effect of initial conditions.

REMARKS

- We provide no formal proof of the consistency of this bootstrap procedure. It might be conjectured from the proofs of bootstrap consistency for unit root tests (Swensen, 2003; Chang and Park, 2003) and for inference in cointegrating regressions (Chang, Park, and Song, 2006). The latter authors argue that their results may hold more generally, e.g., for panel cointegration models.

- Steps 4 and 6 respectively "prewhiten" (or "sieve") and "recolour" the residuals using the sieve bootstrap. Thus, we attempt to generate a valid bootstrap distribution of the data across the time series dimension using a parametric AR approximation to the true DGP. There is, however, no plausible parametric approximation of the dependence structure across the cross-sectional dimension. We therefore resample entire cross-sections of residuals to preserve the cross-sectional dependence structure of the data. The resampling scheme across the cross-sectional dimension is thus similar to block bootstrap procedures (Künsch, 1989).

- The selection of the lag order q in step 3 can be based on any of the well-known selection criteria such as the Akaike Information Criterion or a top-down procedure. The goal of the sieve bootstrap is to prewhiten the residuals across t to obtain random resamples from $\tilde{\epsilon}_{q,i,t}$. Hence, a selection scheme based on the whiteness of $\tilde{\epsilon}_{q,i,t}$ is also an appealing choice.

- It is possible to let q vary over i, $q_i \neq q$, to capture heterogeneity in the error processes. We do not make this possibility explicit in the notation.

- It is also possible to compute the bootstrap P tests based on resamples from prewhitened residuals of a VAR regression of $(\Delta \hat{u}_{i,t}, \Delta \mathbf{x}_{i,t}^{\top})^{\top}$. For a related purpose, Chang, Park, and Song (2006) advocate a similar scheme in order to capture endogeneity of the regressors $\mathbf{x}_{i,t}$. Our simulation results are however similar to the ones to be presented for the sieve bootstrap in the next section. The VAR approach is computationally considerably more expensive and is therefore not discussed in detail. Results are available upon request.

4.4 A Monte Carlo study

We perform a Monte Carlo study of the tests proposed in the previous sections. The main results are as follows. The simple P tests can have high errors in rejection probabilities (ERPs) when the units of the panel exhibit cross-sectional correlation of a general form.[9] Compared to the simple P tests, bootstrapping the P tests as outlined in the previous section can strongly reduce the ERP.

The DGP used here is an extension of a design which has been used in many Monte Carlo studies of (panel) cointegration tests. See Engle and Granger (1987) and, for the extension to the panel data setting, Kao (1999). For simplicity, consider the bivariate case, i.e. $K = 1$.

$$y_{i,t} - \alpha_i - \beta_i x_{i,t} = v_{i,t}, \quad a_1 y_{i,t} - a_2 x_{i,t} = w_{i,t} \quad (i \in \mathbb{N}_N) \tag{4.5}$$

where

$$\begin{aligned} v_{i,t} &= \rho_i v_{i,t-1} + \breve{e}_{zi,t}, \qquad \Delta w_{i,t} = \breve{e}_{wi,t}, \\ \breve{e}_{zi,t} &= e_{zi,t} + \lambda_i \varepsilon_t \qquad\quad \breve{e}_{wi,t} = e_{wi,t} + \pi_i e_{wi,t-1} \end{aligned}$$

and

$$\begin{pmatrix} e_{zi,t} \\ e_{wi,t} \\ \varepsilon_t \end{pmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \psi\sigma & 0 \\ \psi\sigma & \sigma^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

[9] Suppose the DGP is in the null hypothesis set, the tests are performed at a nominal level α and the rejection frequency, or type I error rate, of the test is $R(\alpha)$. Then, $ERP := |R(\alpha) - \alpha|$. The term size distortion is often used synonymously (see Davidson and MacKinnon, 1999).

Remarks

- When $|\rho_i| < 1$ the equilibrium error $v_{i,t}$ in (4.5) is stationary such that $y_{i,t}$ and $x_{i,t}$ are cointegrated with $(1 \quad -\alpha_i \quad -\beta_i)^\top$. Further, ρ_i need not be constant across i. We do, however, choose a common ρ in the simulations to limit the number of experiments and to facilitate the interpretation of the results.

- Solving the system of equations (4.5) for $x_{i,t}$, we can write

$$x_{i,t} = \frac{a_1\alpha_i + a_1 v_{i,t} - w_{i,t}}{a_2 - a_1\beta_i}.$$

  Thus, $x_{i,t}$ is weakly exogenous when $a_1 = 0$.

- The panel is cross-sectionally dependent because of the common factor ε_t and the idiosyncratic factor loadings $\lambda_i \sim \mathcal{U}[\zeta_1, \zeta_2]$, where $\mathcal{U}[\zeta_1, \zeta_2]$ denotes the uniform distribution with lower bound ζ_1 and upper bound ζ_2.[10] Similar factor structures have been employed by Bai and Ng (2004) and Phillips and Sul (2003). Even though we generate cross-sectional dependence via a single factor, the validity of the suggested bootstrap tests does not depend on knowledge of the latent factor structure.

- Since $\lambda_i \neq \lambda = \text{cst.}$ it is not possible to remove the cross-sectional dependence by subtracting time-specific means $N^{-1}\sum_i z_{i,t}$, as would be possible under the stronger assumptions of, e.g., Westerlund (2005).

- If $\pi_i \neq 0$ there is a moving-average component in the errors. In particular, values $-1 < \pi_i < 0$ are well-known to have a potentially severe size-distorting effect on unit root and cointegration tests (Schwert, 1989).

[10] Uniform random numbers are generated using the KM algorithm from which Normal variates are created with the fast acceptance-rejection algorithm, both implemented in GAUSS. Part of the calculations are performed with COINT 2.0 by Peter Phillips and Sam Ouliaris.

The dimensions of the panel are $N \in \{10, 30\}$ and, after having discarded 75 initial observations, $T \in \{50, 100\}$. These are representative for the dimensions of data sets often encountered in macroeconometric applications. For a given cross-sectional dimension we draw the unit specific intercepts as $\alpha_i \sim \mathcal{U}[0, 5]$ and keep them fixed for both T. We choose $\beta_i = \mathcal{U}[1, 2]$, $a_2 = -1$ and $\sigma = 1$ and investigate all combinations of the following values for the parameters of the model: $a_1 \in \{0, 1\}$, $\psi \in \{0, 0.5\}$, $\pi \in \{-0.5, 0\}$ and $\rho \in \{0.8, 0.9, 1\}$. Further, there are three different degrees of cross-sectional dependence: $(\zeta_1, \zeta_2) = (0, 0)$, $(\zeta_1, \zeta_2) = (0, 1)$ and $(\zeta_1, \zeta_2) = (1, 4)$, corresponding to no, "weak" and "strong" cross-sectional dependence (see Mark, Ogaki, and Sul, 2005). That is, 72 experiments are conducted for each of the 4 panel dimensions.

To limit the computational burden, we use $M = 1,000$ replications with $B = 1,000$ bootstrap resamples in each.[11] The p-values are obtained from the Engle and Granger (1987) ADF test, selecting the number of lagged differences J_i in the second stage ADF regressions (4.3) according to the automatic procedure suggested by Ng and Perron (2001). We record a rejection if (i) the P_{χ^2} test statistic exceeds the 5% critical value of the χ^2_{2N} distribution, the $P_{\Phi^{-1}}$ test statistic falls below the 5% quantile of the cdf of the standard normal distribution, -1.645, or (ii) for the bootstrap version of the tests, if equation (4.4) applies.

To summarize, the DGP used here simultaneously addresses several issues that have proved both relevant for empirical work and challenging for unit root and cointegration tests. Hence, while cautioning that interpretation of Monte Carlo results should strictly speaking be confined to the DGP at hand (see Horowitz and Savin, 2000), we are optimistic that tests which perform well under this

[11] To gauge the sampling variability of the experiments we perform selected experiments several times. The rejection rates never differed by more than 2 percentage points and usually by less. This suggests the number of replications is sufficient.

Table 4.1: Rejection Rates of the P Tests.

		P_{χ^2}						$P_{\Phi^{-1}}$					
	N	10			30			10			30		
	T	BS_2	BS_s	Sim	BS_2	BS_s	Sim	BS_2	BS_s	Sim	BS_2	BS_s	Sim
(a) $\lambda_i \sim \mathcal{U}[1,4]$						(i) $a_1 = 0$							
$\rho = 1$	50	.068	.049	.090	.092	.072	.141	.111	.081	.143	.127	.113	.193
	100	.099	.056	.102	.114	.082	.166	.125	.091	.142	.156	.119	.233
$\rho = .9$	50	.138	.092	.228	.156	.133	.313	.208	.157	.319	.235	.216	.452
	100	.433	.294	.548	.477	.341	.691	.521	.393	.679	.691	.437	.815
$\rho = .8$	50	.332	.267	.483	.396	.279	.632	.438	.379	.606	.502	.394	.753
	100	.842	.736	.913	.879	.753	.954	.890	.805	.950	.912	.825	.981
						(ii) $a_1 = 1$							
$\rho = 1$	50	.200	.084	.272	.387	.148	.470	.237	.150	.214	.451	.253	.379
	100	.256	.082	.162	.471	.128	.293	.289	.124	.180	.536	.248	.346
$\rho = .9$	50	.185	.101	.165	.369	.160	.272	.224	.152	.096	.443	.307	.191
	100	.273	.104	.078	.536	.143	.142	.327	.152	.076	.658	.334	.149
$\rho = .8$	50	.229	.116	.155	.408	.211	.206	.303	.202	.093	.546	.415	.106
	100	.452	.150	.138	.749	.300	.251	.550	.288	.162	.839	.593	.295
(b) $\lambda_i \sim \mathcal{U}[0,1]$						(i) $a_1 = 0$							
$\rho = 1$	50	.047	.021	.024	.044	.011	.008	.077	.043	.021	.086	.055	.009
	100	.103	.034	.035	.126	.036	.033	.127	.055	.033	.195	.087	.039
$\rho = .9$	50	.157	.090	.090	.328	.208	.174	.326	.250	.131	.627	.540	.297
	100	.763	.484	.599	.968	.873	.924	.903	.786	.774	.995	.980	.978
$\rho = .8$	50	.569	.423	.486	.922	.773	.862	.804	.732	.630	.987	.974	.948
	100	.999	.987	.998	1.00	1.00	1.00	.999	1.00	1.00	1.00	1.00	1.00
						(ii) $a_1 = 1$							
$\rho = 1$	50	.159	.057	.216	.289	.080	.322	.201	.119	.131	.383	.255	.182
	100	.237	.068	.114	.477	.099	.191	.289	.118	.127	.559	.239	.185
$\rho = .9$	50	.219	.113	.170	.404	.192	.275	.315	.229	.132	.590	.450	.188
	100	.469	.232	.244	.905	.551	.537	.626	.455	.308	.974	.867	.695
$\rho = .8$	50	.418	.227	.346	.790	.569	.625	.618	.470	.346	.942	.889	.630
	100	.972	.791	.908	1.00	.989	1.00	.993	.946	.963	1.00	1.00	1.00
(c) $\lambda_i \equiv 0$						(i) $a_1 = 0$							
$\rho = 1$	50	.030	.018	.015	.044	.006	.005	.059	.041	.011	.074	.033	.005
	100	.081	.030	.019	.098	.015	.013	.092	.049	.016	.172	.063	.010
$\rho = .9$	50	.145	.091	.068	.349	.165	.111	.332	.255	.091	.737	.605	.186
	100	.789	.512	.567	.996	.940	.958	.927	.797	.784	1.00	.999	.995
$\rho = .8$	50	.577	.367	.444	.953	.853	.848	.843	.742	.592	1.00	.998	.964
	100	1.00	.995	.998	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
						(ii) $a_1 = 1$							
$\rho = 1$	50	.157	.062	.196	.271	.085	.308	.198	.129	.132	.370	.153	.170
	100	.265	.071	.142	.496	.090	.209	.304	.110	.124	.564	.215	.199
$\rho = .9$	50	.241	.127	.198	.425	.205	.305	.327	.237	.143	.645	.510	.250
	100	.583	.319	.345	.946	.664	.709	.744	.551	.450	.990	.940	.853
$\rho = .8$	50	.481	.317	.405	.863	.649	.752	.680	.535	.435	.983	.947	.784
	100	.994	.890	.969	1.00	1.00	1.00	.999	.990	.995	1.00	1.00	1.00

Note: $\psi = 0.5$, $\pi_i = -0.5$. $M, B = 1,000$ replications and resamples.
Panels (a), (b) and (c) correspond to "strong", "weak" and no correlation.
Panels (i) and (ii) correspond to the exogenous and endogenous case.
BS_2, BS_s and Sim are the sieve bootstrap with $q = 2$, q data dependent, and simple, resp.

experimental design may be useful for applied studies.

Table 4.1 reports the rejection rates from the Monte Carlo study of the bootstrap and the simple P tests. For brevity, we only give the (representative) results for the serially correlated case ($\pi = -0.5$) and $\psi = 0.5$.[12] In order to illustrate the importance of a suitable lag order q when applying the sieve bootstrap we choose $q = 2$ (columns BS_2) and the data dependent rule $q = [4 \cdot (T/100)^{.25}]$ (columns BS_s). The columns "Sim" refer to the simple P tests.

The main findings may be summarized as follows. The simple P tests overreject in the presence of cross-sectional correlation. The strength of the dependence—see the rows $\rho = 1$ in panels (a) and (b)—does matter. The ERPs are particularly severe in the presence of endogenous regressors (see panels (ii) vs. (i)). Under "strong" correlation (panel (a)) and endogeneity ($a_1 = 1$), the simple P tests even seem to be biased as they reject more frequently for samples generated under H_1 than for samples generated under H_0. On the other hand, the difference in the performance of the tests between the uncorrelated case (c) and the weakly correlated case (b) is small, suggesting that correlation robust tests may be most expedient when one suspects strong forms of dependence in the data. Generally, the sieve bootstrap is capable of removing or at least substantially reducing the ERP. Size control usually is more effective in columns BS_s than in columns BS_2. Thus, as expected, it is necessary to suitably prewhiten the residuals with a data-dependent lag order selection scheme for q. This is in line with many other studies in the nonstationary panel literature. For instance, Hlouskova and Wagner (2006) find that selection of the lag length in panel unit root ADF regressions plays an important role for the behavior of many popular tests.

Concerning the power of the tests, consider rows $\rho = 0.9$, 0.8. Power increases

[12] The full set of results of the study is available upon request.

with T, as expected. The increase in power with growing N is more pronounced when the dependence in the data is smaller. This is intuitive because under strong dependence, the amount of independent information in the panel is smaller for a given N.

4.5 An Empirical Test of the PPP Hypothesis

In this section, we reconsider the Purchasing Power Parity (PPP) hypothesis that has attracted wide attention in the literature (see Taylor and Taylor, 2004, for a recent survey). Assuming the law of one price to hold at least in the long run, its absolute version implies that the ratio of domestic price level $P_{i,t}$ and foreign price level $P^*_{i,t}$ should be close to the exchange rate $S_{i,t}$. Equivalently, the real exchange rate $R_{i,t}$ should be near one. Empirically, denoting natural logs by lowercase letters, the PPP hypothesis therefore postulates that

$$r_{i,t} = p_{i,t} - p^*_{i,t} - s_{i,t} \qquad (i \in \mathbb{N}_N)$$

is a stationary process. Due to factors such as transportation costs, one often allows for non-unitary coefficients to obtain an equation of the form

$$p^*_{i,t} = a_i + \beta_{i1} p_{i,t} + \beta_{i2} s_{i,t} + e_{i,t}, \qquad (i \in \mathbb{N}_N) \qquad (4.6)$$

referred to as the weak PPP hypothesis. Since the variables typically are $I(1)$, the weak PPP hypothesis implies that $e_{i,t}$ is stationary, or, equivalently, that $p^*_{i,t}$, $p_{i,t}$ and $s_{i,t}$ are cointegrated.

We use quarterly Post-Bretton Woods data for a panel of OECD countries, observing the spot exchange rate and the consumer price index ranging from 1973:1 until 1998:2.[13] The numeraire country (with respect to which the spot exchange rate is sampled) is the United States. As has been pointed by, inter

[13]The data is from the IMF Financial Statistics CD-ROM. The list of included countries can be found in Table 4.2.

TABLE 4.2: TESTS FOR WEAK PURCHASING POWER PARITY

Countries	$\hat{a}_i$	$\hat{\beta}_{1i}$	$\hat{\beta}_{2i}$	J_i	t_{ϱ_i-1}
Australia	16.848	0.833	3.982	2	-1.912
Austria	-23.109	1.216	-0.115	2	-2.412
Belgium	-3.907	1.114	-0.205	1	-1.626
Canada	-5.292	0.943	10.491	1	-0.809
Denmark	12.461	0.974	-1.005	0	-0.751
Finland	10.224	0.915	0.455	3	-1.841
France	18.154	0.959	-2.014	1	-0.446
Germany	-40.631	1.425	-3.421	3	-2.778
Greece	38.940	-0.036	0.352	0	-2.273
Ireland	30.309	0.875	-23.705	0	-1.082
Italy	28.505	0.659	0.003	3	-2.815
Japan	6.847	1.098	-0.119	0	-1.222
Netherlands	-21.331	1.290	-6.165	0	-3.010
New Zealand	26.845	0.734	3.613	2	-1.727
Norway	11.707	0.888	0.489	1	-1.500
Portugal	38.443	0.383	0.155	3	-2.821
Spain	27.919	0.657	0.046	1	-2.340
Sweden	14.692	0.766	1.426	3	-2.423
Switzerland	-30.012	1.313	-2.622	0	-1.203
United Kingdom	16.231	0.819	5.693	1	-2.002
United States	numeraire country				

Panel Results		simple	bootstrap critical values for $q =$				
		cr. val	2	4	6	8	10
$\tilde{P}_{\chi^2}$	16.751	55.758	46.018	33.160	38.908	37.408	41.151
$\tilde{P}_{\Phi^{-1}}$	3.483	-1.645	-0.522	1.190	0.383	0.432	-0.052

Note: First 3 columns are estimates of (4.6), J_i are number of lagged differences chosen for unit root tests on $\hat{e}_{i,t}$. t_{ϱ_i-1} is t-statistic from (4.3). Bottom part gives realized test statistics (left), simple (middle) and bootstrap (right) 5% critical values for different AR orders q in the sieve bootstrap procedure.

alia, O'Connell (1998), the choice of a numeraire country naturally induces a common component in the set of regressions (4.6). It is therefore interesting to investigate the effect of using cross-sectional correlation robust tests of the PPP hypothesis. Table 4.2 reports the estimation results. The first three columns give the estimates of a_i, β_{i1} and β_{i2}. The results differ across countries, making it attractive to use a panel procedure that allows for heterogeneity. The number of lagged differences J_i chosen by the procedure of Ng and Perron (2001) is reported in column 4. The fifth column gives the results for the t-

statistics of the individual ADF cointegration regressions (4.3). All statistics fail to produce evidence in favor of PPP, clearly exceeding the critical value of -3.837. The Netherlands have the smallest statistic $t_{\varrho_{\text{Neth}-1}} = -3.010$, implying a p-value of 0.251 (MacKinnon, 1996). The estimates of the panel tests are reported in the bottom of Table 4.2. In view of the individual outcomes, it is not surprising that the simple and bootstrap p-value combination tests also do not reject the null hypothesis of no PPP.

Even though neither tests rejects the null, the critical values differ substantially. In particular, the bootstrap critical values are almost all closer to the realized test statistic, indicating that using the critical values of the simple tests would be overly conservative in the present application. We conclude that using cross-sectional correlation robust tests may well make an important difference in econometric practice.

4.6 Conclusion

We suggest new meta analytic p-value combination tests for panel cointegration. The tests are robust to cross-sectional dependence of an arbitrary form. To achieve robustness, a bootstrap procedure is used. Further advantages compared to other widely used panel cointegration tests are that they are flexible, intuitively appealing and comparatively easy to compute.

The simulation study reveals that the approach is generally effective in improving the reliability of the "simple" P tests. However, care is needed in properly selecting the lag order in the sieve bootstrap scheme. This finding is analogous to the well-known result that the size and power of time series unit root and cointegration tests are fairly sensitive to the selection of the correct number of lagged differences in the ADF regressions (or, selection of lag truncations in estimating long-run variances).

A further interesting finding is that the power gains from using panel tests are the smaller the stronger the cross-sectional correlation in the data. We interpret this as reflecting that combining correlated time series leads to less additional information than combining independent series. This could imply that the impressive power gains reported in simulations of first generation panel unit root tests vs. time series unit root tests are partially an artefact of the experimental design relying on independence of the units. Our empirical study of the weak PPP hypothesis reveals that cross-sectional correlation robust critical values for the P tests can differ substantially from their simple counterparts. It would be interesting to investigate how other commonly used panel cointegration tests such as those developed by Pedroni (2004) or Kao (1999) perform under general forms of cross-sectional correlation.

Another task for future research is to establish whether the bootstrap procedure suggested in this chapter is consistent in the sense of leading to the same limiting distributions for the bootstrap statistics as for the original ones. It may be possible to establish bootstrap consistency along the lines of Chang and Park (2003), Chang, Park, and Song (2006) and Swensen (2003).

Chapter 5

Are PPP Tests Erratic? Some Panel Evidence

Abstract

This chapter[1] examines whether, in addition to standard unit root and cointegration tests, panel approaches also produce test statistics behaving erratically when applied to tests for PPP. We show that if appropriate tests (which are robust to cross-sectional dependence and more powerful than single time series tests) are used, any evidence of erratic behaviour disappears, and strong empirical support is found for PPP.

Keywords: Purchasing Power Parity (PPP), Real Exchange Rates, Erratic Behaviour, Panel Tests

[1]This chapter has been written jointly with Guglielmo Maria Caporale.

5.1 Introduction

Purchasing Power Parity (PPP) is a key concept to the way international economists understand real exchange rate behaviour. Most of them would agree that PPP holds in the long run, if not continuously, at least in some form, and that therefore it represents a valid international parity condition (see, e.g., Taylor and Taylor, 2004, for a critical review of the PPP debate). However, the available empirical evidence has not always been consistent with the PPP condition. Given the wide consensus on the theory, this failure of formal tests to provide support to PPP has mainly been attributed to flaws in the econometric approaches taken. Froot and Rogoff (1995), in particular, highlighted the limitations of the tests used in three successive stages in the time series literature on PPP. Initially, possible non-stationarities were overlooked. Then the null that the real exchange rate follows a random walk (long-run PPP being the alternative) was tested by means of unit root tests which are now well-known to have very low power; cointegration methods, subsequently used, suffered from similar problems. Recently, Caporale, Pittis, and Sakellis (2003) have also argued that classical unit root tests are not informative about PPP. Specifically, they show that the type of stationarity exhibited by the real exchange rate cannot be accommodated by the fixed-parameter autoregressive homoscedastic models normally employed in the literature. In particular, they compute a recursive t-statistic, and show that it exhibits erratic behaviour, suggesting the presence of endemic instability, and of a type of non-stationarity more complex than the unit root one usually assumed. Similar results are reported in the case of trivariate cointegration tests in Chapter 2, where we conclude that the observed erratic behaviour is therefore not due to imposed symmetry/proportionality restrictions.

In order to increase the power of tests of PPP, more recent studies have used panel methods (see, e.g., Wu, 1996, and Papell, 1997, 2002).[2] The present chapter investigates whether erratic behaviour still occurs when panel approaches are taken. If erraticism is found to disappear once more powerful, panel tests are applied, one could then argue that the failure of earlier tests to give support to PPP theory was indeed due to their low power, rather than to incorrect assumptions about the dynamic features of the stochastic process of interest. In this case, panel tests, characterised by much higher power, could be seen as the way forward to settle the PPP debate. The layout of the chapter is the following. Section 2 outlines the panel methods used. Section 3 presents the empirical evidence. Section 4 summarises the main findings and offers some concluding remarks.

5.2 The Panel Tests

This section briefly describes the panel tests considered in this study. It is widely acknowledged that panels of exchange rate data are generally cross-sectionally dependent (O'Connell, 1998). Panel unit root tests relying on the assumption of cross-sectional independence (see, e.g., Levin, Lin, and Chu, 2002, Im, Pesaran, and Shin, 2003, or Choi, 2001) will therefore suffer from size distortion, as recently demonstrated in, for instance, Hlouskova and Wagner (2006). Accordingly, we focus on panel tests which are robust to the presence of cross-sectional dependence. More specifically, we consider the tests put forward by Choi (2006) and Phillips and Sul (2003).

[2] See Caporale and Cerrato (2006) for a critical survey of the empirical literature testing PPP by means of panel methods. Another new development in the literature on real exchange rates is the modelling of nonlinearities (resulting, for instance, from transaction costs – see Taylor, Peel, and Sarno, 2001) in mean reversion. Some studies also allow for structural breaks (see, e.g., Papell, 2002).

Choi (2006)

In the first step, the panel tests of Choi (2006) apply Elliott, Rothenberg, and Stock (1996) GLS detrending to the panel, thereby removing cross-sectional dependence. In the second step, meta-analytic panel tests from, e.g., Choi (2001) can then be used (see also Maddala and Wu, 1999).

Choi (2006) assumes the following two-way error-component model

$$y_{i,t} = \beta_0 + x_{i,t} \quad (i = 1, \ldots, N;\ t = 1, \ldots, T),$$

where

$$x_{i,t} = \mu_i + \lambda_t + v_{i,t},$$

and

$$v_{i,t} = \sum_{l=1}^{p_i} \alpha_{il} v_{i,t-l} + e_{i,t}$$

The test of a panel unit root is formulated as

$$H_0 : \sum_{l=1}^{p_i} \alpha_{il} = 1 \qquad i = 1, \ldots, N$$

against

$$H_1 : \sum_{l=1}^{p_i} \alpha_{il} < 1 \quad \text{for a non-zero fraction } \#i/N$$

The Elliott, Rothenberg, and Stock (1996) GLS estimator of β_0 is given by

$$\hat{\beta}_{0i} = \frac{y_{i,1} + \left(1 - \frac{7}{T}\right) \sum_{t=2}^{T} y_{i,t} - \left(1 - \frac{7}{T}\right) y_{i,t-1}}{1 + (T-1)\left(1 - (1 - \frac{7}{T})\right)^2}$$

Choi (2006) shows that demeaning $y_{it} - \hat{\beta}_{0i}$ cross-sectionally gives, for large T,

$$z_{i,t} := y_{i,t} - \hat{\beta}_{0i} - \frac{1}{N} \sum_{i=1}^{N} (y_{i,t} - \hat{\beta}_{0i}) \simeq v_{i,t} - v_{i,1} - \overline{v}_{.t} + \overline{v}_{.1},$$

where $\overline{v}_{.a} := \frac{1}{N} \sum_{i=1}^{N} v_{ia}$. This expression is independent of β_0, λ_t and μ_i. Moreover, $\overline{v}_{.t}$, $\overline{v}_{.1} \to_p 0$. Hence, $z_{i,t}$ is cross-sectionally independent.

In a second step, one applies meta-analytic panel tests to z_{it}. For instance, run Augmented Dickey-Fuller tests on z_{it}. Then, after having obtained the p-values of the test statistics[3], these may be combined into panel test statistics as follows:

$$P_m = -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\ln(p_i)+1) \tag{5.1}$$

$$Z = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Phi^{-1}(p_i) \tag{5.2}$$

$$L^* = \frac{1}{\sqrt{\pi^2 N/3}}\sum_{i=1}^{N}\ln\left(\frac{p_i}{1-p_i}\right), \tag{5.3}$$

where Φ is the standard normal cumulative distribution function. As $N, T \to \infty$, $P_m,\ Z,\ L^* \Rightarrow N(0,1)$. The tests are consistent because $P_m \to_p \infty$ and $Z,\ L^* \to_p -\infty$ under H_1.

Phillips and Sul (2003)

Phillips and Sul (2003) work with the dynamic panel representation

$$y_{i,t} = \mu_i(1-\rho) + \rho y_{i,t-1} + \sum_{j=1}^{\ell_i}\phi_{ij}\Delta y_{i,t-j} + u_{i,t}, \tag{5.4}$$

where $t = 1,\dots,T,\ i = 1,\dots,N$ and $\rho \in (-1,1]$. They model cross-sectional dependence with a standard normal common time effect θ_t which is allowed to affect the units of the panel heterogeneously:

$$u_{it} = \delta_i\theta_t + \epsilon_{it}.$$

The ϵ_{it} are normal with mean zero and variance σ_i^2. Letting $\mathbf{u}_t = (u_1,\dots,u_N)^\top$, $\boldsymbol{\delta} = (\delta_1,\dots,\delta_N)^\top$ and $\boldsymbol{\Sigma} = diag(\sigma_1^2,\dots,\sigma_N^2)$, we have $Cov(\mathbf{u}_t) = \boldsymbol{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}^\top$. To deal with the cross-sectional dependence in $\mathbf{u}_t$, Phillips and Sul (2003) suggest estimating $\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}$ by computing $M_T = \frac{1}{T}\sum_{t=1}^{T}\hat{\mathbf{u}}_t\hat{\mathbf{u}}_t^\top$, where $\hat{\mathbf{u}}_t$ is obtained

[3] In practice, one can evaluate the numerical distribution functions obtained by MacKinnon (1994, 1996) via response surface regressions.

from the residuals obtained under the null $\rho = 1$ in (5.4), and iteratively solving the system of equations

$$\hat{\boldsymbol{\delta}} = (M_T\hat{\boldsymbol{\delta}} - \hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\delta}})/\hat{\boldsymbol{\delta}}^\top\hat{\boldsymbol{\delta}}, \qquad \hat{\sigma}_i^2 = M_{Tii} - \hat{\delta}_i^2.$$

Using the orthogonal complement matrix $\hat{\boldsymbol{\delta}}_\perp$, one then computes the de-factored series $\mathbf{y}_t^+ = (\hat{\boldsymbol{\delta}}_\perp^\top\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\delta}}_\perp)^{-1/2}\hat{\boldsymbol{\delta}}_\perp^\top\mathbf{y}_t$. Phillips and Sul (2003) show that the above transformation asymptotically removes the dependence in $\mathbf{y}_t$ such that $\mathbf{y}_t^+$ is cross-sectionally independent.

It is then possible to perform, e.g., Fisher-type panel unit root tests:

$$P = -2\sum_{i=1}^{N-1}\ln(p_i), \tag{5.5}$$

using p-values from unit root tests applied to each series $y_{it}^+,\ i = 1, \ldots, N$. In practice, one can obtain the p-values as described in the previous subsection. Under H_0, $P \Rightarrow \chi^2_{2(N-1)}$. Under the alternative, $P \rightarrow_p \infty$.

5.3 Results

We now investigate whether using the panel unit root tests discussed above leads to erratic behaviour of the test statistics, namely whether there are frequent jumps from the rejection to the non-rejection region as new observations are successively added to the sample. We use the dataset also employed by Taylor (2002), which includes annual data for the nominal exchange rate, CPI and the GDP deflator. This dataset is particularly useful for our purposes because it covers a long period, ranging from 1892 through to 1996. The countries contained in our panel are: Argentina, Australia, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Mexico, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom. We use the United States as the reference country throughout (see Taylor, 2002, for further details

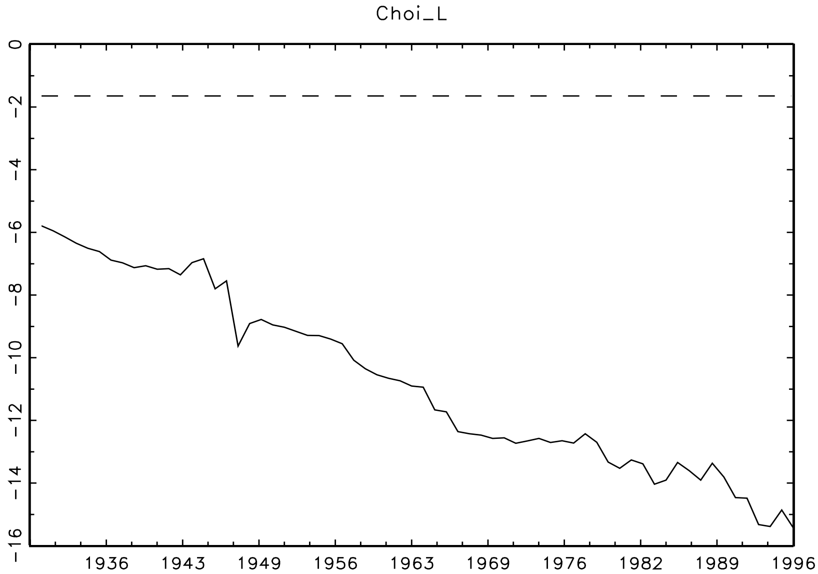


FIGURE 5.1: TEST STATISTIC SERIES FOR L^* FOR VARIOUS N

on data sources and definitions). In order to investigate possible parameter instability, we create a time series of test statistics resulting from the successive estimation of (5.1)–(5.3) and (5.5). That is, we use the first k observations to produce the first set of statistics, where we let $k = 40$ to discard estimates which are likely to be affected by small-sample distortions. We then add an extra observation to compute the second set of statistics based on $k + 1$ data points, and repeat the process until all T available observations have been used to yield $T - k + 1$ estimates of the test statistics.

We report results obtained using CPI data to construct the real exchange rate series in Figures 5.1 to 5.4, where we plot the test statistic series against the endpoint of the sample used to construct the statistics. The dashed lines indicate the appropriate critical values at the 5% level.[4]

It is fairly apparent that there is little evidence of erratic behaviour in the panel

[4]The findings were very similar when the GDP deflator was used instead of the CPI series (they are not reported here for the sake of brevity).

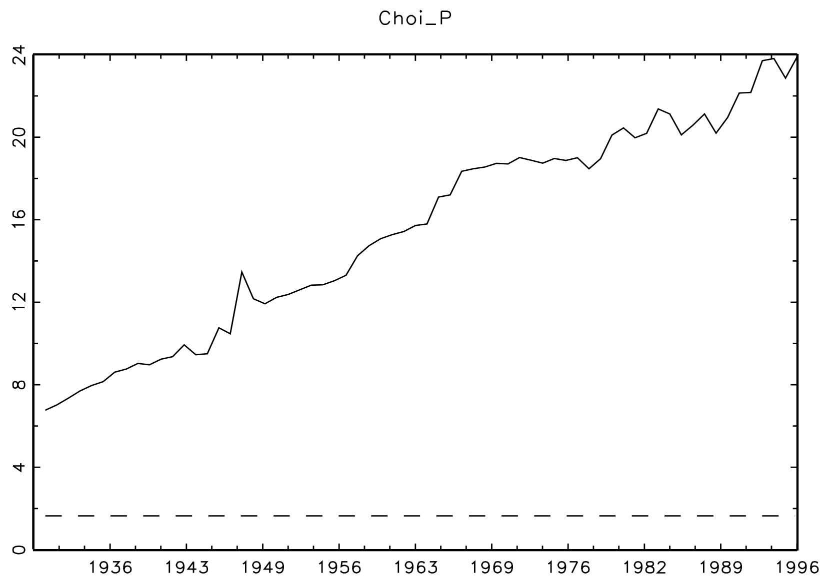


FIGURE 5.2: TEST STATISTIC SERIES FOR P_m FOR VARIOUS N

test statistic series. Rather, the test statistics seem to be approaching their respective probability limits under the alternative. In other words, it appears that using suitably designed (i.e., robust to cross-sectional dependence) panel tests, with much higher power compared to standard unit root tests, removes erraticism of the test statistics, and provides strong evidence in favour of PPP.

5.4 Conclusion

This chapter has examined whether, in addition to standard unit root and cointegration tests, panel approaches also produce test statistics behaving erratically when applied to testing for PPP. We have shown that if appropriate tests (which are robust to cross-sectional dependence and more powerful) are used, any evidence of erratic behaviour disappears, and strong empirical support is found for PPP. This suggests that power is the critical issue in testing PPP, rather than considering more complicated dynamic structures. Although

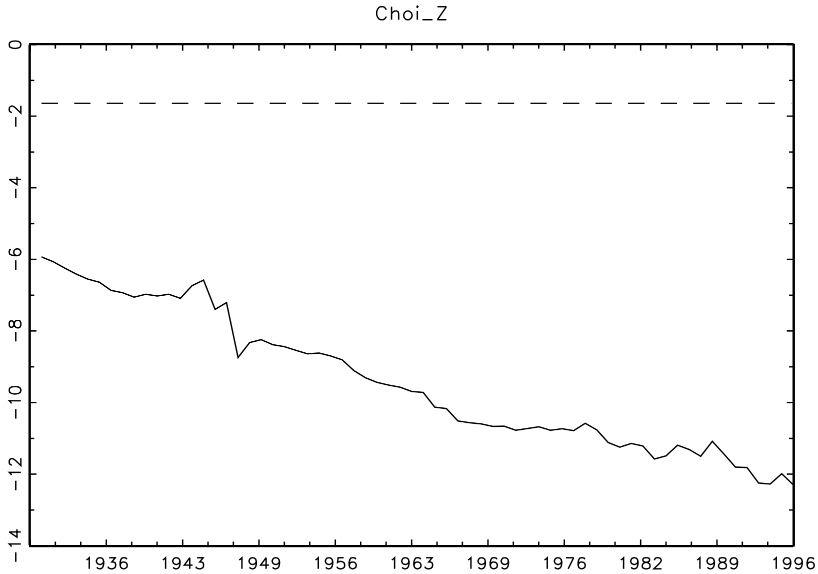


FIGURE 5.3: TEST STATISTIC SERIES FOR Z FOR VARIOUS N

nonlinear modelling also seems to be a very promising direction for future research on real exchange rates (see, e.g., Taylor and Peel, 2000), addressing the power problem is confirmed here to be of crucial importance, and panel approaches may to be able to provide conclusive evidence of the adequacy of PPP as a theory of real exchange rate determination, provided sufficiently long runs of data are used and cross-sectional dependence is tackled appropriately.

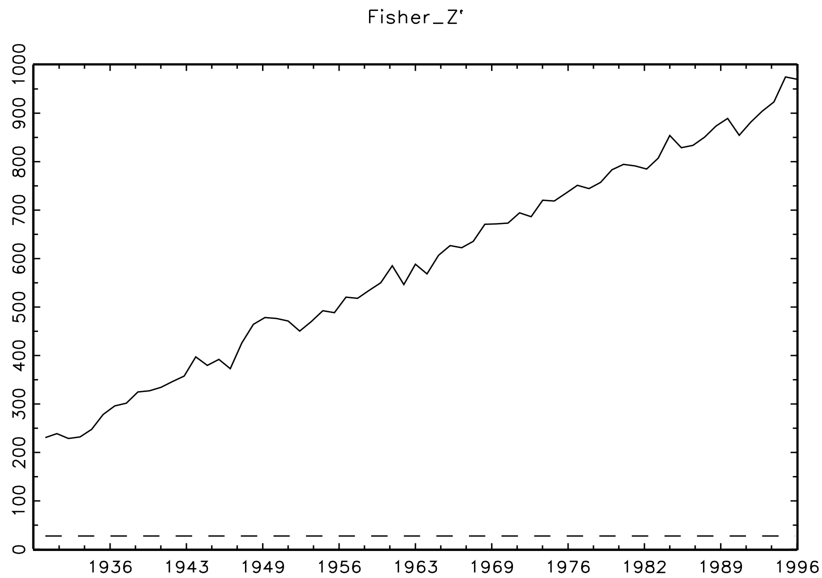


FIGURE 5.4: TEST STATISTIC SERIES FOR P FOR VARIOUS N

Chapter 6

The Error-in-Rejection Probability of Meta-Analytic Panel Tests

Abstract

Meta-analytic panel unit root tests such as Fisher's χ^2 test, which consist of pooling the p-values of time series unit root tests, are widely applied in practice. Recently, several Monte Carlo studies have found these tests' Error-in-Rejection Probabilities (or, synonymously, size distortion) to increase with the number of series in the panel. We investigate this puzzling finding by modelling the finite sample p-value distribution of the time series tests with local deviations from the asymptotic p-value distribution. We find that the size distortions of the panel tests can be explained as the cumulative effect of small size distortions in the time series tests.[1]

Keywords: Panel Unit Root Tests, Meta-Analysis, Error-in-Rejection Probability

[1]I would like to thank Sebastian Herr for very helpful discussion.

Meta-analysis is a useful tool to efficiently combine related information.[2] In recent years, the meta-analytic testing approach has been fruitfully applied to nonstationary panels: Consider the testing problem on the panel as consisting of N testing problems for each unit of the panel. That is, conduct N separate time series tests and obtain the corresponding p-values of the test statistics. Then, combine the p-values of the N tests into a single panel test statistic. Among others, Maddala and Wu (1999), Choi (2001) and Phillips and Sul (2003) propose meta-analytic panel unit root and cointegration tests. The tests are intuitive, relatively easy to compute and allow for a considerable amount of heterogeneity in the panel.

Via Monte Carlo experiments, the above-cited authors show that their meta-analytic tests can be substantially more powerful than separate time series tests on each unit in the panel, justifying the use of panel tests. Disturbingly, however, Choi (2001), Hlouskova and Wagner (2006) and Chapter 3, inter alia, find the Error-in-Rejection Probability (ERP) (or, synonymously, size distortion) to be increasing in N. That is, the (absolute) difference between the estimated rejection probability (or type I error rate) $R(\alpha, N)$ and the nominal significance level α, $ERP_N(\alpha) := |R(\alpha, N) - \alpha|$, gets larger with N. A priori, this finding is counterintuitive, since more information should improve the performance of the panel tests.

We argue that this behavior may be explained as the cumulative effect of arbitrarily small ERPs in the underlying time series test statistics composing the panel test statistics. Under a simple H_0, assuming continuous distribution functions of the test statistics, p-values of test statistics should be distributed uniformly on the unit interval, denoted $\mathcal{U}[0, 1]$ (see, e.g., Bickel and Doksum, 2001, Sec. 4.1). We model size-distorted time series tests by deviations from the null distribution of the test statistics' p-values. The analytical and simulation

[2]See Hedges and Olkin (1985) for a nice introduction to the topic.

evidence reported in the following sections corroborate our conjecture.

6.1 The P-Value Combination Test

We briefly review the p-value combination test whose ERP is investigated subsequently.[3] We discuss the example of a panel unit root test. The conclusions might however be valid also for other applications of the meta test. Denote by p_i the marginal significance level, or p-value, of a time series unit root test applied to the ith unit of the panel. Let θ_{i,T_i} be a unit root test statistic on unit i for a sample size of T_i . Let F_{T_i} denote the null distribution function of the test θ_{i,T_i}. Since the tests considered here are one-sided, $p_i = F_{T_i}(\theta_{i,T_i})$ if the test rejects for small values of θ_{i,T_i} and $p_i = 1 - F_{T_i}(\theta_{i,T_i})$ if the test rejects for large values of θ_{i,T_i}. We only consider time series tests with the null of a unit root.

We test the following null hypothesis:

$$H_0: \text{ The time series } i \text{ is unit-root nonstationary} \qquad (i \in \mathbb{N}_N), \tag{6.1}$$

against the alternative

$$H_1: \text{ For at least one } i\text{, the time series is stationary.}$$

($(i \in \mathbb{N}_N)$ is shorthand for $i = 1, \ldots, N$.) The N p-values of the individual time series tests, p_i $(i \in \mathbb{N}_N)$, are combined as follows to obtain a test statistic for panel (non-) stationarity:

$$P_{\chi^2} = -2 \sum_{i=1}^{N} \ln p_i \tag{6.2}$$

The P_{χ^2} test, via pooling p-values, provides a convenient test for panel (non-) stationarity by imposing minimal homogeneity restrictions on the panel. For

[3] Similar results for other widely used meta-analytic tests such as the inverse normal test are available upon request.

instance, the panel can be unbalanced. For further discussion see Choi (2001) or Chapter 3. The following lemma gives the asymptotic distribution of the test.

LEMMA 1 *(Distribution of the P_{χ^2} test).*
Under the null of panel nonstationarity and assuming continuous distribution functions of the θ_{i,T_i}, the P_{χ^2} test is, as $T_i \to \infty$ $(i \in \mathbb{N}_N)$, asymptotically distributed as

$$P_{\chi^2} \to_d \chi^2_{2N}$$

PROOF. See page 49. □

The decision rule is to reject the null of panel nonstationarity when P_{χ^2} exceeds the critical value from a χ^2_{2N} distribution at the desired significance level. The test has a well-defined asymptotic distribution (for $T \to \infty$) for any finite N. This feature is attractive because in many applications, the assumption that N, the number of units in the panel, goes to infinity may not be a natural one.

6.2 The Error-in-Rejection Probability of the Combination Test

As should be clear from the previous discussion, any unit root for which p-values are available can be used to compute the P_{χ^2} test statistic. Popular choices include the Augmented Dickey-Fuller test (Dickey and Fuller, 1979). It is well-known that using the (first-order) asymptotic approximation F, a functional of Brownian Motions and possibly nuisance parameters, to the exact, finite T_i null distribution of the test statistics, F_{T_i}, need not be accurate. This is because the null hypothesis (6.1) is not a simple one (and the available test statistics are not pivotal). H_0 is satisfied by all unit-root nonstationary processes

$$y_{i,t} = y_{i,t-1} + u_{i,t}, \qquad (i \in \mathbb{N}_N)$$

TABLE 6.1: SIMULATED TYPE I ERROR RATES FOR THE P_{χ^2} TEST.

N	5	10	25	30	50	60	100
Maddala and Wu (1999)			.044		.107		.131
Choi (2001)	.050	.070	.090		.090		.130
Chapter 3		.035	.031		.021		.014
Hlouskova and Wagner (2006)		.090	.110		.120		.145
Choi [forthcoming]				.051		.042	.037

Note: All results are for the nominal 5% level.

where the errors $u_{i,t}$ can be from a wide class of dependent and heterogeneous sequences. See, for instance, the fairly general strong mixing conditions on $u_{i,t}$ of Phillips (1987). Hence, the p-values of the test need no longer be uniformly distributed on the unit interval, even if the true Data Generating Process (DGP) of the time series is from the null hypothesis set of unit-root nonstationary processes. Thus, the assumptions required for validity of Lemma 1 need no longer be met.

As we argue in this section, this fact can explain the counterintuitive finding of a deteriorating performance of the P_{χ^2} test with increasing N. Table 6.1 summarizes selected Monte Carlo results on the ERP of the P_{χ^2} test reported in the literature.[4] Most authors find $R(\alpha, N) - \alpha$ to increase with N, while Chapter 3 and Choi report an inverse relationship. All find $ERP_N(\alpha) = |R(\alpha, N) - \alpha|$ to increase with N.

We propose the following modelling assumption to investigate this behavior.

ASSUMPTION 5 *(Generalized p-value distribution).*
For finite T_i, the p-values are distributed as $\tilde{p}_i \sim \mathcal{U}[a, b]$, where $a \geqslant 0$, $b \leqslant 1$ and $a < b$, $(i \in \mathbb{N}_N)$.

[4] The differences stem from the length of the underlying time series, the type of nonstationarity test applied to the time series, as well as the design of the DGP.

Since the exact, finite T_i distribution of the test statistics is generally unknown, so is the exact p-values' distribution. The assumption is, however, convenient for modelling purposes. First, letting $a \to 0$ and $b \to 1$, it comprises the asymptotic result as a limiting case. Second, it is easy to characterize the ERP of a single time series test in terms of a and b. More precisely, since a rejection at level α is equivalent to a p-value $p < \alpha$,

$$\mathsf{P}(F_{\tilde{p}_i} < \alpha) = R(\alpha, 1) = \begin{cases} 0 & \text{for} \quad a > \alpha \\ \frac{\alpha - a}{b - a} & \text{for} \quad a < \alpha \text{ and } b > \alpha \\ 1 & \text{for} \quad b < \alpha \end{cases}$$

In particular, it is possible to model "oversized" unit root tests by taking $\tilde{p}_i \sim \mathcal{U}[0, b]$, where $b < 1$. Intuitively, we remove the p-values corresponding to the test statistics speaking most strongly in favor of H_0. Conversely, $\tilde{p}_i \sim \mathcal{U}[a, 1]$, $a > 0$ represents an "undersized" test. The following lemma derives the density function of $-2 \ln \tilde{p}_i$ under Assumption 1.

LEMMA 2 *(Distribution of $-2 \ln \tilde{p}_i$).*
Under $\tilde{p}_i \sim \mathcal{U}[a, b]$, the density of $-2 \ln \tilde{p}_i$ is given by

$$f_{-2 \ln \tilde{p}_i}(y) = \begin{cases} 0 & \text{for} \quad y \in (-\infty, -2 \ln b) \\ \frac{1}{2(b-a)} e^{-\frac{y}{2}} & \text{for} \quad y \in [-2 \ln b, -2 \ln a] \\ 0 & \text{for} \quad y \in (-2 \ln a, \infty), \end{cases}$$

taking $-\ln a = \infty$ for $a = 0$.

PROOF. Again, we can apply the transformation theorem for absolutely continuous r.v.s. Using the notation from the proof of Lemma 1, we still have $\tilde{p}_i = g^{-1}(y) = e^{-\frac{1}{2}y}$ and hence $|g^{-1'}(y)| = \frac{1}{2} e^{-\frac{y}{2}}$. $f_{\tilde{p}_i}$ follows immediately from Assumption 1 as $f_{\tilde{p}_i}(g^{-1}(y)) = \frac{1}{b-a}$ for $g^{-1}(y) \in [a, b]$ and 0 otherwise. The support of the r.v. $-2 \ln \tilde{p}_i$ follows from solving g^{-1} for the lower and upper bounds of $\tilde{p}_i$. It is verified elementarily that $f_{-2 \ln \tilde{p}_i}(y)$ satisfies $\int_{\mathbb{R}} f_{-2 \ln \tilde{p}_i}(\tilde{y}) d\tilde{y} = 1$. □

$f_{-2 \ln \tilde{p}_i}(y)$ contains the density of the χ^2_2 distribution as a special case with $a = 0$ and $b = 1$. We now study the ERP of the P_{χ^2} test for the case $N = 1$, denoted $ERP_1(\alpha)$. Let c_{α_2} be the critical value of the χ^2_2-distribution at nominal level

α, i.e. $\int_0^{c_{\alpha_2}} \frac{1}{2} e^{-\frac{1}{2}\tilde{y}} d\tilde{y} = 1 - \alpha \Rightarrow c_{\alpha_2} = -2\ln\alpha$. Then,

$$\begin{aligned} R(\alpha, 1) &= 1 - \int_{-\infty}^{-2\ln\alpha} f_{-2\ln\tilde{p}_i}(\tilde{y}) d\tilde{y} \\ &= \frac{\alpha}{b} \end{aligned}$$

Let us consider a specific example. We investigate the "oversized" case, $\tilde{p}_i \sim \mathcal{U}[0, 0.9]$, and $\alpha = 0.05$. Then, $ERP_1(0.05) = \left|\frac{0.05(1-0.9)}{0.9}\right| = \frac{0.05(1-0.9)}{0.9} \approx 0.005$, yielding an ERP which would be considered small in most Monte Carlo analyses.

For $N = 2$, we derive the following lemma in the appendix:

Lemma 3 *(Density function of $f_{-2\sum_{i=1}^{2}\ln\tilde{p}_i}(y)$).*

$$f_{-2\sum_{i=1}^{2}\ln\tilde{p}_i}(y) = \begin{cases} 0 & \text{for} \quad y \in (-\infty, -2\ln b) \\ \frac{1}{4(b-a)^2} e^{-\frac{y}{2}}(y + 4\ln b) & \text{for} \quad y \in [-4\ln b, -2\ln a - 2\ln b] \\ \frac{1}{4(b-a)^2} e^{-\frac{y}{2}}(-y - 4\ln a) & \text{for} \quad y \in (-2\ln a - 2\ln b, -4\ln a] \\ 0 & \text{for} \quad y \in (-4\ln a, \infty), \end{cases}$$

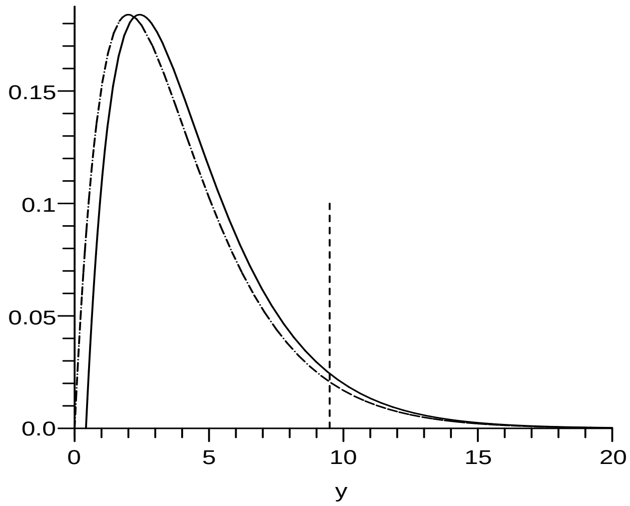
taking $-\ln a = \infty$ for $a = 0$.

Continuing the above example, we compute $ERP_2(0.05)$ as

$$\begin{aligned} ERP_2(0.05) &= |R(0.05, 2) - 0.05| = R(0.05, 2) - 0.05 \\ &= 1 - \int_{-\infty}^{c_{\alpha_4}} f_{-2\sum_{i=1}^{2}\ln\tilde{p}_i}(\tilde{y}) d\tilde{y} - 0.05 \\ &\approx 0.009 \end{aligned}$$

Note that $ERP_2(0.05) > ERP_1(0.05)$. For illustration, Figure 6.1 displays $f_{-2\sum_{i=1}^{2}\ln\tilde{p}_i}(y)$ and the density function of the χ_4^2 distribution. The generalized p-value distribution lies to the right of the χ_4^2 distribution, as expected. The dashed line indicates the 0.95 quantile of the χ_4^2 distribution. $f_{-2\sum_{i=1}^{2}\ln\tilde{p}_i}(y)$ has probability mass of more than 0.05 to the right of c_{α_4}.

To analyze the ERP of the P_{χ^2} test for general N, we require the cumulative distribution function of the r.v. $-2\sum_{i=1}^{N}\ln\tilde{p}_i$ under Assumption 1. In keeping



DOTTED: χ^2_4-DENSITY, SOLID: $f_{-2\sum_{i=1}^{2}\ln\tilde{p}_i}(y)$, DASHED: $\chi^2_{4,.95}$ CRITICAL VALUE

FIGURE 6.1: THE DENSITY FUNCTIONS FOR THE CASE $N = 2$

with most applications in the literature, we assume independence across i. It is then possible to write the density of $-2\sum_{i=1}^{N}\ln\tilde{p}_i$ as the convolution of $f_{-2\ln\tilde{p}_i}$ $(i \in \mathbb{N}_N)$ (see, e.g., Shiryaev, 1996, pp. 241)

$$\begin{aligned} f_{-2\sum_{i=1}^{N}\ln\tilde{p}_i}(y) &= f_{-2\ln\tilde{p}_1} * \cdots * f_{-2\ln\tilde{p}_N}(y) \\ &= \frac{e^{-\frac{y}{2}}}{2^N(b-a)^N}\varphi_{-2\ln b,-2\ln a} * \cdots * \varphi_{-2\ln b,-2\ln a}, \end{aligned}$$

where φ is the indicator function of y on the interval $I = [-2\ln b, -2\ln a]$. Introducing a suitable standardization factor r_N, the convolution for general N can be written as a function of the indicator functions of y on the unit interval,

$$f_{-2\sum_{i=1}^{N}\ln\tilde{p}_i}(y) = r_N\frac{e^{-\frac{y}{2}}}{2^N(b-a)^N}\varphi_{0,1} * \cdots * \varphi_{0,1}.$$

By a Central Limit Theorem, the sum of N centered and standardized uniform r.v.s converges to a standard normal r.v. Using a further scaling constant s_N, we expect that the density of $f_{-2\sum_{i=1}^{N}\ln\tilde{p}_i}(y)$ can be well approximated for N sufficiently large by an expression of the form $r_N s_N\frac{e^{-\frac{y}{2}}}{2^N(b-a)^N}\phi_N(y)$. Here, $\phi_N(y)$

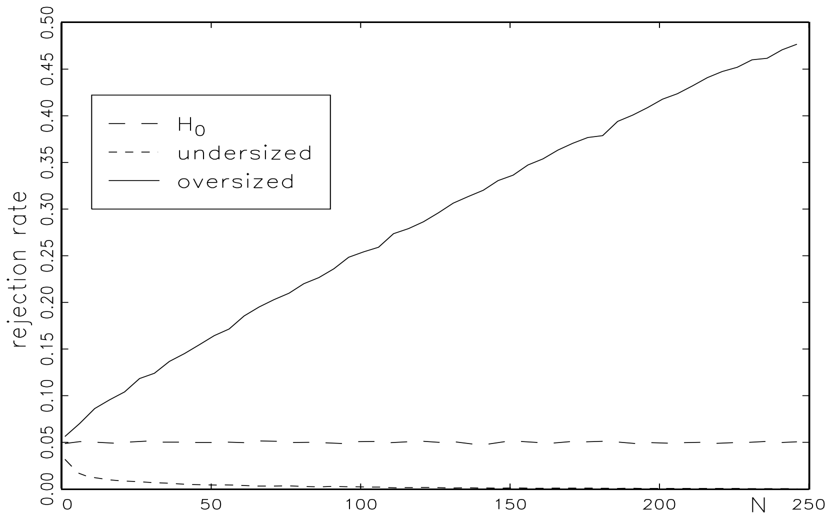


FIGURE 6.2: REJECTION RATES OF P_{χ^2} TEST AT 5% AS A FUNCTION OF N

is the density function of the standard normal distribution (whose argument also depends on N).

The exact computation however quickly becomes cumbersome for large N. We shall therefore rely on simulation to further illustrate the effect of increasing N. For each $N \in \{1, 6, 11, \ldots, 246\}$ we generate p-values according to $\mathcal{U}[0.02, 1]$, $\mathcal{U}[0, 1]$ and $\mathcal{U}[0, 0.9]$, corresponding to under-, correctly, and oversized tests. Based on $R = 50,000$ replications, Figure 6.2 displays the rejection rates of the P_{χ^2} test as a function of N when using the 5% critical values of the appropriate χ^2_{2N} distribution.

The figure confirms the conjectures resulting from the theoretical analysis. When the p-values are, as they should be under H_0, distributed uniformly on the unit interval, the ERP of the P_{χ^2} test is excellent uniformly in N. Conversely, even for small deviations from the nominal size of the time series tests, the ERPs clearly increase in N. Reassuringly, the simulation result for $N = 1$ is virtually indistinguishable from the analytical result above. Similarly, Figure 6.3 reveals that the ERP of the panel test is the higher the stronger the component test statistics are oversized, as one would expect.

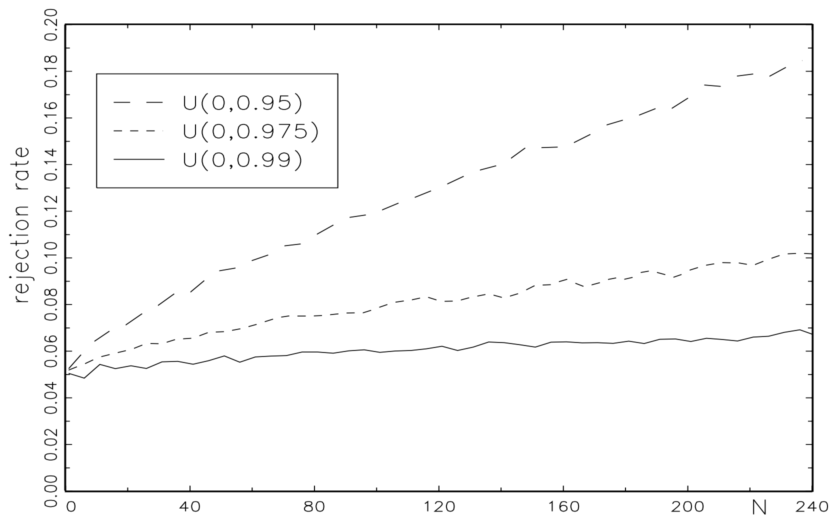


Figure 6.3: Rejection Rates of P_{χ^2} test at 5% for differing degrees of "oversizedness"

6.3 Conclusion

We show that meta-analytic panel tests can have arbitrarily large Errors-in-Rejection Probabilities (size distortions) even when the underlying time series tests have only slight Errors-in-Rejection Probabilities. The recommendation for empirical practice therefore is to use critical values which take into account as well as possible the shape of the exact (but generally unknown) finite sample distribution of the test statistics. One way to achieve this is to compute correction factors depending on T using response surface regressions (MacKinnon, 1991). Even though we discuss the application of the P_{χ^2} test to testing problems for nonstationary panel data, the conclusions may hold more generally for other applications and other meta-analytic test statistics.

Appendix

Proof of Lemma 3

The convolution integral is given by

$$\begin{aligned} f_{-2\sum_{i=1}^{2}\ln\tilde{p}_i}(y) &= f_{-2\ln\tilde{p}_i} * f_{-2\ln\tilde{p}_i} \\ &= \int_{\mathbb{R}} \frac{1}{4(b-a)^2} e^{-\frac{1}{2}x}\varphi_{-2\ln b,-2\ln a} e^{-\frac{1}{2}(y-x)}\varphi_{-2\ln b,-2\ln a} dx \\ &= \frac{1}{4(b-a)^2} e^{-\frac{1}{2}y} \int_{\mathbb{R}} \varphi_{-2\ln b,-2\ln a}\varphi_{-2\ln b,-2\ln a} dx \qquad (6.A.1) \end{aligned}$$

Since we consider the convolution of two densities with support $I = [-2\ln b, -2\ln a]$, the arguments have to satisfy $y-x, x \in [-2\ln b, -2\ln a]$, implying the following weak inequalities: $x \leqslant y + 2\ln b$, $x \geqslant y + 2\ln a$, $y \leqslant -4\ln a$ and $y \geqslant -4\ln b$. Together, these require that

$$\begin{aligned} x &\leqslant \min\{y + 2\ln b, -2\ln a\} =: M(y) \text{ and} \\ x &\geqslant \max\{-2\ln b, y + 2\ln a\} =: m(y) \end{aligned}$$

That is, we distinguish the following cases in (6.A.1):

1. For $y \in (-\infty, -4\ln b)$, we have $f_{-2\sum_{i=1}^{2}\ln\tilde{p}_i}(y) = 0$.
2. We have $-2\ln b \geqslant y + 2\ln a$ for $y \in [-4\ln b, -2\ln a - 2\ln b]$ and hence $m(y) = -2\ln b$ and $M(y) = y + 2\ln b$. Thus,

$$\begin{aligned} f_{-2\sum_{i=1}^{2}\ln\tilde{p}_i}(y) &= \frac{1}{4(b-a)^2} e^{-\frac{1}{2}y} \int_{-2\ln b}^{y+2\ln b} dx \\ &= \frac{1}{4(b-a)^2} e^{-\frac{y}{2}}(y + 4\ln b). \end{aligned}$$

3. We have $-2\ln b \leqslant y + 2\ln a$ for $y \in (-2\ln a - 2\ln b, -4\ln a)$ and hence $m(y) = y + 2\ln a$ and $M(y) = -2\ln a$. Thus,

$$\begin{aligned} f_{-2\sum_{i=1}^{2}\ln\tilde{p}_i}(y) &= \frac{1}{4(b-a)^2} e^{-\frac{1}{2}y} \int_{y+2\ln a}^{-2\ln a} dx \\ &= \frac{1}{4(b-a)^2} e^{-\frac{y}{2}}(-y - 4\ln a). \end{aligned}$$

4. For $y \in (-4\ln a, \infty)$, we have $f_{-2\sum_{i=1}^{2}\ln\tilde{p}_i}(y) = 0$. $\square$

Chapter 7

Mixed Signals Among Panel Cointegration Tests

Abstract

Time series cointegration tests, even in the presence of large sample sizes, often yield conflicting conclusions ("mixed signals") as measured by, inter alia, a low correlation of empirical p-values (see Gregory, Haug, and Lomuto, 2004, *Journal of Applied Econometrics*). Using their methodology, we present evidence suggesting that the problem of mixed signals persists for popular panel cointegration tests. As expected, there is weaker correlation between residual and system-based tests than between tests of the same group.

Keywords: Panel cointegration tests, Monte Carlo comparison

7.1 Introduction

An extensive battery of tests is available to investigate the unit root and cointegration properties of economic time series. Typically, however, an applied researcher has little practical guidance as to which test to use, as most tests test very similar hypotheses. It would therefore be reassuring if rejection or acceptance of a particular economic hypothesis did not depend on which of the tests is used. For instance, in the context of hypothesis testing with stationary variables it is well-known that the classical likelihood ratio, Lagrange multiplier and Wald tests are asymptotically numerically equivalent under quite general conditions (Davidson and MacKinnon, 1993, Ch. 13).

As analytical characterizations of the correlations of the various test statistics for cointegration are difficult to obtain, Gregory, Haug, and Lomuto (2004) analyze this question by means of Monte Carlo methods. They generate replications of two independent random walks and test the null of no cointegration using the popular residual-based tests by Engle and Granger (1987) and Phillips and Ouliaris (1990) as well as the system-based λ_{trace} and λ_{max} tests (Johansen, 1988). They then calculate p-values from the empirical distribution of the test statistics by taking rank order of the latter and dividing by M, the number of replications. Disturbingly, for most pairs of tests, virtually any combination of p-values can arise. That is, while the combinations should ideally cluster around the 45°-line, it frequently occurs that a particularly strong rejection using, say, the λ_{trace}-test is associated with a clear acceptance of the null using, say, the Engle and Granger (1987) Augmented Dickey-Fuller (ADF)-test. The main conclusion is that using different tests is likely to yield conflicting conclusions in applications.

In recent years, the cointegration methodology has been extended to panel data. Pedroni (2004) and Kao (1999) generalize residual-based tests, Larsson, Lyhagen, and Löthgren (2001) extend the Johansen (1988) tests, while McCoskey and Kao (1998) propose a test for the null of panel cointegration in the spirit of Shin (1994). Chapter 3 extends the p-value combination panel unit root tests of Maddala and Wu (1999) and Choi (2001) to the panel cointegration case.

Under cross-sectional independence all the above-mentioned panel tests provide a means to better exploit the variation in the data. Furthermore, Phillips and Moon (1999) show that panel data can help mitigate the spurious regression phenomenon. The contribution of this chapter is to investigate whether the availability of panel data is also useful for obtaining more consistent decisions among the competing tests. To shed light on this question we adopt the methodology suggested by Gregory, Haug, and Lomuto (2004) and extend it to the panel data setting.

The remainder of the chapter is organized as follows. Section 2 briefly reviews the panel cointegration tests compared in this chapter. Section 3 describes the simulation setup of the comparative study and reports the results. Section 4 concludes.

7.2 Panel Cointegration Tests

We give the key statistics of the various tests that are considered. For more details, refer to the original contributions. Furthermore, Banerjee (1999) or Baltagi and Kao (2000) provide surveys of the literature. We focus on tests with the null of no panel cointegration.

Pedroni (2004)

Pedroni (2004) derives seven different tests for panel cointegration. These may be categorized according to what information on the different units of the panel is pooled. The "Group-Mean" Statistics are essentially means of the conventional time series tests (see Phillips and Ouliaris, 1990). The "Within" Statistics separately sum the numerator and denominator terms of the corresponding time series statistics. Let $A_i = \sum_{t=1}^{T} \tilde{e}_{i,t}\tilde{e}'_{i,t}$, where $\tilde{e}_{i,t} = (\Delta\hat{e}_{i,t}, \hat{e}_{i,t-1})'$ and T is sample size. The $\hat{e}_{i,t}$ are obtained from heterogenous Engle/Granger-type first stage OLS multivariate time series regressions of one of the variables $\mathbf{x}_{ik}$ on the remaining $\mathbf{x}_{i,-k}$, possibly including some deterministic regressors. We consider the "Group-ρ", "Panel-ρ" and (nonparametric) "Panel-t"-test statistics which are given by, respectively,

$$\begin{aligned}
\tilde{Z}_{\hat{\rho}_{NT}-1} &= \sum_{i=1}^{N} A_{22i}^{-1}(A_{21i} - T\hat{\lambda}_i), \\
Z_{\hat{\rho}_{NT}-1} &= \left(\sum_{i=1}^{N} A_{22i}\right)^{-1} \sum_{i=1}^{N}(A_{21i} - T\hat{\lambda}_i) \quad \text{and} \\
Z_{\hat{t}_{NT}} &= \left(\tilde{\sigma}_{NT}^2 \sum_{i=1}^{N} A_{22i}\right)^{-1/2} \sum_{i=1}^{N}(A_{21i} - T\hat{\lambda}_i).
\end{aligned}$$

The expressions $\hat{\lambda}_i$ and $\tilde{\sigma}_{NT}^2$ estimate nuisance parameters from the long-run conditional variances. After proper standardization, all statistics have a standard normal limiting distribution. The decision rule is to reject the null hypothesis of no panel cointegration for large negative values.

Kao (1999)

Kao (1999) proposes five different panel extensions of the time series $(A)DF$-type tests. We focus on those that do not require strict exogeneity of the regressors. More specifically,

$$
\begin{aligned}
DF^*_\rho &= \frac{\sqrt{N}T(\hat\rho - 1) + \dfrac{3\sqrt{N}\hat\sigma^2_\nu}{\hat\sigma^2_{0\nu}}}{\sqrt{3 + \dfrac{36\hat\sigma^4_\nu}{5\hat\sigma^4_{0\nu}}}} \quad \text{and} \\
DF^*_t &= \frac{t_\rho + \dfrac{\sqrt{6N\hat\sigma^2_\nu}}{2\hat\sigma_{0\nu}}}{\sqrt{\dfrac{\hat\sigma^2_{0\nu}}{2\hat\sigma^2_\nu} + \dfrac{3\hat\sigma^2_\nu}{10\hat\sigma^2_\nu}}}.
\end{aligned}
$$

Here, $\hat\rho$ is the estimate of the AR(1) coefficient of the residuals from a fixed effects panel regression and t_ρ is the associated t-statistic. The remaining terms play a role similar to the nuisance parameter estimates in the Pedroni (2004) tests. Again, both tests are standard normal under the null of no panel cointegration and reject for large negative values.

Larsson, Lyhagen, and Löthgren (2001)

The panel cointegration test of Larsson, Lyhagen, and Löthgren (2001) applies a Central Limit Theorem to the set of N λ_{trace} test statistics (Johansen, 1988) for each unit in the panel. (See also (7.2) below.) Defining $\overline{\lambda}_{trace} = N^{-1}\sum_{i=1}^N \lambda_{trace,i}$, their panel cointegration test statistic is given by

$$
\Upsilon_{LR} = \sqrt{N}\left(\frac{\overline{\lambda}_{trace} - \mathsf{E}[\overline{\lambda}_{trace}]}{\sqrt{\mathsf{Var}[\overline{\lambda}_{trace}]}}\right).
$$

Under some conditions, including $\sqrt{N}T^{-1} \to 0$, Larsson, Lyhagen, and Löthgren (2001) can show that $\Upsilon_{LR} \xrightarrow{T,N} \mathcal{N}(0,1)$. The moments are obtained by stochastic simulation and are tabulated in the paper. The null hypothesis of no cointegration at a level α is rejected if the test statistic exceeds the $(1-\alpha)$-quantile of the standard normal distribution, i.e. for large values.

The Tests from Chapter 3

The main idea of the testing principle has been used in meta analytic studies for a long time (cf. Fisher, 1970; Hedges and Olkin, 1985). Consider the testing problem on the panel as consisting of N testing problems for each unit of the panel. That is, conduct N separate time series cointegration tests and obtain the corresponding p-values of the test statistics. The test statistics are obtained by combining the p-values of the N tests into panel test statistics as follows:

$$P_{\chi^2} = -2\sum_{i=1}^{N} \ln(p_i) \tag{7.1a}$$

$$P_{\Phi^{-1}} = N^{-\frac{1}{2}}\sum_{i=1}^{N} \Phi^{-1}(p_i), \tag{7.1b}$$

where Φ^{-1} denotes the inverse of the cumulative distribution function (cdf) of the standard normal distribution. When considered together we refer to Eqs. (7.1a) and (7.1b) as P tests from now on. Assuming continuous distribution functions of the time series test statistics under H_0, as $T_i \to \infty$ for all i, the test statistics are asymptotically distributed as

$$P_{\chi^2} \to_d \chi^2_{2N}$$
$$P_{\Phi^{-1}} \to_d \mathcal{N}(0,1),$$

where χ^2_{2N} is a χ^2 random variable with $2N$ degrees of freedom. The decision rule is to reject the null of no panel cointegration when P_{χ^2} exceeds the critical value from a χ^2_{2N} distribution at the desired significance level. On the other hand, for (7.1b) one would reject for large negative values of $P_{\Phi^{-1}}$.

We obtain the p-values from the ADF cointegration tests (Engle and Granger, 1987) as provided by MacKinnon (1996). That is, the p-values are from the t-statistic of $\gamma_i - 1$ in the OLS regression

$$\Delta\hat{u}_{i,t} = (\gamma_i - 1)\hat{u}_{i,t-1} + \sum_{p=1}^{P} \nu_p \Delta\hat{u}_{i,t-p} + \epsilon_{i,t}.$$

Here, $\hat{u}_{i,t}$ is the usual residual from a first stage multivariate OLS time series regressions of one of the variables $\mathbf{x}_{ik}$ on the remaining $\mathbf{x}_{i,-k}$. Alternatively, one could capture serial correlation by the semiparametric approach of Phillips and Ouliaris (1990). Finally, we obtain the p-values for the Johansen (1988) λ_{trace} and λ_{max} tests provided in MacKinnon, Haug, and Michelis (1999). That is, we test for the presence of h cointegrating relationships by estimating the number of significantly non-zero eigenvalues of the matrix $\hat{\Pi}_i$ estimated from the Vector Error Correction Model

$$\Delta \mathbf{x}_{i,t} = -\Pi_i \mathbf{x}_{i,t-P} + \sum_{p=1}^{P-1} \Gamma_{i,p} \Delta \mathbf{x}_{i,t-p} + \boldsymbol{\epsilon}_{i,t}$$

by the λ_{trace}-test

$$\lambda_{trace,i}\left(h\right) = -T \sum_{k=h+1}^{K} \ln\left(1 - \hat{\pi}_{k,i}\right) \tag{7.2}$$

and the λ_{max}-test

$$\lambda_{max,i}\left(h|h+1\right) = -T \ln\left(1 - \hat{\pi}_{h+1,i}\right). \tag{7.3}$$

Here, $\hat{\pi}_{k,i}$ denotes the kth largest eigenvalue of $\hat{\Pi}_i$. In (7.2), the alternative is a general one, while one tests against $h+1$ cointegration relationships in (7.3).

7.3 Do Panel Cointegration Tests Produce "Mixed Signals"?

We now use the panel cointegration tests outlined in the previous section to investigate the extent to which different widely used panel cointegration tests yield the same decision for a given (artificial) sample. Gregory, Haug, and Lomuto (2004) observe mixed signals, i.e. a relatively high test statistic for one test and a relatively low test statistic for another, for time series cointegration tests.[1] This effect is particularly strong when comparing residual- and system-based tests.

It might be conjectured that the availability of panel data, leading to standard (often, standard normal) null distributions of the test statistics, could help alleviate this problem. To shed light on this question, we adopt the methodology of Gregory, Haug, and Lomuto (2004).[2] More precisely, we generate many, say M, replications of two integrated time series for each of the N units in the panel. For each replication, we store the different panel cointegration test statistics. The extent to which the different tests yield identical decisions is measured by two related criteria. First, we compute empirical p-values of the tests by taking rank order of the test statistics and dividing by M. We then compute the correlation of the empirical p-values for each pair of tests. If both have the same null and the same alternative, the correlation should therefore ideally be close to one, i.e. a strong rejection of one test should also be a strong one of the other. Second, we record all the instances of each pair of tests rejecting jointly. The critical values are either taken from the asymptotic distribution of the tests or the empirical distribution arising from the replications under the null. Thus, when testing a sample generated under the null at the 5% level, all pairs of tests should ideally jointly reject in close to 5% of the replications.

We compare the tests of Kao (1999), Pedroni (2004) and Larsson, Lyhagen, and Löthgren (2001) presented in the previous section. We further include the two P tests. For each, we use both Engle and Granger's (1987) ADF test with one

[1]Berndt and Savin (1977) study the related problem of conflicting decisions among the classical hypothesis tests in linear regression models. A crucial difference is that the numerical relationship between the criteria is well understood for these simpler models. Furthermore, in this context the situation is resolved asymptotically.

[2]Gregory, Haug, and Lomuto (2004) complement their simulation study with an extensive analysis of all applications of the cointegration methodology published in the *Journal of Applied Econometrics* in recent years. While such an approach has obvious appeal it is not yet promising in the panel data context due to the small number of empirical applications. We therefore exclusively rely on artificial data.

lagged difference ($P_{\chi^2 DF}$ and $P_{\Phi^{-1}DF}$) as well as Johansen's (1988) λ_{trace} test for $h = 0$ versus $h \leqslant 2$ cointegrating relationships ($P_{\chi^2 J}$ and $P_{\Phi^{-1}J}$). Following Gregory, Haug, and Lomuto (2004), we choose relatively large time series dimensions to limit size distortions. More specifically, $T \in \{250, 500, 1000, 2000\}$ and $N \in \{10, 20, 50, 100, 150\}$. The Data Generating Process (DGP) is similar to the one used by Engle and Granger (1987). The extension to the panel data setting is discussed in Kao (1999). For simplicity we only consider the bivariate case:

DGP

$$x_{i,1t} - \alpha_i - \beta x_{i,2t} = z_{i,t}, \qquad a_1 x_{i,1t} - a_2 x_{i,2t} = w_{i,t}$$

where

$$z_{i,t} = \rho z_{i,t-1} + e_{zi,t}, \qquad \Delta w_{i,t} = e_{wi,t}$$

and

$$\begin{pmatrix} e_{zi,t} \\ e_{wi,t} \end{pmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \psi\sigma \\ \psi\sigma & \sigma^2 \end{bmatrix} \right)$$

Remarks

- When $|\rho| < 1$ the equilibrium error in the first equation is stationary such that x_{i1t} and x_{i2t} are cointegrated with $\boldsymbol{\beta}_i = (1 \quad -\alpha_i \quad -\beta)'$.
- When writing the above DGP as an error correction model (see, e.g., Gonzalo, 1994) it is immediate that x_{i2t} is weakly exogenous when $a_1 = 0$.

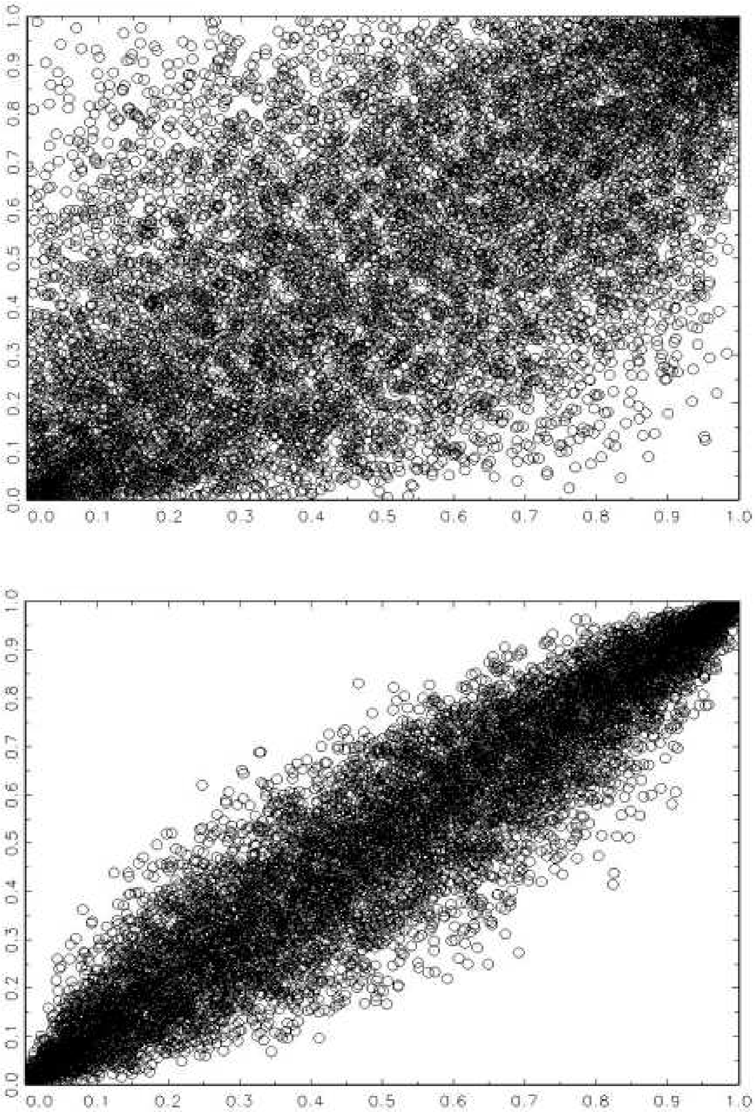
We consider the parameter space $\beta = 2$, $a_2 = -1$, $\sigma \in \{0.5, 1\}$, $\psi \in \{-0.5, 0, 0.5\}$ and $a_1 \in \{0, 1\}$. This implies that, for instance, the Pedroni (2004) and P tests cannot exhibit their comparative advantage of being able to detect cross-sectional heterogeneity in the slope coefficients. Similarly, a bivariate system necessarily has at most one cointegrating relationship. Thus, the Larsson, Lyhagen, and Löthgren (2001) test has no opportunity to detect multiple cointegration. But, the Kao (1999) tests require a common β for all i. Hence, in order to be able to validly compare all tests under both the null and the alternative we use this simple DGP. We carry out the experiments under both the null and the alternative.[3] For the latter we set $\rho = 0.98$. The fraction of cointegrated series in the panel is either zero or one. For a given cross-sectional dimension we draw the unit specific intercepts as $\alpha_i \sim \mathcal{U}[0, 10]$ and keep them fixed for all T. The number of replications for each experiment is $M = 10,000$.

Here, we report the (representative) results for $a_1 = 1,\ \sigma = 1,\ \psi = 0$.[4] Table 7.1 shows the correlation of the empirical p-values for $N = 50$. Panels (a) and (b) consider $T = 250$ and $T = 2000$, respectively. Within each of the panels there is a fairly high correlation among the different residual-based tests (rows 2-8) and, especially, among the different system-based tests (rows 1, 9-10). The pattern is not uniform, though. For the residual-based tests, the correlation ranges from roughly 30% ($P_{\chi^2 DF}$ and $Z_{\hat{t}_{NT}}$) to almost 95% (DF^*_ρ and $Z_{\hat{\rho}_{NT}-1}$). For a graphical illustration, see the scatter plot of the empirical p-values for these cases in Figure 7.1. Panel (a), depicting the correlation of $P_{\chi^2 DF}$ and $Z_{\hat{t}_{NT}}$, shows that, even within the group of residual-based tests, virtually any combination of empirical p-values can arise. On the other hand, Panel (b) reveals that for some cases the p-values cluster around the 45°-line, indicating a close correspondence.

Furthermore, different tests by the same author do not seem to be any more related than tests by different authors. Across the two groups the correlation typically is substantially lower, with several entries even being negative (see, e.g., the first column). Finally, compare Panels (a) and (b). Increasing the

[3] Uniform random numbers are generated using the KM algorithm from which Normal variates are created with the fast acceptance-rejection algorithm, both implemented in GAUSS. Part of the calculations are performed with COINT 2.0 by Peter Phillips and Sam Ouliaris.

[4] The full set of results of the finite sample study is available upon request.



upper panel: $P_{\chi^2 DF}$ and $Z_{\hat{t}_{NT}}$ lower panel: DF^*_ρ and $Z_{\hat{\rho}_{NT}-1}$

FIGURE 7.1: CORRELATION OF EMPIRICAL p-VALUES

time series dimension barely affects the correlation of the empirical p-values. (Similar results obtain for increasing N.)

We provide some further insights in Table 7.2. Using 5% size-adjusted critical values we report the fraction of each pair of test rejecting jointly.[5] The case

[5]Horowitz and Savin (2000) correctly point out that size-adjusted critical values are usually of little use for applied work. Here, however, we use them to avoid spurious results

TABLE 7.1: CORRELATION OF THE EMPIRICAL p-VALUES UNDER THE NULL

	Υ_{LR}	DF_t^*	$Z_{\hat{\rho}NT-1}$	$\tilde{Z}_{\hat{\rho}NT-1}$	DF_ρ^*	$Z_{\hat{t}_{NT}}$	$P_{\chi^2 DF}$	$P_{\Phi^{-1}DF}$	$P_{\chi^2 J}$	$P_{\Phi^{-1}J}$
Υ_{LR}	1.00									
DF_t^*	-.055	1.00			*(a)*	$T=250$				
$Z_{\hat{\rho}NT-1}$	.115	.445	1.00							
$\tilde{Z}_{\hat{\rho}NT-1}$	.264	.312	.698	1.00						
DF_ρ^*	.098	.514	.944	.658	1.00					
$Z_{\hat{t}_{NT}}$	-.087	.935	.486	.341	.492	1.00				
$P_{\chi^2 DF}$	.235	.314	.583	.927	.599	.304	1.00			
$P_{\Phi^{-1}DF}$	.213	.466	.764	.919	.806	.439	.898	1.00		
$P_{\chi^2 J}$	.984	-.059	.116	.268	.099	-.089	.245	.213	1.00	
$P_{\Phi^{-1}J}$	.961	-.045	.106	.242	.090	-.078	.205	.198	.898	1.00
Υ_{LR}	1.00									
DF_t^*	-.096	1.00			*(b)*	$T=2000$				
$Z_{\hat{\rho}NT-1}$	.131	.466	1.00							
$\tilde{Z}_{\hat{\rho}NT-1}$	.346	.320	.652	1.00						
DF_ρ^*	.094	.552	.949	.614	1.00					
$Z_{\hat{t}_{NT}}$	-.112	.938	.505	.359	.530	1.00				
$P_{\chi^2 DF}$	.265	.330	.545	.929	.561	.330	1.00			
$P_{\Phi^{-1}DF}$	.242	.487	.736	.915	.782	.467	.896	1.00		
$P_{\chi^2 J}$	.984	-.097	.135	.351	.099	-.111	.279	.245	1.00	
$P_{\Phi^{-1}J}$	.964	-.090	.119	.318	.083	-.107	.229	.222	.903	1.00

Note: $N=50,\ \rho=1,\ \psi=0,\ \sigma=1,\ \delta=1$ and $a_1=1$. $M=10,000$ replications.

considered in Panel (a) corresponds to Panel (a) of Table 7.1. The entries under (b) give results under the alternative of panel cointegration. As expected from Table 7.1, no pair of tests achieves a fraction of joint rejections of 5%. Reassuringly, the combinations having a high correlation of empirical p-values also have a relatively high fraction of joint rejections. However, in spite of fairly high correlation (take $P_{\chi^2 J}$ and $P_{\Phi^{-1}J}$ with more than 90%) we still observe pair of tests jointly rejecting for a rather small fraction of samples (1.6% for this example). That is, conflicting testing decisions are not uncommon. As all tests reject more frequently under the alternative, the fraction of joint rejections of course increases (see Panel b). Nevertheless, there is still a large amount of disagreement especially across groups of tests.

that could arise if, say, two tests were both heavily oversized and would therefore also frequently reject jointly.

Table 7.2: Fraction of joint rejections under H_0 and H_1

	Υ_{LR}	DF_t^*	$Z_{\hat{\rho}_{NT}-1}$	$\tilde{Z}_{\hat{\rho}_{NT}-1}$	DF_ρ^*	$Z_{\hat{t}_{NT}}$	$P_{\chi^2 DF}$	$P_{\Phi^{-1} DF}$	$P_{\chi^2 J}$	$P_{\Phi^{-1} J}$
					(a)	$\rho = 1$				
Υ_{LR}	.050									
DF_t^*	.002	.050								
$Z_{\hat{\rho}_{NT}-1}$	.005	.011	.050							
$\tilde{Z}_{\hat{\rho}_{NT}-1}$	.007	.008	.021	.050						
DF_ρ^*	.004	.015	.037	.020	.050					
$Z_{\hat{t}_{NT}}$	.002	.036	.012	.011	.014	.050				
$P_{\chi^2 DF}$	.006	.008	.016	.036	.017	.008	.050			
$P_{\Phi^{-1} DF}$	.006	.012	.025	.036	.028	.013	.032	.050		
$P_{\chi^2 J}$	.043	.002	.004	.007	.004	.002	.007	.006	.050	
$P_{\Phi^{-1} J}$	.040	.002	.004	.007	.004	.002	.005	.006	.016	.050
					(b)	$\rho = .98$				
Υ_{LR}	.191									
DF_t^*	.191	.999								
$Z_{\hat{\rho}_{NT}-1}$	.190	.975	.975							
$\tilde{Z}_{\hat{\rho}_{NT}-1}$	.183	.872	.869	.873						
DF_ρ^*	.176	.857	.858	.809	.858					
$Z_{\hat{t}_{NT}}$	.191	.999	.975	.873	.858	1.00				
$P_{\chi^2 DF}$	.135	.516	.516	.516	.504	.516	.516			
$P_{\Phi^{-1} DF}$	.169	.750	.750	.749	.733	.750	.513	.750		
$P_{\chi^2 J}$	.164	.174	.173	.166	.161	.174	.124	.155	.174	
$P_{\Phi^{-1} J}$	.173	.227	.225	.215	.208	.227	.153	.197	.028	.227

Note: $N = 50,\ T = 250,\ \psi = 0,\ \sigma = 1,\ \delta = 1$ and $a_1 = 1$.
$M = 10,000$ replications. Size-adjusted 5% critical values.

Comparing the results with Gregory, Haug, and Lomuto (2004), we state that the consensus in test decisions among panel data cointegration tests generally does not seem to be higher than among time series cointegration tests. Thus, it seems all but unlikely that a researcher will find conflicting evidence when applying some pairs of panel cointegration tests to a given dataset. The issue is not resolved asymptotically. A possible explanation of this phenomenon could be that the complexities inherent to panel data—such as treatment of cross-sectional heterogeneity—lead to different implicit alternatives of the tests. Consequently, we observe a rather low correlation of empirical p-values and fractions of joint rejections when the data is generated under the null.

7.4 Conclusion

We perform a simulation study to investigate whether several widely used panel cointegration tests yield the same acceptance or rejection decisions. Broadly in accordance with the evidence presented by Gregory, Haug, and Lomuto (2004) for time series tests, the panel versions also exhibit a low correlation of empirical p-values under the null. The persistence of the phenomenon even at $T = 2000$ indicates that this problem does not seem to be resolved asymptotically. When analyzing the relative frequency of joint rejections, we constrain the tests to have the desired size by using size-adjusted critical values. Low fractions of joint rejections (relative to the size of the tests) show that the tests do not reject for the same samples. This phenomenon is less prevalent under the alternative.

The practical upshot is that researchers are likely to be confronted with conflicting test decisions when using different tests in applied work. Given that there rarely is a compelling theoretical reason to prefer one test over another in practice, this issue is rather troublesome. More research clarifying the theoretical relationship between the different tests would be welcome.

Chapter 8

For Which Countries did PPP hold? A Multiple Testing Approach

Abstract

We use recent advances in multiple testing to identify those countries for which Purchasing Power Parity (PPP) held over the last century. The approach controls the multiplicity problem inherent in simultaneously testing for PPP on several time series, thereby avoiding spurious rejections. It has higher power than traditional multiple testing techniques by exploiting the dependence structure between the countries with a bootstrap approach. We use a sieve bootstrap approach to account for nonstationarity under the null hypothesis. Our empirical results show that, plausibly, controlling for multiplicity in this way leads to a number of rejections of the null of no PPP that is intermediate between that of traditional multiple testing techniques and that which results if one tests the null on each single time series at some level α.

Keywords: Multiple Testing, Bootstrap, PPP, Panel Data

8.1 Introduction

Purchasing Power Parity (PPP) is among the most popular theories to explain the long run behaviour of exchange rates. Not least because it is ready-made for empirical implementation, it has been investigated by a host of econometric techniques. So-called "stage-two" tests (Froot and Rogoff, 1995) test the hypothesis that the real exchange rate follows a random walk. The alternative is that the real exchange rate is a stationary process, i.e. that PPP holds in the long run. Typically, researchers would obtain real exchange rate data over a certain time span for several countries and conduct appropriate unit root tests on each series (see, e.g., Taylor, 2002). It is then argued that PPP holds for those countries for which the null is rejected.

Unfortunately, this simple and intuitive way of investigating the validity of PPP is problematic from a statistical point of view. In particular, it ignores the issue of "multiplicity." To illustrate the problem, consider the following artificial numerical example. Suppose one has exchange rate data (relative to some reference country) on a panel of, say, $N = 20$ countries. Also assume for simplicity that the units are independent and that PPP does not hold for any of the units. When conducting tests on each unit at, say, the $\alpha = 0.05$ level, one might expect to erroneously find evidence in favor of PPP once on average, because $1/20 = 0.05$. However, the event of a rejection is a Bernoulli random variable with "success" probability 0.05. Hence, the number of k rejections in N tests, is a Binomial random variable with probability function

$$P_k = \binom{N}{k} \alpha^k (1-\alpha)^{N-k}.$$

Therefore, the probability of (at least) one erroneous rejection, also known as the Familywise Error Rate[1] ($FWER$), equals

$$P_{k\geqslant 1} = \sum_{j=1}^{20} \binom{20}{j} 0.05^j (1-0.05)^{20-j} = 0.6415.$$

Even if PPP does not hold for any of the countries in the panel, one will falsely find some evidence of it with a rather high probability. Of course, the problem only worsens if one adds more units to the panel.

This so-called “multiplicity” problem, while not widely recognized in econometrics (Savin, 1984), has of course been realized long ago in the statistics literature (see Lehmann and Romano, 2005). Several solutions to controlling the $FWER$ at some specified level α have been suggested. Among the most popular are the Bonferroni and the Holm (1979) procedure. These procedures have however been less successful in econometric applications because ensuring $FWER \leqslant \alpha$ typically comes at the price of reducing the ability to identify false hypotheses. That is, the procedures are conservative or have low “power.”[2] Hence, often quite reasonably, researchers have tended to ignore the issue of multiplicity.

Recently, panel data methods have become popular to test for PPP. See for instance Wu (1996), Papell and Theodoridis (2001), Papell (2002) or Murray and Papell (2005). Typically, these panel unit root tests formulate the null of the entire panel being nonstationary. The alternative quite often is that of a stationary panel (see, for instance, Harris and Tzavalis, 1999; Levin, Lin, and Chu, 2002; Breitung, 2000). These panel tests also have power against “mixed” panels, where only some fraction of the units is actually stationary (see Taylor and Sarno, 1998; Karlsson and Löthgren, 2000; Boucher Breuer, McNown, and Wallace, 2001). Hence, erroneous conclusions on the number of countries for

[1]More generally, the j-$FWER$ is defined as $P_{k\geqslant j}$, the probability of j or more false rejections.

[2]For a discussion of “power” in a multiple testing framework see Romano and Wolf (2005), Sec. 2.2.

which PPP holds remain possible. (Concluding from a rejection of a panel unit test that *all* units are stationary is closely related to the erroneous inference that a rejection in an F test of the "significance of a regression" implies that *all* coefficients are nonzero.)

As a partial remedy, Maddala and Wu (1999) and Choi (2001) draw on the meta analytic literature (see Hedges and Olkin, 1985) to provide panel unit root tests having the more conservative alternative that some nonzero fraction of the panel is stationary. However, their approach neither allows to identify which nor how many of the countries in the panel have a stationary real exchange rate.

Recently, there has been substantial research on improving the ability of multiple testing approaches to detect false hypotheses while still controlling the *FWER*. Notably, Romano and Wolf (2005) have put forward a bootstrap scheme that exploits the dependence structure of the statistics in order to improve the power of the multiple test. In the present chapter, we propose an adaptation of the Romano and Wolf (2005) approach to identify those countries of a panel of real exchange rate data for which the Purchasing Power Parity condition holds.

The plan of the chapter is as follows. Section 2 offers a brief statement of the PPP condition and presents the general multiple testing approach of Romano and Wolf (2005). Section 3 discusses the bootstrap approach employed in this chapter. The empirical results are in Section 4. Section 5 concludes.

8.2 The Multiple Testing Approach

Our goal is to identify those countries of a panel for which the Purchasing Power Parity (PPP) relation held over the sample period. Let $p_{i,t}$ be the (log) price level in country i and period t, where $i = 1, \ldots, N$ and $t = 1, \ldots, T$,

p_t^* the "foreign" (log) price level of the reference country in the panel and $s_{i,t}$ the (log) nominal exchange rate between the currencies of country i and the reference country. The real exchange rate is then given by

$$r_{i,t} = p_{i,t} - p_t^* - s_{i,t} \qquad (i = 1, \ldots, N)$$

Testing the strong PPP hypothesis is naturally formulated (see Rogoff, 1996) as a unit root test on the real exchange rate. A vast number of unit root tests have been suggested in the literature (see Phillips and Xiao, 1998, for a survey), many of which have been applied to the PPP question. We will use the standard augmented Dickey and Fuller (1979) test (see also Said and Dickey, 1984). We do so because it is still the most popular unit root test and, more importantly, the bootstrap versions of the test required for the multiple testing scheme have desirable properties (Swensen, 2003; Chang and Park, 2003). Accordingly, we investigate PPP by testing the individual hypotheses

$$H_i : \varrho_i = 0 \qquad \text{vs.} \qquad H_i' : \varrho_i < 0 \quad (i = 1, \ldots, N) \quad (8.1)$$

where

$$\Delta r_{i,t} = \varrho_i r_{i,t-1} + \sum_{j=1}^{J_i} \nu_j \Delta r_{i,t-j} + \epsilon_{J_i,i,t}. \quad (8.2)$$

The number of lagged differences J_i required to capture serial correlation in $r_{i,t}$, is allowed to vary across i. Our test statistic is given by $\hat{\tau}_i = \hat{\varrho}_i / s.e.(\hat{\varrho}_i)$, the t-statistic of ϱ_i in (8.2), where $\hat{\varrho}_i$ is the usual OLS estimator and $s.e.(\hat{\varrho}_i)$ the associated standard error.

We aim to determine those countries $I \subset \{1, \ldots, N\}$ for which $r_{i,t}$ is a stationary process. As argued in the Introduction, in order to provide reliable statistical inference in the sense of controlling the *FWER*, it is important to take into account the multiplicity inherent in testing in a panel setting. We now present the general multiple testing framework used here, making suitable adjustments to adapt the procedure to the PPP testing case.

First, relabel the test statistics from smallest to largest, such that $\hat{\tau}_{r_1} \leqslant \hat{\tau}_{r_2} \leqslant \cdots \leqslant \hat{\tau}_{r_N}$. (The smaller a Dickey-Fuller test statistic, the stronger the evidence in favor of stationarity.) Form a joint rectangular confidence region for the vector $(\varrho_{r_1}, \ldots, \varrho_{r_N})^\top$. The region is of the form

$$(-\infty,\ \hat{\varrho}_{r_1} + s.e.(\hat{\varrho}_{r_1}) \cdot d_1] \times \cdots \times (-\infty,\ \hat{\varrho}_{r_N} + s.e.(\hat{\varrho}_{r_N}) \cdot d_1], \tag{8.3}$$

where one chooses d_1 so as to ensure a joint asymptotic coverage probability $1-\alpha$.[3] The bootstrap method to appropriately choose d_1 in the present problem will be discussed below. The decision rule is to reject a particular hypothesis H_{r_n} if the corresponding confidence interval satisfies $0 \notin (-\infty,\ \hat{\varrho}_{r_n} + s.e.(\hat{\varrho}_{r_n}) \cdot d_1]$. Romano and Wolf (2005) show that if the confidence region (8.3) has coverage probability $1-\alpha$, then this method asymptotically controls the $FWER$ at level α, $\lim_T FWER \leqslant \alpha$. Crucially, the method does not stop there. In order to improve the ability of the method to detect false hypotheses, one can construct further confidence regions after having rejected, say, the first N_1 hypotheses. In a second step, one forms a confidence region for the remaining $N - N_1$ coefficients $(\varrho_{r_{N_1+1}}, \ldots, \varrho_{r_N})^\top$. This is again constructed to have nominal joint coverage probability $1-\alpha$ and is of the form

$$(-\infty,\ \hat{\varrho}_{r_{N_1+1}} + s.e.(\hat{\varrho}_{r_{N_1+1}}) \cdot d_2] \times \cdots \times (-\infty,\ \hat{\varrho}_{r_N} + s.e.(\hat{\varrho}_{r_N}) \cdot d_2],$$

potentially leading to the rejection of some further N_2 hypotheses. This step-down process can be repeated until no further hypotheses are rejected. Romano and Wolf (2005) show that the d_j should ideally be chosen as

$$d_j \equiv d_j(1-\alpha, P) = \inf\left\{x : \Pr_P\left[\max_{R_{j-1}+1 \leqslant n \leqslant N}\left(\frac{\hat{\varrho}_{r_n} - \varrho_{r_n}}{s.e.(\hat{\varrho}_{r_n})}\right) \leqslant x\right] \geqslant 1-\alpha\right\},$$

where $R_{j-1} = \sum_{k=0}^{j-1} N_k$ and $R_0 = 0$. In practice, however, P and hence d_j are unknown. Fortunately, Romano and Wolf (2005, Thms. 3.1 and 4.1) show that d_j can often be estimated consistently with the bootstrap without affecting asymptotic control of the $FWER$.

[3]As recommended by Romano and Wolf (2005) we use the studentized version of their method. For a discussion of the "basic" approach, see Sec. 3 of their paper.

8.3 The Bootstrap Algorithm

We now outline the bootstrap approach to obtain an estimator $\hat{d}_j$ employed in this chapter.

1. Fit an autoregressive process to $\Delta r_{i,t}$ $(i = 1, \ldots, N;\ t = 2, \ldots, T)$. It is natural to use the Yule-Walker procedure because it always yields an invertible representation (Brockwell and Davis, 1991, Secs. 8.1–2). Letting $\overline{\Delta r_i} := (T_i - 1)^{-1} \sum_{t=2}^{T_i} \Delta r_{i,t}$, compute the empirical autocovariances of $\Delta r_{i,t}$ up to order q,

$$\hat{\gamma}_i(\ell) := \frac{1}{T_i - 1 - \ell} \sum_{t=2}^{T_i - \ell} (\Delta r_{i,t} - \overline{\Delta r_i})(\Delta r_{i,t+\ell} - \overline{\Delta r_i}),$$

where $i = 1, \ldots, N;\ \ell = 1, \ldots, q$.[4] Defining

$$\hat{\Gamma}_{i,q} := \begin{pmatrix} \hat{\gamma}_i(0) & \cdots & \hat{\gamma}_i(q-1) \\ \vdots & \ddots & \vdots \\ \hat{\gamma}_i(q-1) & \cdots & \hat{\gamma}_i(0) \end{pmatrix}$$

and $\hat{\gamma}_i := (\hat{\gamma}_i(1), \ldots, \hat{\gamma}_i(q))^\top$, obtain the AR coefficient vector as

$$(\hat{\phi}_{q,i,1}, \ldots, \hat{\phi}_{q,i,q})^\top := \hat{\Gamma}_{i,q}^{-1} \hat{\gamma}_i. \qquad (i = 1, \ldots, N)$$

2. The residuals are, as usual, given by

$$\hat{\epsilon}_{q,i,t} := \Delta r_{i,t} - \sum_{\ell=1}^{q} \hat{\phi}_{q,i,\ell} \Delta r_{i,t-\ell},$$

for $i = 1, \ldots, N$; $t = q + 2, \ldots, T$. Center $\hat{\epsilon}_{q,i,t}$ to obtain

$$\tilde{\epsilon}_{q,i,t} := \hat{\epsilon}_{q,i,t} - \frac{1}{T_i - q - 1} \sum_{g=q+2}^{T_i} \hat{\epsilon}_{q,i,g}$$

for $i = 1, \ldots, N$; $t = q + 2, \ldots, T$.

[4]In practice, q can be chosen with a data-dependent criterion such as Akaike's.

3. Resample nonparametrically from $\tilde{\epsilon}_{q,i,t}$ to get $\epsilon^*_{q,i,t}$. To preserve the empirical cross-sectional dependence structure, jointly resample residual vectors

$$\tilde{\epsilon}_{q,\cdot,t} := (\tilde{\epsilon}_{q,1,t}, \ldots, \tilde{\epsilon}_{q,N,t}). \qquad (t = q+2, \ldots, T)$$

   See Chapter 4 for evidence of the good performance of this step to account for cross-sectional dependence.

4. Recursively construct the bootstrap samples as[5]

$$\Delta r^*_{q,i,t} = \sum_{\ell=1}^{q} \hat{\phi}_{q,i,\ell} \Delta r^*_{q,i,t-\ell} + \epsilon^*_{q,i,t}$$

   for $i = 1, \ldots, N$, $t = q+2, \ldots, T$.

5. It is necessary to impose the null of a unit root when generating the artificial data in bootstrap unit root tests to achieve consistency (Basawa, Mallik, McCormick, Reeves, and Taylor, 1991). Accordingly, impose the null of nonstationarity by integrating $\Delta r^*_{i,t}$ to obtain $r^*_{i,t}$.

6. For each sample $r^*_b := \big((r^*_{b,1,1}, \ldots, r_{b,1,T})^\top, \ldots, (r^*_{b,N,1}, \ldots, r^*_{b,N,T})^\top\big)$, compute the test statistics τ^*_{b,r_n}, $r_n = R_{j-1}+1, \ldots, N$, and

$$\max^*_{b,j} := \max_{R_{j-1}+1 \leqslant n \leqslant N} (\tau^*_{b,r_n} - \hat{\tau}_{r_n}).$$

7. Repeat steps 3 to 6 many, say B, times.

8. Compute $\hat{d}_j$ as the $1-\alpha$ quantile of the B values $\max^*_{1,j}, \ldots, \max^*_{B,j}$.

Chang and Park (2003) and Swensen (2003) show that the above sieve bootstrap scheme yields asymptotically valid bootstrap ADF tests in the sense that using the α quantile of the bootstrap distribution of the τ^*_{b,r_n} as critical value asymptotically gives a test with size α. By a continuous mapping theorem argument, the bootstrap also consistently estimates the distribution of the $\max^*_{b,j}$ and hence $\hat{d}_j$.

[5]We run the recursion for 30 initial observations before using the $\Delta r^*_{q,i,t}$ to mitigate the effect of initial conditions.

TABLE 8.1: EMPIRICAL RESULTS

country	$\hat{\tau}_i$	p-value	Holm criterion
Mexico	-4.334	< 0.001	0.0026
Finland	-4.136	0.001	0.0028
Argentina	-3.632	0.006	0.0029
Italy	-3.344	0.015	0.0031
Norway	-3.285	0.018	0.0033
Sweden	-3.202	0.022	0.0036
UK	-2.996	0.038	0.0038
Belgium	-2.980	0.040	0.0042
Germany	-2.957	0.042	0.0046
France	-2.929	0.045	0.0050
Brazil	-2.561	0.104	0.0056
Australia	-2.544	0.108	0.0063
Netherlands	-2.498	0.119	0.0071
Portugal	-2.391	0.147	0.0083
Canada	-2.202	0.207	0.0100
Spain	-2.118	0.238	0.0125
Denmark	-2.058	0.262	0.0167
Switzerland	-1.349	0.604	0.0250
Japan	-1.323	0.617	0.0500

8.4 Results

We now present the empirical results of an application of the modified Romano and Wolf (2005) methodology to the PPP condition. We revisit the dataset used by Taylor (2002), which includes annual data for the nominal exchange rate, CPI and the GDP deflator. This dataset is particularly useful for our purposes because it covers a long period, ranging from 1892 through to 1996. The countries contained in our panel are given in Table 8.1. We use the United States as the reference country throughout and report results using CPI price series. See Taylor (2002) for further details on data sources and definitions.

Using standard ADF unit root tests, we find rejections for 9 out of 19 countries at the 5% critical value -2.94. See the first column of Table 8.1. (The entries are sorted for later use.) The number of lagged differences J_i in (8.2) is chosen with the data-dependent criterion of Ng and Perron (2001). The findings of Taylor (2002) are very similar.[6] Evidence in favor of PPP is therefore at best mixed. Taylor (2002) then argues that it may be possible to find more rejections in favor of PPP by employing more powerful techniques. Our goal, on the other hand, is to investigate whether some of the rejections are spurious in the sense that they would not occur when taking into account the multiplicity of the testing problem.

As a preliminary step, we report results for the more classical techniques to control the *FWER*, namely the Bonferroni and the Holm (1979) procedures. Recall that the former rejects H_i if the p-value $\hat{p}_i$ corresponding to the test statistic $\hat{\tau}_i$ satisfies $\hat{p}_i \leqslant \alpha/N$. The Holm (1979) procedure first sorts the p-values from smallest to largest, $\hat{p}_{r_1} \leqslant \cdots \leqslant \hat{p}_{r_N}$. Relabel the hypotheses accordingly as H_{r_n}. Then, reject H_{r_n} at level α if $\hat{p}_{r_j} \leqslant \alpha/(N-j+1)$ for all $r_j = 1, \ldots, r_n$.[7] The cutoff value for the first hypothesis is identical for both methods, but unlike the Bonferroni method, the Holm (1979) procedure uses gradually less challenging criteria for $H_{r_2}, \ldots, H_{r_N}$. Nevertheless, it often has low power because it also fails to exploit the dependence structure between the statistics.

The limit distribution of the ADF test statistics is a functional of Brownian motions that cannot be evaluated analytically to obtain p-values. We therefore rely on response surface regressions suggested by MacKinnon (1994, 1996) to obtain numerical distribution functions of the test statistics. We report results in columns 2 and 3 of Table 8.1.

As expected, the number of rejections is now much lower. After controlling for multiplicity, we only observe rejections for Mexico and Finland for either

[6] The small differences can be explained by different interpolation schemes for missing wartime data, other lag selection criteria as well as the fact that we balance our panel.

[7] See Lehmann and Romano (2005) for a proof that the Bonferroni and the Holm method control the *FWER* at level α.

method. These results indeed suggest that the Bonferroni and Holm procedures are conservative.

We therefore now turn to the results of the Romano and Wolf (2005) approach. The algorithm presented in Section 2 yields $\hat{d}_1 = 4.050$, leading to a rejection for Mexico and Finland. In the second round, we obtain $\hat{d}_2 = 3.429$, implying evidence in favor of PPP for Argentina. Next, we find $\hat{d}_3 = 3.252$ such that we reject for Italy and Norway. Finally, $\hat{d}_4 = 3.075$ means that we also reject the null in the case of Sweden.

Observe that the number of rejections is intermediate between the results for the Holm and Bonferroni methods and that of the individual country results. In view of the above discussion, we find that this result is rather plausible. Furthermore, the ability of the Romano and Wolf (2005) method to detect several false hypotheses in a stepwise fashion proved instrumental in improving upon the more traditional multiple testing methods.

8.5 Conclusion

We have used recent advances in the multiple testing literature to attempt to identify those countries for which Purchasing Power Parity (PPP) held over the last century. The approach controls the multiplicity problem inherent in simultaneously testing for PPP on several time series, thereby avoiding spurious rejections. It has higher power than traditional multiple testing techniques by exploiting the dependence structure between the countries with a bootstrap approach. We use a sieve bootstrap approach to account for nonstationarity under the null hypothesis. On the other hand, our empirical results show that, plausibly, controlling for multiplicity leads to fewer rejections of the null of no PPP than if one tests the null on each single time series at some level

α. Specifically, we find rejections of the null of no PPP for Mexico, Finland, Argentina, Italy, Norway and Sweden.

Several open issues remain. Hlouskova and Wagner (2006) point out that bootstrapping in a nonstationary framework is a "delicate issue." It would therefore be interesting to investigate the performance of other resampling techniques in the present problem. Consider, for instance, block bootstrapping as in Psaradakis (2006).

Obviously, the present framework is fairly general and could be applied to other macroeconomic questions such as savings-investment correlation or spot and forward exchange rates (Mark, Ogaki, and Sul, 2005) that have hitherto been dealt with using panel techniques. Similarly, it is possible to accommodate problems that imply testing for cointegration.

Chapter 9

OLS-based Estimation of the Disturbance Variance under Spatial Autocorrelation

Abstract

We[1] investigate the OLS-based estimator s^2 of the disturbance variance in the standard linear regression model with cross section data when the disturbances are homoskedastic, but spatially correlated. For the most popular model of spatially autoregressive disturbances, we show that s^2 can be severely biased in finite samples, but is asymptotically unbiased and consistent for most types of spatial weighting matrices as sample size increases.

Keywords: regression, spatial error correlation, bias, variance

[1]This chapter has been written jointly with Walter Krämer.

9.1 Introduction

We consider the standard linear regression model

$$y = X\beta + u,$$

where y is $N \times 1$, X is nonstochastic $N \times K$ with rank K and β is unknown $K \times 1$. The components of u have expected value $E(u) = 0$ and a common variance $E(u_i^2) = \sigma^2$. The OLS estimate for β is $\hat{\beta} = (X'X)^{-1}X'y$, and the OLS-based estimate for σ^2 is

$$s^2 = \frac{1}{N-K}(y - X\hat{\beta})'(y - X\hat{\beta}) = \frac{1}{N-K}u'Mu, \tag{9.1}$$

where $M = I - X(X'X)^{-1}X'$. It has long been known that s^2 is in general (and contrary to $\hat{\beta}$) biased whenever $V := Cov(u)$ is no longer a multiple of the identity matrix. Krämer (1991) and Krämer and Berghoff (1991) show that this problem disappears asymptotically for certain types of temporal correlation such as stationary AR(1)-disturbances, although it is clear from Kiviet and Krämer (1992) that the relative bias of s^2 might still be substantial for any finite sample size. The present chapter extends these analyses to the case of spatial correlation, where we allow the disturbance vector u to be generated by the spatial autoregressive scheme

$$u = \rho W u + \epsilon, \tag{9.2}$$

where ϵ is a $N \times 1$ random vector with mean zero and scalar covariance matrix $\sigma_\epsilon^2 I$ and W is some known $N \times N$-matrix of nonnegative spatial weights with $w_{ii} = 0$ $(i = 1, \ldots, N)$. Such patterns of dependence are often entertained when the objects under study are positioned in some "space," whether geographical or sociological (in some social network, say) and account for spillovers from one unit to its neighbors, whichever way "neighborhood" may be defined. They date back to Whittle (1954) and have become quite popular in econometrics recently. See Anselin and Florax (1995) or Anselin (2001) for surveys of this literature.

The coefficient ρ in (9.2) measures the degree of correlation, which can be both positive and negative. Below we focus on the empirically more relevant case of positive disturbance correlation, where

$$0 \leqslant \rho \leqslant \frac{1}{\lambda_{\max}} \tag{9.3}$$

and where $\lambda_{\max}$ is the Frobenius-root of W (i.e. the unique positive real eigenvalue such that $\lambda_{\max} \geqslant |\lambda_i|$ for arbitrary eigenvalues λ_i). The disturbances are then given by

$$u = (I - \rho W)^{-1}\epsilon, \tag{9.4}$$

so

$$V = Cov(u) = \sigma_\epsilon^2[(I - \rho W)'(I - \rho W)]^{-1} \tag{9.5}$$

which reduces to $V = \sigma_\epsilon^2 I$ whenever $\rho = 0$.

Of course, for our analysis to make sense, the main diagonal of V should be constant, i.e.

$$V = \sigma^2\Sigma, \tag{9.6}$$

where Σ is the correlation matrix of the disturbance vector.[2] It is therefore important to clarify that many, though not all, spatial autocorrelation schemes are compatible with homoskedasticity. Consider for instance the following popular specification for the weight matrix known as "one ahead and one behind:"

$$\tilde{W} := \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \ddots & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 & 1 \\ 1 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

[2]Note that $\sigma^2 = Var(u_i)$ need not be equal to $\sigma_\epsilon^2 = Var(\epsilon_i)$, unless $\Sigma = I$. In the sequel, we keep σ_ϵ^2 fixed, so σ^2 will in general vary with W and N.

and renormalize the rows such that the row sums are 1. Then it is easily seen that $E(u_i^2)$ is independent of i, and analogous results hold for the more general "j ahead and j behind" weight matrix W which has non-zero elements in the j entries before and after the main diagonal, with the non-zero entries equal to $j/2$. This specification has been considered by, for instance, Kelejian and Prucha (1999) and Krämer and Donninger (1987).

As another example, consider the equal-weight matrix (see, e.g., Kelejian and Prucha (2002), Lee (2004), Case (1992) or Kelejian, Prucha, and Yuzefovich (2006)), defined by

$$W^{EW} = (w_{ij}^{EW}) = \begin{cases} \frac{1}{N-1} & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases}. \tag{9.7}$$

One easily verifies that, for $|\rho| < 1$,

$$(I - \rho W^{EW})^{-1} = \delta_1 J_N + \delta_2 I_N, \tag{9.8}$$

where

$$\delta_1 = \frac{\rho}{(N-1+\rho)(1-\rho)}, \; \delta_2 = \frac{N-1}{N-1+\rho} \tag{9.9}$$

and J_N is an $(N \times N)$ matrix of ones. Without loss of generality, let $\sigma_\epsilon^2 = 1$. We then have, using symmetry of W,

$$\begin{aligned} V &= [(I - \rho W^{EW})'(I - \rho W^{EW})]^{-1} \\ &= (I - \rho W^{EW})^{-1}(I - \rho W^{EW})^{-1} \\ &= (\delta_1 J_N + \delta_2 I_N)^2. \end{aligned} \tag{9.10}$$

Carrying out the multiplication, it is seen that

$$E(u_i^2) = (\delta_1^2 + \delta_2^2)^2 + (N-1)\delta_1^2 \quad \text{for} \quad i = 1, \ldots, N.$$

So V is homoskedastic. It is straightforward to extend this result to the case where W is block-diagonal with B blocks of dimension $(R \times R)$, defined as

$$W_R^{EW} = (w_{R,ij}^{EW}) = \begin{cases} \frac{1}{R-1} & \text{for } i \neq j \\ 0 & \text{for } i = j, \end{cases} \tag{9.11}$$

where $N = BR$. We therefore conclude that our analysis is applicable in many relevant spatial econometric specifications.

9.2 The relative bias of s^2 in finite samples

We have

$$\begin{aligned} E\left(\frac{s^2}{\sigma^2}\right) &= E\left(\frac{1}{\sigma^2(N-K)}u'Mu\right) \\ &= \frac{1}{\sigma^2(N-K)}tr(MV) \\ &= \frac{1}{N-K}tr(M\Sigma). \end{aligned} \tag{9.12}$$

Watson (1955) and Sathe and Vinod (1974) derive the (attainable) bounds

$$\begin{aligned} &\text{mean of } N-K \text{ smallest eigenvalues of } \Sigma \\ &\qquad \leqslant E\left(\frac{s^2}{\sigma^2}\right) \leqslant \\ &\text{mean of } N-K \text{ largest eigenvalues of } \Sigma, \end{aligned} \tag{9.13}$$

which shows that the bias can be both positive and negative, depending on the regressor matrix X, whatever Σ may be. Finally, Dufour (1986) points out that the inequalities (9.13) amount to

$$0 \leqslant E\left(\frac{s^2}{\sigma^2}\right) \leqslant \frac{N}{N-K} \tag{9.14}$$

when no restrictions are placed on X and Σ. Again, these bounds are sharp and show that underestimation of σ^2 is much more of a threat in practise than overestimation.

The problem with Dufour's bounds is that they are unnecessarily wide when extra information on V is available. Here we assume a disturbance covariance matrix V as in (9.5) and show first that the relative bias of s^2 depends crucially on the interplay between X and W. In particular, irrespective of sample size and of the weighting matrix W, we can always produce a regressor matrix X such that $E(s^2/\sigma^2)$ becomes as close to zero as desired. To see this, let W be symmetric[3] and let

$$W = \sum_{i=1}^{N} \lambda_i \omega_i \omega_i' \tag{9.15}$$

be the spectral decomposition of W, with the eigenvalues λ_i in increasing order and ω_i the corresponding orthonormal eigenvectors. Now it is easily seen that

$$\lim_{\rho \to 1/\lambda_N} E\left(\frac{s^2}{\sigma^2}\right) = 0 \tag{9.16}$$

whenever

$$M\omega_N = 0. \tag{9.17}$$

This follows from

$$V = \sigma_\epsilon^2 \left[\sum_{i=1}^{N} \frac{1}{(1-\rho\lambda_i)^2} \omega_i \omega_i'\right] \tag{9.18}$$

and

$$\Sigma = \frac{1}{\sigma^2} V = \frac{1}{\sum_{i=1}^{N} \frac{1}{(1-\rho\lambda_i)^2} \omega_{i1}^2} \sum_{i=1}^{N} \frac{1}{(1-\rho\lambda_i)^2} \omega_i \omega_i', \tag{9.19}$$

where ω_{i1}^2 is the $(1,1)$-element of $\omega_i \omega_i'$ (under homoscedasticity, we could select any diagonal element of $\omega_i \omega_i'$) and

$$\sigma^2 = \sigma_\epsilon^2 \sum_{i=1}^{N} \frac{1}{(1-\rho\lambda_i)^2} \omega_{i1}^2. \tag{9.20}$$

Multiplying the numerator and denominator of (9.19) by $(1-\rho\lambda_N)^2$, we obtain

$$\Sigma = \frac{1}{\sigma^2} V = \frac{1}{\sum_{i=1}^{N} \frac{(1-\rho\lambda_N)^2}{(1-\rho\lambda_i)^2} \omega_{i1}^2} \sum_{i=1}^{N} \frac{(1-\rho\lambda_N)^2}{(1-\rho\lambda_i)^2} \omega_i \omega_i', \tag{9.21}$$

[3]Notice that for all the homoskedastic examples considered above, row-normalization does not destroy symmetry of W.

which tends to

$$\frac{1}{\omega_{N1}^2}\omega_N\omega_N' \tag{9.22}$$

as $\rho \to 1/\lambda_N$. Given W, one can therefore choose X to be $(N \times 1)$ and equal to ω_N. Then, M is by construction orthogonal to ω_N, which implies that $tr(M\Sigma)$ and therefore also $E(s^2/\sigma^2)$ tend to zero as $\rho \to 1/\lambda_N$.

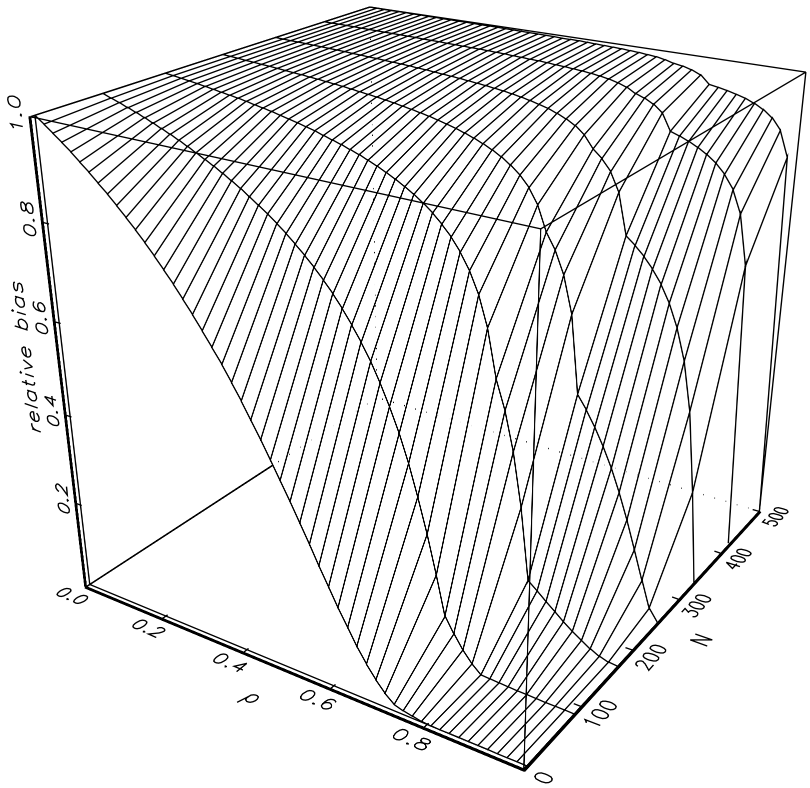


FIGURE 9.1: THE RELATIVE BIAS OF s^2 AS A FUNCTION OF ρ AND N

For illustration, consider the following example. The largest eigenvalue λ_N of a row-normalized matrix such as $\tilde{W}/2$ is 1. (This follows immediately from Theorem 8.1.22 of Horn and Johnson (1985).) It is then readily verified that $\omega_N = \iota := (1, \ldots, 1)'$ is (up to the usual multiple) the eigenvector corresponding to λ_N. Now, if $X = \iota$, $M\omega_N = (I - \frac{1}{N}\iota\iota')\iota = 0$. Figure 9.1 shows the behaviour of the relative bias as $\rho \to 1/\lambda_N = 1$. We see that (9.16) holds for any given N. Also, pointwise in ρ, the relative bias vanishes as $N \to \infty$, as one would expect. We now rigorously establish the latter property.

9.3 Asymptotic bias and consistency

From (9.14), it is clear that, for any V, the relative upward bias of s^2 must vanish as $N \to \infty$. A sufficient condition for the relative downward bias to disappear as well is that the largest eigenvalue of Σ, μ_N, is

$$\mu_N = o(N). \tag{9.23}$$

This is so because, using $\sum_{i=1}^{N} \mu_i = \sum_{i=1}^{N-K} \mu_i + \sum_{i=1}^{K} \mu_{i+N-K} = N$, we have

$$\begin{aligned}\text{mean of } N-K \text{ smallest eigenvalues of } \Sigma &= \frac{N}{N-K} - \frac{1}{N-K}\sum_{i=1}^{K}\mu_{i+N-K} \\ &\geqslant \frac{N}{N-K} - \frac{K}{N-K}\mu_N\end{aligned}$$

and the right-hand side tends to 1 when (9.23) holds as $N \to \infty$.

Condition (9.23) also guarantees consistency. From (9.1), we have

$$s^2 = \frac{1}{N}u'Mu = \frac{1}{N}u'u - \frac{1}{N}u'Hu, \tag{9.24}$$

where $H = X(X'X)^{-1}X'$. Since $u'u/N \xrightarrow{p} \sigma^2$, it remains to show that

$$\frac{1}{N}u'Hu \xrightarrow{p} 0. \tag{9.25}$$

To this purpose, consider

$$\begin{aligned}E\left(\frac{1}{N}u'Hu\right) &= E\left(\frac{1}{N}\varepsilon'\Sigma^{1/2}H\Sigma^{1/2}\varepsilon\right) \qquad (\text{where } \varepsilon = \Sigma^{-1/2}u) \\ &= \frac{\sigma^2}{N}tr(\Sigma^{1/2}H\Sigma^{1/2}) \\ &= \frac{\sigma^2}{N}tr(H\Sigma) \\ &\leqslant \frac{\sigma^2}{N}K \cdot \mu_N, \end{aligned} \tag{9.26}$$

where the inequality follows from the fact that $H\Sigma$ has rank K (since rank $(H) = K$). Since no eigenvalue of $H\Sigma$ can exceed μ_N, and $H\Sigma$ has exactly K nonzero eigenvalues, the inequality follows from the well known fact that the trace of a matrix equals the sum of its eigenvalues. By assumption, $\mu_N/N \to 0$ as $N \to \infty$, so in view of (9.26), $E(u'Hu/N) \to 0$. As $u'Hu$ is nonnegative, this in turn implies $u'Hu/N \xrightarrow{p} 0$ and therefore the consistency of s^2.

The crucial condition (9.23) is a rather mild one; in the present context, it obviously depends on the weighting matrix W. From (9.6) and (9.18), we have

$$\mu_N = \frac{\sigma_\epsilon^2}{\sigma^2(1-\rho\lambda_N)^2}, \tag{9.27}$$

so the condition (9.23) obtains whenever

$$\sigma^2(1-\rho\lambda_N)^2 N \to \infty \tag{9.28}$$

For row-normalized weight matrices, $\lambda_N \equiv 1$ irrespective of N, so (9.28) holds trivially, provided σ^2 remains bounded away from zero. This in turn follows from the fact that, in view of (9.20),

$$\sigma^2 \geqslant \frac{\sigma_\epsilon^2}{(1-\rho\lambda_N)^2} \sum_{i=1}^{N} \omega_{i1}^2, \tag{9.29}$$

where

$$\sum_{i=1}^{N} \omega_{i1}^2 = 1 \tag{9.30}$$

as $\Omega = (\omega_1, \ldots, \omega_N)$ satisfies $\Omega\Omega' = I$.

As another example, consider the "one ahead and one behind" matrix adapted to a "non-circular world" where the $(1, N)$ and $(N, 1)$ entries of $\tilde{W}$ are set to zero, such that after row-normalization,

$$W' := \begin{pmatrix} 0 & 0.5 & 0 & \cdots & 0 & 0 \\ 0.5 & 0 & \ddots & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & \cdots & 0 & 0.5 & 0 \end{pmatrix}.$$

Ord (1975) shows that the eigenvalues of W' are then given by

$$\lambda_i' = \cos\left(\frac{\pi i}{N+1}\right),\ i = 1, \ldots, N,$$

so

$$\lambda_i' \in [-1, 1],\ i = 1, \ldots, N.$$

Chapter 10

Concluding Remarks

The previous eight chapters report my research of the last 18 months. Since the main messages have already been stated in the Introduction, I would now like to give a brief outlook on possible extensions of the work presented here.

First, the PPP Condition was used throughout as a common device to illustrate the methods and issues discussed in this thesis. Obviously, there are many other macroeconometric topics to which the machinery developed here could be applied. As such, the multiple testing approach developed in Chapter 8 might be applied to questions such as savings and investment correlations or the unbiasedness of the forward exchange rate, to name a few. Similarly, panels of investment data are quite obviously cross-sectionally dependent, so that applying the robust tests of Chapter 4 would be an interesting exercise. On a more formal level, as already mentioned in the Conclusion of that Chapter, it would be desirable to analytically establish the consistency of the bootstrap approach put forward there.

Also, I use two rather distinct and competing approaches to deal with dependence among different units such as countries, viz. the more flexible, but potentially less efficient bootstrap and the spatial econometric approach which assumes a lot more structure for the dependence. The spatial econometric approach has so far mainly been used in a cross-sectional framework. There is a recent literature on panel data models under spatial correlation. See, e.g., Baltagi, Song, and Koh (2003) or Kapoor, Kelejian, and Prucha (forthcoming). However, these contributions deal with stationary and small-T panels and are therefore not directly applicable to the problem of testing, e.g., the null of no PPP.

Obviously, the spatial econometric approach of modelling dependence may also be useful in a nonstationary setting. Future work might therefore investigate whether a cross-fertilization of these hitherto fairly independent literatures proves fruitful. As a sort of converse, one might investigate whether popular panel unit root or cointegration tests remain "robust" if cross-sectional dependence is driven by structures typically assumed in the spatial econometrics literature rather than by factor models often assumed in nonstationary models (see Baltagi, Bresson, and Pirotte, 2007).

Bibliography

ANSELIN, L. (2001): "Rao's score test in spatial econometrics," *Journal of Statistical Planning and Inference*, 97(1), 113–139.

ANSELIN, L., AND R. FLORAX (1995): "Small sample properties of tests for spatial dependence in regression models: Some further results," in *New Directions in Spatial Econometrics*, ed. by L. Anselin, and R. Florax, pp. 21–74. Springer, Berlin.

BAI, J., AND S. NG (2004): "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, 72(4), 1127–1177.

BALTAGI, B. H. (2001): *Econometric Analysis of Panel Data*. John Wiley & Sons, New York.

BALTAGI, B. H., G. BRESSON, AND A. PIROTTE (2007): "Panel Unit Root Tests and Spatial Dependence," *Journal of Applied Econometrics*, 22(2), 329–360.

BALTAGI, B. H., AND C. KAO (2000): "Nonstationary Panels, Cointegration in Panels and Dynamic Panels: A Survey," in *Advances in Econometrics*, ed. by B. H. Baltagi, vol. 15, chap. 2, pp. 7–51. Elsevier, Amsterdam.

BALTAGI, B. H., S. H. SONG, AND W. KOH (2003): "Testing Panel Data Regression Models with Spatial Error Correlation," *Journal of Econometrics*, 117, 123–150.

BANERJEE, A. (1999): "Panel Data Unit Roots and Cointegration: An Overview," *Oxford Bulletin of Economics and Statistics*, 61(S1), 607–629.

BANERJEE, A., M. MARCELLINO, AND C. OSBAT (2005): "Testing for PPP: Should We Use Panel Methods?," *Empirical Economics*, 30(1), 77–91.

BASAWA, I. V., A. K. MALLIK, W. P. MCCORMICK, J. H. REEVES, AND R. L. TAYLOR (1991): "Bootstrapping Unstable First-Order Autoregressive Processes," *Annals of Statistics*, 19(2), 1098–1101.

BERNDT, E. R., AND N. E. SAVIN (1977): "Conflict Among Criteria for Testing Hypothesis in the Multivariate Linear Regression Model," *Econometrica*, 45(5), 1263–1277.

BICKEL, P. J., AND K. A. DOKSUM (2001): *Mathematical Statistics*, vol. 1. Prentice Hall, Upper Saddle River, New Jersey, 2nd edn.

BIERENS, H. J. (2005): *Introduction to the Mathematical and Statistical Foundations of Econometrics*. Cambridge University Press, Cambridge.

BOUCHER BREUER, J., R. MCNOWN, AND M. S. WALLACE (2001): "Misleading Inferences from Panel Unit Root Tests with an Illustration from Purchasing Power Parity," *Review of International Economics*, 9(3), 482–493.

BREITUNG, J. (2000): "The Local Power of Some Unit Root Tests for Panel Data," in *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, ed. by B. H. Baltagi, pp. 161–177. Elsevier, Amsterdam.

BREITUNG, J., AND S. DAS (2005): "Panel Unit Root Tests Under Cross Sectional Dependence," *Statistica Neerlandica*, 59(4), 414–433.

BREITUNG, J., AND M. H. PESARAN (2007): "Unit Roots and Cointegration in Panels," in *The Econometrics of Panel Data*, ed. by L. Matyas, and P. Sevestre. Kluwer Academic Publishers, Dordrecht, 3rd edn., To appear.

BROCKWELL, P. J., AND R. A. DAVIS (1991): *Time Series: Theory and Methods*. Springer, New York, 2nd edn.

CAPORALE, G. M., AND M. CERRATO (2006): "Panel Data Tests of PPP: A

Critical Overview," *Applied Financial Economics*, 16(1–2), 73–91.

Caporale, G. M., and N. Pittis (2002): "Unit Roots Versus Other Types of Time-Heterogeneity, Parameter Time Dependence and Superexogeneity," *Journal of Forecasting*, 21(3), 207–223.

Caporale, G. M., N. Pittis, and P. Sakellis (2003): "Testing for PPP: The Erratic Behaviour of Unit Root Tests," *Economics Letters*, 80, 277–284.

Case, A. (1992): "Neighborhood Influence and Technological Change," *Regional Science and Urban Economics*, 22(2), 491–508.

Chang, Y., and J. Y. Park (2003): "A Sieve Bootstrap for the Test of a Unit Root," *Journal of Time Series Analysis*, 24(4), 379–400.

Chang, Y., J. Y. Park, and K. Song (2006): "Bootstrapping Cointegrating Regressions," *Journal of Econometrics*, (forthcoming).

Cheung, Y.-W., and K. S. Lai (1993): "Finite Sample Sizes of Johansen's Likelihood Ratio Tests for Cointegration," *Oxford Bulletin of Economics and Statistics*, 55(3), 313–328.

——— (1994): "Mean reversion in real exchange rates," *Economics Letters*, 46(3), 251–256.

Choi, I. (2001): "Unit Root Tests for Panel Data," *Journal of International Money and Finance*, 20(2), 249–272.

——— (2006): "Combination Unit Root Tests for Cross-Sectionally Correlated Panels," in *Econometric Theory and Practice: Frontiers of Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*, ed. by D. Corbae, S. N. Durlauf, and B. E. Hansen, chap. 11, pp. 311–333. Cambridge University Press.

Coakley, J., N. Kellard, and S. Snaith (2005): "The PPP Debate: Price Matters!," *Economics Letters*, 88, 209–213.

Davidson, R., and J. G. MacKinnon (1993): *Estimation and Inference in*

Econometrics. Oxford University Press, New York.

——— (1999): "The Size Distortion of Bootstrap Tests," *Econometric Theory*, 15, 361–376.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74(366), 427–431.

DUFOUR, J.-M. (1986): "Bias of s^2 in Linear Regression with Dependent Errors," *The American Statistican*, 40(4), 284–285.

ELLIOTT, G., T. ROTHENBERG, AND J. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64(4), 813–836.

ENGLE, R., AND C. GRANGER (1987): "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55(2), 251–76.

FISHER, R. (1970): *Statistical Methods for Research Workers*. Oliver and Boyd, London, 14th edn.

FRANKEL, J. (1986): "International Capital Mobility and Crowding Out in the U.S. Economy: Imperfect Integration of Financial Markets or Goods Markets?," in *How Open is the U.S. Economy?*, ed. by R. Hafer. Lexington Books, Lexington.

FROOT, K. A., AND K. ROGOFF (1995): "Perspectives on PPP and Long-Run Real Exchange Rates," in *The Handbook of International Economics*, ed. by G. Grossman, and K. Rogoff, vol. 3, chap. 32, pp. 1647–1688. Elsevier, Amsterdam.

GONZALO, J. (1994): "Five Alternative Methods of Estimating Long-Run Equilibrium Relationships," *Journal of Econometrics*, 60(1–2), 203–233.

GREGORY, A., AND B. E. HANSEN (1996): "Residual-Based Tests for Cointegration in Models with Regime Shifts," *Journal of Econometrics*, 70, 99–126.

GREGORY, A. W., A. A. HAUG, AND N. LOMUTO (2004): "Mixed Signals

Among Tests for Cointegration," *Journal of Applied Econometrics*, 19(1), 89–98.

Gutierrez, L. (2003): "On the Power of Panel Cointegration Tests: A Monte Carlo Comparison," *Economics Letters*, 80(1), 105–111.

Hansen, B. E. (1992): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business & Economic Statistics*, 10, 321–335.

Harris, R. D., and E. Tzavalis (1999): "Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed," *Journal of Econometrics*, 91(2), 201–226.

Haug, A. A. (1996): "Tests for Cointegration: A Monte Carlo Comparison," *Journal of Econometrics*, 71, 89–115.

Hedges, L. V., and I. Olkin (1985): *Statistical Methods for Meta-Analysis*. Academic Press, San Diego.

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(1), 85–116.

Holm, S. (1979): "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 6(1), 65–70.

Horn, R. A., and C. R. Johnson (1985): *Matrix Analysis*. Cambridge University Press, Cambridge.

Horowitz, J. L., and N. Savin (2000): "Empirically Relevant Critical Values for Hypothesis Tests: A Bootstrap Approach," *Journal of Econometrics*, 95, 375–389.

Hsiao, C. (2003): *Analysis of Panel Data*. Cambridge University Press, New York, second edn.

Im, S., M. Pesaran, and Y. Shin (2003): "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics*, 115, 53–74.

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12(2–3), 231–254.

——— (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59(6), 1551–1580.

Kao, C. (1999): "Spurious Regression and Residual Based Tests for Cointegration in Panel Data," *Journal of Econometrics*, 90(1), 1–44.

Kapoor, M., H. H. Kelejian, and I. R. Prucha (forthcoming): "Panel Data Models with Spatially Correlated Error Components," *Journal of Econometrics*.

Karlsson, S., and M. Löthgren (2000): "On the Power and Interpretation of Panel Unit Root Tests," *Economics Letters*, 66(3), 249–255.

Kelejian, H. H., and I. R. Prucha (1999): "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 40(2), 509–533.

——— (2002): "2SLS and OLS in a Spatial Autoregressive Model with Equal Spatial Weights," *Regional Science and Urban Economics*, 32(6), 691–707.

Kelejian, H. H., I. R. Prucha, and Y. Yuzefovich (2006): "Estimation Problems in Models with Spatial Weighting Matrices Which Have Blocks of Equal Elements," *Journal of Regional Science*, 46(3), 507–515.

Kiviet, J., and W. Krämer (1992): "Bias of s^2 in the Linear Regression Model with Autocorrelated Errors," *The Review of Economics and Statistics*, 74(2), 362–365.

Krämer, W. (1991): "The Asymptotic Unbiasedness of S^2 in the Linear Regression Model with AR(1)-Disturbances," *Statistical Papers*, 32(1), 71–72.

Krämer, W., and S. Berghoff (1991): "Consistency of s^2 in the Linear

Regression Model with Correlated Errors," *Empirical Economics*, 16(3), 375–377.

Krämer, W., and C. Donninger (1987): "Spatial Autocorrelation Among Errors and the Relative Efficiency of OLS in the Linear Regression Model," *Journal of the American Statistical Association*, 82(398), 577–579.

Krämer, W., and F. Marmol (2004): "The Power of Residual-Based Tests for Cointegration when Residuals are Fractionally Integrated," *Economics Letters*, 82, 63–69.

Kremers, J. J., N. R. Ericsson, and J. J. Dolado (1992): "The Power of Cointegration Tests," *Oxford Bulletin of Economics and Statistics*, 54(3), 325–348.

Künsch, H. R. (1989): "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241.

Larsson, R., J. Lyhagen, and M. Löthgren (2001): "Likelihood-Based Cointegration Tests in Heterogeneous Panels," *The Econometrics Journal*, 4(1), 109–142.

Lee, L.-F. (2004): "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models," *Econometrica*, 72(6), 1899–1925.

Lehmann, E. L., and J. P. Romano (2005): *Testing Statistical Hypotheses*. Springer, New York, 3rd edn.

Levin, A., C.-F. Lin, and C.-S. J. Chu (2002): "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties," *Journal of Econometrics*, 108(1), 1–24.

Li, H., and G. Maddala (1997): "Bootstrapping Cointegrating Regressions," *Journal of Econometrics*, 80(2), 297–318.

Lothian, J. R., and M. P. Taylor (1996): "Real Exchange Rate Behavior:

The Recent Float from the Perspective of the Past Two Centuries," *Journal of Political Economy*, 104(3), 488–509.

——— (1997): "Real Exchange Rate Behavior," *Journal of International Money and Finance*, 16(6), 945–954.

MacKinnon, J. G. (1991): "Critical Values for Cointegration Tests," in *Long-run Economic Relationships: Readings in Cointegration*, ed. by R. F. Engle, and C. W. Granger, chap. 13, pp. 267–276. Oxford University Press, Oxford.

——— (1994): "Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests," *Journal of Business & Economic Statistics*, 12(2), 167–176.

——— (1996): "Numerical Distribution Functions for Unit Root and Cointegration Tests," *Journal of Applied Econometrics*, 11(6), 601–618.

MacKinnon, J. G., A. A. Haug, and L. Michelis (1999): "Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration," *Journal of Applied Econometrics*, 14(5), 563–577.

Maddala, G., and S. Wu (1999): "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test," *Oxford Bulletin of Economics and Statistics*, 61(S1), 631–652.

Mark, N. C., M. Ogaki, and D. Sul (2005): "Dynamic Seemingly Unrelated Cointegrating Regression," *Review of Economic Studies*, 72(3), 797–820.

McCoskey, S., and C. Kao (1998): "A Residual Based Test of the Null of Cointegration in Panel Data," *Econometric Reviews*, 17(1), 57–84.

Moon, H. R., and B. Perron (2004): "Testing for a Unit Root in Panels with Dynamic Factors," *Journal of Econometrics*, 122, 81–126.

Murray, C. J., and D. H. Papell (2005): "Do Panels Help Solve the Purchasing Power Parity Puzzle?," *Journal of Business & Economic Statistics*, 23(4), 410–415.

Ng, S., and P. Perron (2001): "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, 69(6), 1519–1554.

O'Connell, P. (1998): "The Overvaluation of Purchasing Power Parity," *Journal of International Economics*, 44(1), 1–19.

Ord, K. (1975): "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70(349), 120–126.

Osterwald-Lenum, M. (1992): "A Note With Quantiles of The Asymptotic Distribution of The Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics*, 54(3), 461–472.

Papell, D. H. (1997): "Searching for Stationarity: Purchasing Power Parity Under the Current Float," *Journal of International Economics*, 43, 313–332.

——— (2002): "The Great Appreciation, the Great Depreciation, and the Purchasing Power Parity Hypothesis," *Journal of International Economics*, 57(1), 51–82.

Papell, D. H., and H. Theodoridis (2001): "The Choice of Numeraire Currency in Panel Tests of Purchasing Power Parity," *Journal of Money, Credit and Banking*, 33(3), 790–803.

Pedroni, P. (2004): "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis," *Econometric Theory*, 20(3), 597–625.

Phillips, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55(2), 277–301.

Phillips, P. C. B., and H. R. Moon (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67(5), 1057–1111.

PHILLIPS, P. C. B., AND S. OULIARIS (1990): "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58(1), 165–193.

PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for Unit Roots in Time Series Regression," *Biometrika*, 75(2), 335–46.

PHILLIPS, P. C. B., AND D. SUL (2003): "Dynamic Panel Estimation and Homogeneity Testing under Cross Section Dependence," *The Econometrics Journal*, 6, 217–259.

PHILLIPS, P. C. B., AND Z. XIAO (1998): "A Primer on Unit Root Testing," *Journal of Economic Surveys*, 12(5), 423–469.

PLOBERGER, W., W. KRÄMER, AND K. KONTRUS (1989): "A New Test for Structural Stability in the Linear Regression Model," *Journal of Econometrics*, 40, 307–318.

PSARADAKIS, Z. (2006): "Blockwise Bootstrap Testing for Stationarity," *Statistics & Probability Letters*, 76(6), 562–570.

ROGOFF, K. (1996): "The Purchasing Power Parity Puzzle," *Journal of Economic Literature*, 34(2), 647–668.

ROMANO, J. P., AND M. WOLF (2005): "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica*, 73(4), 1237–1282.

SAID, S. E., AND D. A. DICKEY (1984): "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71(3), 599–607.

SATHE, S., AND H. VINOD (1974): "Bounds on the Variance of Regression Coefficients due to Heteroscedastic or Autoregressive Errors," *Econometrica*, 42(2), 333–340.

SAVIN, N. E. (1984): "Multiple Hypothesis Testing," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. Intriligator, vol. 2, chap. 14, pp. 827–879. North-Holland Publishing, Amsterdam.

SCHWERT, G. (1989): "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business & Economic Statistics*, 7(2), 147–159.

SHIN, Y. (1994): "A Residual-Based Test of the Null of Cointegration against the Alternative of No Cointegration," *Econometric Theory*, 10(1), 91–115.

SHIRYAEV, A. N. (1996): *Probability*. Springer, Berlin, 2nd edn.

SPANOS, A. (1986): *Statistical Foundations of Econometric Modelling*. Cambridge University Press, New York.

SWENSEN, A. R. (2003): "Bootstrapping Unit Root Tests for Integrated Processes," *Journal of Time Series Analysis*, 24(1), 99–126.

TAYLOR, A., AND M. P. TAYLOR (2004): "The Purchasing Power Debate," *Journal of Economic Perspectives*, 18(4), 135–158.

TAYLOR, A. M. (2002): "A Century of Purchasing-Power Parity," *The Review of Economics and Statistics*, 84(1), 139–150.

TAYLOR, M. P., AND D. PEEL (2000): "Nonlinear Adjustment, Long-Run Equilibrium and Exchange Rate Fundamentals," *Journal of International Money and Finance*, 19, 33–53.

TAYLOR, M. P., D. PEEL, AND L. SARNO (2001): "Non-Linear Mean Reversion in Real Exchange Rates: Towards a Solution to the Purchasing Power Parity Puzzles," *International Economic Review*, 42, 1015–1042.

TAYLOR, M. P., AND L. SARNO (1998): "The Behavior of Real Exchange Rates During the Post-Bretton Woods Period," *Journal of International Economics*, 46(2), 281–312.

WATSON, G. (1955): "Serial Correlation in Regression Analysis I," *Biometrika*, 42(3/4), 327–341.

WESTERLUND, J. (2005): "New Simple Tests for Panel Cointegration," *Econometric Reviews*, 24(3), 297–316.

WHITTLE, P. (1954): "On Stationary Processes in the Plane," *Biometrika*, 41(3/4), 434–449.

Wu, Y. (1996): "Are Real Exchange Rates Stationary? Evidence from a Panel Data Test," *Journal of Money, Credit and Banking*, 28(1), 54–63.

Erklärung

Ich versichere, dass ich diese Dissertation selbständig verfasst habe. Bei der Erstellung der Arbeit habe ich mich ausschließlich der angegebenen Hilfsmittel bedient. Die Dissertation ist nicht bereits Gegenstand eines erfolgreich abgeschlossenen Promotions- oder sonstigen Prüfungsverfahrens gewesen.

Dortmund, den 21. Februar 2007

Christoph Hanck