International Journal of Conflict and Violence

The Analysis of Non-Stationary Pooled Time Series Cross-Section Data

Christoph Birkel, Federal Criminal Police Office (BKA), Wiesbaden, Germany

Vol. 8 (2) 2014

The Analysis of Non-Stationary Pooled Time Series Cross-Section Data

Christoph Birkel, Federal Criminal Police Office (BKA), Wiesbaden, Germany

It is common in macro-level research on violent crime to analyze datasets combining a cross-section (N units) with a time-series (T periods) dimension. A large body of methodological literature accumulated since the 1990s raises questions regarding the validity of conventional models for such Pooled Time Series Cross-Section (PTCS) data in the presence of non-stationarity (stochastic trends). Extant research shows that conventional techniques lead to consistent estimates only under specific conditions, and standard procedures for statistical inference do not apply. The approaches proposed in the literature to test for stochastic trends and cointegration (see the introduction to this issue) are reviewed, as well as methods for estimation and inference in the non-stationary PTCS context. A host of procedures have been developed, including methods to take cross-section dependence and/or structural breaks simultaneously into account. Thus all the tools needed for valid analyses of non-stationary PTCS data are now available, although many of them need large samples to perform well. The general approach to the analysis of non-stationary PTCS data is illustrated using a data set with robbery rates for eleven West German federal states 1971–2004. Several meaningful long-run relationships are identified and estimated.

It is quite common in macro-level research on violent crime to analyze datasets combining a cross-section (N units) with a time-series (T periods) dimension, where the number of observation periods is approximately as large as or even larger than the number of units. Examples include analyses of the effect of imprisonment rates on violent crime rates at the level of the US federal states (Vieraitis, Kovandzic, and Marvell 2007) and studies on the effect of economic inequality on murder rates among developed countries (Jacobs and Richardson 2008).[1] While such a design is, in principle, quite powerful, a large body of methodological research accumulated since the 1990s raises questions regarding the validity of conventional

models for such Pooled Time-Series Cross-Section (PTCS) data in the presence of stochastic trends.[2] At the same time, it is known from single time-series analyses that violent crime rates are, in fact, often non-stationary (for example Hale 1998). This issue is often dealt with by estimating models in first differences (for example, Entorf and Winker 2005). The cost of this approach is a loss of information regarding long-run relationships. But there are some exceptions: One example is a paper by Entorf and Spengler (2000), who analyze crime rates for eleven German federal states. The most advanced application of non-stationary PTCS methods is the recent analysis of Swedish crime rates at county level by Blomquist and Westerlund (2014).

---

1 To be sure, micro-level data-sets might also have the property that T ≥ N, so that the discussion given here would also apply to them. But more often, N is much larger than T for micro data, and it is reasonable to treat T as fixed. In this case, different analytic results apply (Bond, Nauges, and Windmeijer 2005), which are not the topic of the present paper. Suffice to say – with respect to the popular Arellano-Bond

dynamic panel models – that the first-differenced GMM estimator performs poorly under non-stationarity, which is not necessarily (but sometimes) the case for system-GMM estimators (Blundell and Bond 1998; Han and Phillips 2010; Binder, Hsiao, and Pesaran 2003). Furthermore, non-stationarity might also be an issue in the analysis of count data as well as binary and categorical variables

(and, more generally, every kind of longitudinal analysis), but this is a question far beyond the scope of this paper.

2 An overview is given in Breitung and Pesaran (2008). In what follows, the term "panel" is often used synonymously for PTCS data, as is done in the econometric literature.

The purpose of this paper is to describe the consequences of non-stationarity for conventional PTCS analyses and to give a non-technical overview of approaches to estimation and inference for non-stationary PTCS data. Since research has shown that non-stationary PTCS methods are sensitive with respect to cross-section-dependence (which might be due to common shocks, common latent factors driving the individual series, or spatial autocorrelation) and structural breaks (such as shifts in the mean of the series), these problems are also dealt with. In section 1, I review analytic results and evidence from simulation studies regarding the behavior of conventional estimators for PTCS data in the presence of non-stationarity. Next, testing for unit roots (2.) and cointegration (3.) is discussed. In section 4, approaches to the estimation of long-run relationships are presented. An example using PTCS data for the West German federal states for 1971–2004 follows (5.). A brief discussion concludes (6.). Throughout the paper it is assumed that the reader has studied the introduction to cointegration and error-correction modelling by Helmut Thome in this issue.

1. Properties of OLS and Fixed-Effect Regressions with Non-Stationary Panel Data

First, I would like to summarize some analytic results that are corroborated by simulation evidence (Entorf 1997; Kao and Chiang 2000; Chen, McCoskey, and Kao 1999; Coakley, Fuertes, and Smith 2001; Urbain and Westerlund 2011) regarding standard estimators in this situation. The behavior of panel estimators under non-stationarity depends on the cointegration properties of the variables and the degree of homogeneity of long-run relationships. Several cases have been studied:

1.1. No Cointegration

In the first scenario, there is no cointegration between the left-hand side variable and the regressors: this is the classical "spurious regression" case. In contrast to single time-series analysis, in the PTCS context, there might be a long-run relationship between the variables that can be

consistently estimated even when the residuals are non-stationary: this is what Phillips and Moon (1999) call the "long-run average regression coefficient". This result is due to the fact that the pooling of series for several units attenuates noise, which restores the consistency of standard estimators. Thus, if there is no long-run relationship between two random walks, the panel estimator will converge to zero as T and N approach infinity in sequence. This consistency property is shown by Phillips and Moon (1999, 2000) for simple OLS regressions with driftless random walks without intercepts and heterogeneous, randomly varying long-run relationships.

This finding for heterogeneous (that is, unit-specific) long-run relationships also holds for regressions with detrended data.[3] Furthermore, it also pertains to the fixed-effect (FE) estimator when applied to pure random walks (Phillips and Moon 1999, 1090). An exception is the case of cross-section dependencies in both left-hand side and right-hand side variables due to common non-stationary factors (see 2.2.1. below) that are cointegrated across units (Urbain and Westerlund 2011, 124). Here, the OLS estimator behaves as in the classical "spurious regression" case in the analysis of time series for a single unit. Additionally, for FE regressions with drifting random walks and homogeneous coefficients, results mirror those obtained in the single time-series case, that is, the coefficient obtained is a consistent estimate of the ratio of the drift parameters (Entorf 1997).

For each of the situations where "spurious regression" does not occur, the distribution of the estimator is normal, but the variance depends on the specific scenario and cannot be estimated by the usual formulae. Thus, conventional t-tests can be highly misleading, as shown by the simulations of Kao (1999), for example.

1.2. The Cointegrated Case

If a cointegration relationship exists for all units, standard OLS and FE estimators are also consistent estimators of the long-run average relationship (which is not the average of

---

[3] That is, they have been purged from a deterministic time trend by regressing them on a time index.

the cointegration parameters, if they vary across units).[4] Nonetheless, small-sample biases exist if the regressors are not strongly exogenous (in the sense that the error term of the regression is not correlated with contemporaneous or past random shocks on the independent variable), which is often the case; further biases arise if there is residual serial correlation which varies across units (Pedroni 2000; Phillips and Moon 1999; Phillips and Moon 2000; Kao and Chiang 2000). These results presume cross-section independence; if there are common stationary or non-stationary factors in the residuals, the FE estimator for homogeneous cointegration parameters is biased (Bai and Kao 2006; Bai, Kao, and Ng 2009). If there are cross-section dependencies due to common non-stationary factors in the dependent as well as the independent variables, on the other hand, the (homogeneous) cointegration parameters can be consistently estimated via OLS or FE regression (Urbain and Westerlund 2011).

The variance of the estimators depends on several features: the degree of homogeneity of (true) long-run relationships, the specification of deterministic components (unit-specific intercepts and/or time trends), and – in case of a common factor structure – the factor loadings (Phillips and Moon 1999; Kao and Chiang 2000; Baltagi, Kao, and Chiang 2000; Urbain and Westerlund 2011; Bai and Kao 2006). In any case, standard errors cannot be estimated in the usual way.

The bottom line of the research reviewed here is threefold: First, when the data are non-stationary, standard OLS or FE regressions produce consistent parameter estimates only under very specific circumstances; second, the properties of estimators are dependent on the presence or absence of cointegration; third, conventional estimates for standard errors as routinely reported by standard statistical software

do not allow valid inference under non-stationarity. Therefore, for proper inference it is necessary to ascertain if the series are non-stationary, and if so, if they are cointegrated. Depending on the results of these preliminary analyses, appropriate models have to be estimated.

2. Unit-Root Tests

2.1. Unit-Root Tests Requiring Cross-Section Independence

The first unit-root tests for PTCS data proposed in the literature presuppose that the observations for unit i are not correlated with those for unit j. If this assumption is violated, these tests often exhibit a rate of alpha errors far above the specified nominal level, as shown in Monte Carlo simulations (see below).

2.1.1. ADF-Type Tests

There are three well-known adaptations of the Augmented Dickey-Fuller (ADF) test for single time series to the PTCS context: those of Levin, Lin, and Chu (2002, in the following referred to as LL), Breitung (2000), and Im, Pesaran, and Shin (2003; IPS). All these tests are based on estimating the parameters of an equation of the form

$$(1)^5 \quad \Delta y_{it} = \alpha_{i0} + \alpha_{1i}t + \rho_i y_{it-1} + \sum_{p=1}^{P} \omega_{ip} \Delta y_{it-p} + e_{it}$$

After running (1) using OLS, the null-hypothesis $\rho_i = 0$ is tested, which implies non-stationarity of the form of a stochastic trend. Inference on ρ_i is complicated by the PTCS structure of the data. To account for this, all tests employ complicated computations involving several steps, resulting in test statistics which asymptotically obey the standard normal distribution.

Besides computational details, which I will not discuss here, the three tests differ in the formulation of the alternative hypothesis: In the tests proposed by Levin, Lin, and

---

4 This is due to the fact that the former is the *ratio* of the expectation of the long-run covariance to the expectation of the long-run variance of the regressor, while the latter is the *expectation of the ratio* of the long-run covariance to the long-run variance. In general, $E(x)/E(y) \neq E(x/y)$.

5 α_{i0} and $\alpha_{1i}t$ are optional, depending on assumptions regarding the alternative hypothesis (see section 3 of the introduction to this issue). Lagged

values of Δy_{yt} are added – if necessary – as regressors to ensure that the residuals are not serially correlated. The formulation of the test equation in first differences chosen here is equivalent to the formulation in levels presented in the introductory essay by Helmut Thome in this issue: This can be seen by rearranging the most simple form of (1), without intercept, time trend, and lagged values of Δy_{yt}. For this purpose, ρ_i in (1) is designated ρ^*_i here. Then

$\Delta y_{it} = \rho^*_i y_{it-1} + e_{it} \rightarrow y_{it} - y_{it-1} = \rho^*_i y_{it-1} + e_{it} \rightarrow y_{it} = \rho^*_i y_{it-1} + e_{it} + y_{it-1} \rightarrow y_{it} = (1 + \rho^*_i)y_{it-1} + e_{it}$. The expression in brackets might be combined to ρ_i: $(1 + \rho^*_i) = \rho_i$; then the equation reads $y_{it} = \rho_i y_{it-1} + e_{it}$, the PTCS-analogue to (3) in the introduction. When $\rho_i = 1$, then $\rho^*_i = 1 - \rho_i = 1 - 1 = 0$. Thus, testing $\rho^*_i = 0$ in the original equation with Δy_{yt} on the left side is equivalent to testing $\rho_i = 1$ in the rearranged equation.

Chu (2002) and Breitung (2000), the alternative hypothesis holds that data for *all* units follow identical stationary ($\rho_{i=1} = \rho_{i=2} = \ldots = \rho_{i=N} < 0$) or trend-stationary ($\rho_{i=1} = \rho_{i=2} = \ldots = \rho_{i=N} < 0$, $\alpha_{1i} \neq 0$ for all i) autoregressive processes. But the null hypothesis might also be wrong in other cases: for example, if *some,* but not *all* units exhibit stationarity; or if all series are stationary, but follow heterogeneous autoregressive processes. In contrast, the alternative hypothesis of Im, Pesaran, and Shin (2003) explicitly allows for heterogeneity by requiring only at least one series to be stationary and assuming that the autoregressive properties of the stationary series might vary.

2.1.2. Combination Tests

Maddala and Wu (1999) proposed to combine the p-values of individual unit-root tests into a single test statistic using meta-analytic methods. They utilize the fact that such combinations follow a well-defined distribution. Specifically, the following test statistic is computed:

$$(2), \quad \lambda = -2 \sum_{i=1}^{N} \ln(\pi_i)$$

where π_i is the significance level of a unit-root test for unit i. Any unit-root test for single time-series might be used for this procedure. For fixed N, λ follows asymptotically (as $T \to \infty$) a χ^2-distribution with 2N degrees of freedom. If the number of cross-sections is large, it is advisable to use modified test statistics (called P_m and Z) developed by Choi (2001), which are valid under an asymptotic theory which assumes that N also approaches infinity (but slower than T). The exact formulation of the null and the alternative hypothesis depends on the unit-root test chosen, but generally, the significance of the test statistic implies that at least one unit is stationary.

2.1.3. Testing the Null of Stationarity

There might be situations in which it is attractive to view stationarity or trend-stationarity as the null hypothesis. Hadri (2000) proposes such a test. Here, the series are

decomposed into deterministic components, a random walk, and a stationary white noise error term.[6] When the series are stationary, the variance of the non-stationary component σ_u will be zero. Therefore, the ratio of the variances of the random-walk and stationary white noise components (σ_u and σ_e) will also be zero – this is the null hypothesis of Hadris tests, implying that the data for all units are generated by a stationary or trend-stationary process. The alternative hypothesis holds that the ratio is greater than zero, meaning that all units exhibit unit-root processes. Hadri proposes two test statistics that are constructed using estimates of σ^2_u and σ^2_e, one under the assumption of homogeneous variances and one for the more realistic case of heterogeneous variances. These are based on regressions of the series on an intercept or an intercept and a time trend. For the case of serially correlated errors in these regressions, Hadri suggests using non-parametric heteroscedasticity- and autocorrelation-consistent variance estimators.

2.2. Panel Unit-Root Tests for PTCS-Data Exhibiting Cross-Sectional Correlation

Cross-section dependence affects the size of panel unit-root tests (see below).[7] Subtracting the period-specific mean from the data before applying a unit-root test removes cross-section correlation only if it is due to one common component (such as a common shock) with exactly identical influence on all units, which is not very plausible in most cases. Thus, panel unit-root tests have to be modified to take cross-section dependence explicitly into account – these are the so-called "second generation unit-root tests" (Hurlin and Mignon 2004).

2.2.1. The Common Factor Approach

One approach assumes that a common factor structure is the source of cross-section dependence: Here, the data are assumed to be generated by the following process:

$$(3), \quad y_{it} = \alpha_i + \beta_i t + \boldsymbol{\lambda}_i' \mathbf{F}_t + e_{it}$$

---

6 With "white noise" I refer to a series free of autocorrelation; in other words the observation for period t is not correlated with prior observations.

7 With "size" I refer to the rate of alpha errors of a test. A test is said to have a correct size if the rate of alpha errors corresponds to the nominal significance level chosen. In other words the test should wrongly reject the null hypothesis in at most 5 percent of

cases if a significance level of 5 percent is chosen. Otherwise, I speak of "size distortions." If the actual rate of alpha errors is higher, for example, a test is said to be "oversized."

where $\mathbf{F_t}$ is a vector of common factors and $\boldsymbol{\lambda}_i$ a vector of associated factor loadings; such a common factor might be a national trend driving regional rates of violent crime, for example.[8] Thus, the data consist of deterministic components (intercept and time trend), common factors, and an idiosyncratic part e_{it}. The common factors might be stationary or non-stationary.

The most simple approach for this set-up are extensions of Maddala-Wu-type tests (designated $C\tilde{P}$ and $C\tilde{Z}$) as well as the IPS test (CIPS and CIPS*), developed by M. Hashem Pesaran, where it is assumed that there is one common stationary factor. In a first-step "cross-sectionally augmented Dickey-Fuller-Test" (CADF) this common factor is approximated by the lagged period-specific cross-section mean (Pesaran 2007). There is also a proposal to extend the approach to the case of several common stationary factors (Pesaran, Smith, and Yamagata 2008). Here, lagged period-specific cross-section means of other non-stationary variables which contain the same common factors as the variable of interest are also entered in the test equation.

A more explicit modelling of one or more stationary or non-stationary common factors has been proposed by Jushan Bai and Serena Ng in their "Panel Analysis of Non-stationarity in Idiosyncratic and Common Components" (PANIC) approach (Bai and Ng 2004). The procedure of Moon and Perron (2004) is very similar. Here, the common factors and the factor loadings are estimated. These estimates are then used for different purposes: Bai and Ng develop methods for separately testing the estimated common factors and idiosyncratic parts for unit-roots, while Moon and Perron are interested only in the behavior of the individual-specific component, and therefore apply a unit-root test to the defactored data. Thus, only the approach of Bai and Ng is able to detect non-stationarity if it is due to common factors; the tests of Moon and Perron as well as those of Pesaran will wrongly reject the hypothesis of a unit-root here, because the non-stationary common factor is removed before applying the test.

2.2.2. The Bootstrapping Approach

A more general alternative which does not presume a specific source of cross-section dependence is to conduct statistical inference based on empirical critical values obtained via bootstrapping. Chang (2004), for example, proposes a resampling scheme that preserves the cross-section correlation structure as well as the autoregressive properties of the residuals. Chang's approach is based on the assumption that the cross-section correlation is due to spatial dependencies in stationary components of the series.

Palm, Smeekes, and Urbain (2011), in contrast, develop a very general bootstrapping approach for simplified versions of the LL and IPS tests, which assumes a data-generating process as in (3), where the factors might be stationary or have a unit-root. Also, the approach also applies if there are (non-contemporaneous) dynamic dependencies between the idiosyncratic components e_{it}, if there is a correlation between e_{it} and e_{jt-1}, for example.

2.3. Structural Breaks

As in single time-series analysis, the power of panel unit-root tests is reduced by breaks (Im, Lee, and Tieslau 2005; Sethrapramote 2004).

There are several proposals for panel unit-root tests that take structural shifts into account.[9] Im, Lee, and Tieslau (2005), for example, assume a break in the form of a one-time shift of the mean, possibly at a different date for each unit. The advantage of the proposed test is that its distribution is invariant to the break date. The authors also consider a procedure to identify the break dates if they are not known – with the drawback that the distribution of the test statistic is then no longer invariant to the break date (Sethrapramote 2004, 42). In reaction to the latter problem, several authors have modified this approach (Tam 2006; Westerlund 2006a), suggesting alternative procedures for the determination of the break dates and bootstrapping procedures for the case of cross-section correlation. Tam also considers the case of a possible shift of the time trend

---

8 Blomquist and Westerlund (2014) report evidence for such a nationwide trend in Swedish county-level rates of property crime.

9 Besides the contributions reviewed here, see also Murray and Papell (2000); Jönsson (2005); Breitung

and Candelon (2005); Harris, Leybourne, and McCabe (2005).

(if the data are assumed to have a time trend under the alternative hypothesis).

There are also suggestions for testing the null hypothesis of trend stationarity, possibly with a break: The approaches of Carrion-i-Silvestre, Barrio-Castro, and López-Bazo (2005) and Hadri and Rao (2008) are extensions of Hadri's stationarity test, augmenting the test equation with break-dummies to account for shifts in the mean and/or the time trend. In both papers, analytic results allowing the computation of the asymptotic expectations and variances of the test statistics – which are dependent on the break date here – are provided, so that it is possible to implement an appropriate test. If the break points are not known, they are determined empirically using methods similar to those suggested by Tam. Furthermore, Hadri and Rao propose a modified test based on the sum of two test statistics computed before and after the break, the distribution of which is not dependent on the breaks. Here it is assumed that the break consists of a mean shift only or a combined shift of the mean and the trend slope. For the case of cross-section dependence, the authors suggest bootstrapping procedures.

2.4. Small Sample Properties

Numerous simulation studies on the finite sample behavior of panel unit-root tests have been published (see the papers cited above and Hlouskova and Wagner 2006; Banerjee, Marcellino, and Osbat n.d.; O'Connell 1998; Maddala and Wu 1999; Gengenbach, Palm, and Urbain 2004; Gutierrez n.d.; Baltagi, Bresson, and Pirotte 2007; Sethrapramote 2004; Westerlund and Breitung 2013), but their results are, due to the variety of setups used, difficult to compare. Nonetheless it emerges that the size of panel unit-root tests is sensitive to first-order moving-average errors,[10] and tends to be distorted if the number of cross-sections is large compared to the time dimension. For small samples, for example with 10 cross-sections and 25 periods, power is generally modest, especially in the specification with deterministic trend. Among the first generation tests, the LL test

and Breitung's test often, but by no means uniformly, perform best – especially if the autoregressive behavior of the data is homogeneous across units, as assumed under the alternative hypothesis of these tests. If the latter is not the case, the IPS and Maddala-Wu tests often outperform LL – which might, on the other hand, have more power if applied to nearly non-stationary data (i.e. if ρ_i in (1) is very close to 0) (Westerlund and Breitung 2013). According to the findings of Hlouskova and Wagner (2006), furthermore, Hadri's stationarity test is badly oversized as soon as the residuals are not totally free from serial correlation.

It is interesting to know how the first generation tests behave under cross-section dependency. This seems to depend on the specific strength and type of contemporaneous correlation: O'Connell (1998) reports large upward size distortions for the LL test: for average cross-section correlations of 0.9, he often finds rejection rates of more than 50 percent at a nominal level of 5 percent, especially if the number of cross-sections is large. Similarly, Banerjee, Marcellino, and Osbat (n.d.). find considerable size distortions if there is cointegration across units, that is, if y_{it} and y_{jt} are cointegrated (which is the case if there is a non-stationary common factor while the idiosyncratic part in (3), e_{it}, is stationary).

On the other hand, Hlouskova and Wagner (2006) as well as Maddala and Wu (1999) find only small increases in size for moderate cross-section correlations (of up to 0.6 in the case of Hlouskova and Wagner). In this situation, the relative ranking of the tests does not change. Baltagi, Bresson, and Pirotte (2007), who study the behavior of several tests for three models of spatial autocorrelation (in none of which common factors appear), report considerable size distortions (rejection rates up to 20 percent) only for strong spatial autoregressive processes, while they are weak for spatial moving-average and – especially – spatial error-component models. The performance of the conventional tests studied – LL, Breitung, IPS, Maddala-Wu

---

10 A first order moving-average-process is a form of autocorrelation where the observed value at time t is affected by the random shock on the series at time t-1.

and Choi – is very similar. Thus, it seems that cross-section dependence does not universally affect the size of panel unit-root tests.

Among second generation tests, unit-root tests assuming a factor structure sometimes show size distortions. This is the case for the CIPS test and Bai and Ng's approach if the time dimension is small, while bootstrapping procedures minimize them. Moon and Perron's tests perform best in terms of power, but tests assuming common factors generally need large samples to achieve satisfactory power. This finding applies also to bootstrapping procedures and approaches that take breaks into account.

Thus, the main conclusion to be drawn from the simulation evidence is that the performance of panel unit-root tests is moderate for the relatively small data sets which are employed often in comparative social research, especially if the data are subject to cross-section dependence and/or breaks.

2.5. Issues in the Application of Panel Unit Root Tests

When applying panel unit-root tests, one has to decide on the inclusion of an intercept and/or a time trend in the test equation. Furthermore, one has to select the number of lagged differences (p_i in (1)) to be included as regressors if a parametric correction for serial correlation is used. Regarding the first issue, the considerations pointed out in the contribution by Helmut Thome in this issue apply. With respect to the selection of p_i, it is common to use a general-to-specific-approach. Here, the test equation is computed sequentially, starting with a maximum p_i, p_{max}, which is chosen depending on T, and to step-wise reduce p_i until the t-Test for coefficient of the highest lag is significant. For the determination of p_{max}, the following formula performs well (Hayashi 2000, 594f.):

$$(4) \quad p_{max} = \text{int}\left[12\left(\frac{T}{100}\right)^{1/4}\right]$$

Alternatively, information criteria can be used. Their drawback is that they tend to select a too small p_i, especially the Schwarz information criterion (SIC), if there is a specific form of serial correlation in the errors (negative moving-average processes) (Im, Pesaran, and Shin 2003, 68). Ng

and Perron suggested modified information criteria which perform better in selecting p_i (Ng and Perron 2001).

2.6. Remarks Regarding the Choice of a Test and the Interpretation of Unit-Root Tests

Which of the conventional unit-root tests reviewed here should be chosen by the applied researcher? The answer is, first, dependent on the asymptotic theory deemed to be plausible: If the cross-sections can be viewed as a finite universe, the fixed-N, T→∞ asymptotics of Maddala and Wu are appropriate; generally, also the sequential limit theories of the other tests are valid in fixed-N situations, but this does not apply vice versa. Second, one should check if it is plausible to assume homogeneous autocorrelation properties across units under the alternative hypothesis: if this is the case, either the LL or Breitung's test might be applied; otherwise, IPS and the Maddala/Wu/Choi tests might be considered. To my view, it is difficult to imagine situations where the units exert literally identical dynamic behavior, as assumed by the former tests. Nonetheless, in view of the simulation results reviewed above, there might be situations where the tests of LL or Breitung are an option: these seem to outperform IPS in terms of power if the series exhibit strong autocorrelation. Regarding the stationarity test of Hadri, the simulation studies show that it tends to exhibit strongly inflated rates of alpha errors as soon as there is some autocorrelation in the data, making it useless in practical applications.

But which test should one choose if cross-section correlation is present? Here, the nature of cross-section correlation is crucial: if it is plausible to assume that it is due to common factors, factor analytic methods might be applied. Among these, the PANIC approach is conceptually most convincing, because both common as well as idiosyncratic components are tested. The drawback is that it needs large samples (especially large T) to perform well. For moderately-sized samples with a not too small time dimension (T ≥ 20, say), CIPS* might be considered, although the presumption of one stationary common factor is somewhat restrictive. If the common factor model is not plausible, bootstrapping might be considered, although one has to be willing to assume that the source of cross-section correlation is stationary when using Chang's test, while the

more general test of Palm and others has yet to be fully developed. The bootstrapping approach might also be chosen if the sample size is small, because – according to simulation studies – tests based on the common factor approach tend to over-reject in small samples.

Finally, in analyses of PTCS data, breaks in the series might be an issue. In PTCS analyses of crime statistics (crime rates, imprisonment rates, etc.) at the sub-national level, break dates are known in most cases, because they are due to changes in legislation and/or registration procedures. Here, the original approach of Im and coauthors might be a choice. In cross-national research, however, breaks at an unknown date are a realistic possibility, because it is difficult to gather comprehensive information on changes in legislation and recording procedures for every country in the sample. Thus, the procedures suggested by Tam and Westerlund might be considered, which also allow cross-section correlation to be taken into account.

I would like to close the section on unit-root tests with a caveat: Much care is needed in interpreting the results of panel unit-root tests, because the rejection of the null hypothesis suggests only that *at least one panel member* is stationary. But the test does not tell us for *how many units* the series are stationary. It would be important to know if, in fact, *all* series are stationary, or if there is a mixture of stationary and non-stationary processes. One way to check this would be to test also the null hypothesis that all units are stationary. But, as pointed out above, the only widely available panel stationarity test by Hadri (2000) performs extremely badly. So this is only a theoretical possibility. There are proposals for testing fractions of the units sequentially (Smeekes 2010) and approaches to unit-by-unit testing which control the size of the test (which would otherwise be inflated due to multiple testing) (Moon and Perron 2012; Hanck 2009). But, these procedures do not perform well in identifying the stationary units in PTCS data as long as T is not large (< 100)

(Smeekes 2010). Thus, their usefulness is questionable for many situations. It remains only to exert caution if there are indications that non-stationarity is an issue for at least a part of the units, even if a formal test rejects the unit-root-hypothesis.

3. Testing for Cointegration
3.1. Testing for Cointegration in the Absence of Cross-Section Dependence and Breaks

The cointegration tests proposed by Pedroni (1999, 2004) are widely used.[11] These tests are based on the residuals of a regression of the following form:

$$(5) \quad y_{it} = \alpha_i + \delta_i t + \beta_{1i} x_{1it} + \beta_{2i} x_{2it} + \ldots + \beta_{Mi} x_{Mit} + e_{it}$$

If there is cointegration between y and x, these residuals (e_{it}) should be stationary; thus, testing e_{it} for a unit-root amounts to testing for cointegration. Generally, Pedroni's tests allow the cointegration coefficients and the variances of the series to be heterogeneous and the regressors to be endogenous, although they should not be cointegrated.

He considers two basic approaches for the construction of panel cointegration tests: For the first it is assumed that the residuals of the individual static cointegration regressions follow identical autoregressive processes under the alternative hypothesis. These "panel" statistics are constructed based on pooled regressions. The second approach allows for heterogeneous serial correlation properties, therefore, the "group" tests are based on averages of individual test statistics. For each of these two types, panel cointegration tests based on the ADF test and the semi-parametric cointegration tests studied by Phillips and Ouliaris (1990) are developed.

Westerlund (2007) argued that there might be gains in power if the cointegration test is carried out as a test on the so-called error correction parameters γ_i in the following panel-error correction model:

---

11 In some situations it might be of interest to test the null hypothesis that two series are cointegrated; tests of this kind are not reviewed here; see McCoskey and Kao (1998) for such a test.

$$(6) \quad \Delta \mathbf{y}_{it} = \alpha_{1i} + \alpha_{2i}t + \gamma_i(\mathbf{y}_{i,t-1} - \vartheta_i'\mathbf{X}_{it}) + \sum_{j=1}^{p_i} \beta_{ij}\Delta \mathbf{y}_{i,t-j} + \sum_{j=0}^{p_i} \delta_{ij}'\Delta \mathbf{X}_{i,t-j} + \varepsilon_{it}$$

Here, θ_i is a vector with unit-specific cointegration parameters. In the case of cointegration, the γ_i have to be negative ($\gamma_i < 0$). Thus, testing the hypothesis $\gamma_i = 0$ amounts to a cointegration test. The advantage of this approach is that the restriction of the long- and short-run dynamics to be identical, which is implicit in residual-based cointegration tests, can be avoided. He develops two tests: a test on the error correction parameter, and one based on the product of the error correction term and the number of time periods. Both come in a "panel" (P_ϑ, P_α) and a "group" (G_ϑ, G_α) version, so that four different tests result. A drawback of these tests is that strict exogeneity is required, which might be relaxed to weak exogeneity by adding leads of the first differences of the regressors.[12]

3.2. Dealing with Cross-Section Dependence and Breaks

The tests reviewed so far assume cross-section independence, an assumption which might be violated in many cases. Structural breaks also affect the performance of cointegration tests. According to simulations, at least Pedroni's t-tests lose power if there are breaks (Banerjee and Carrion-i-Silvestre 2006; Gutierrez 2005). Many suggestions for dealing with cross-section dependence also consider the issue of breaks.[13] The types of breaks studied are level shifts (implying a change of the intercept in estimation equations), changes of the time trend (change of the coefficient for a time index), and changes of the cointegration relationship (change of the cointegration parameter). None of the approaches suggested in the literature considers all possible combinations of these types of breaks, but they are all tailored to specific situations.

3.2.1. Allowing for Breaks in Absence of Cross-Section Dependence

To deal with breaks in the mean and/or the cointegration parameters in absence of cross-section correlation, there are two proposals to extend the approach of Gregory and Hansen (1996) for single time-series: Westerlund (2006b) constructs four test statistics which are cross-sectional sums of Gregory/Hansen-type test statistics for a mean-

shift. Gutierrez (2005), in contrast, bases his tests on standardized sums of the *p-values* for Gregory/Hansen tests.

3.2.2. Testing for Cointegration in the Presence of Breaks and Cross-Section Dependence

Di Iorio and Fachin (2007) also adapt the Gregory/Hansen approach to the PTCS context in a test that uses the mean or the median of individual test statistics. They suggest computing critical values using a bootstrapping procedure, which also accounts for cross-section dependence.

Banerjee and Carrion-i-Silvestre (2006) consider every possible combination of single breaks in the mean, the slope, and the time trend of the cointegration relation, except a change restricted to the time trend and a simultaneous break in all three parameters. Specifically, they adapt two of Pedroni's parametric test statistics in order to take the breaks into account. For the case of cross-section dependence, Banerjee and Carrion-i-Silvestre use Bai and Ng's PANIC approach. Here, it is necessary to assume a common break date for all units; if it is not known, it has to be estimated. A similar test, which also follows Bai and Ng, has been proposed by Westerlund and Edgerton (2008), who also develop a procedure for the determination of the date of the break. But their test presumes that the common factors are stationary, and allows only for shifts in the intercept or in the intercept and the cointegration parameter.

3.3. Small Sample Behavior

The results from various Monte-Carlo studies (Banerjee, Marcellino, and Osbat 2004; Banerjee and Carrion-i-Silvestre 2006; Gengenbach, Palm, and Urbain 2005; Gutierrez 2003; Gutierrez 2005; Örsal 2007; Pedroni 2004; Wagner and Hlouskova 2007; Westerlund 2006b; Westerlund 2007; Westerlund and Basher 2008) are anything but clear-cut, but some general observations can be made: First, most tests show size distortions in the presence of residual serial correlation, especially Pedroni's parametric tests. One of Westerlund's tests on the error correction parameters, P_α, rejects too often when the regressors are endo-

---

12 For the concept of weak exogeneity see Enders (2004, 368).

13 One exception is the bootstrapping procedure developed by Westerlund (2007) to be applied with

his error-correction tests in the case of cross-section correlation.

genous. The effect of cross-section dependence on size seems to be moderate in most cases. With respect to power, Pedroni's parametric statistics often perform best, but even for these, power is often quite low for small T (< 25), sometimes of the same magnitude as size. Similar results apply to other tests which I do not present here. Cointegration tests that accommodate breaks are generally well-sized, but need large T ($= 100$ or even 200) to achieve reasonable power, with the exception of Di Iorio and Fachin (2007).

3.4. Practical Considerations

When implementing residual-based cointegration tests, one has to specify the deterministic components (intercepts and time trends) of the static regression equation (5) as well as the number of lagged differences of the residuals to include in the ADF-type equation for testing the residuals for a unit-root. Here, similar considerations apply as in the case of unit-root tests. When serial correlation is accounted for non-parametrically, one has to choose a lag length for the band width of the kernel estimator; Pedroni suggests to do this in dependence on T according to the formula $\mathrm{int}[K=4(T/100)^{2/9}]$ (Pedroni 2004, 608).

3.5. Remarks Regarding the Choice of Test and the Interpretation of Cointegration Tests

How should one proceed in applied research, especially macro-level criminological research? Here, the difficulty arises that the time dimension of the data sets at hand – often between 20 and 50 periods – lies in the region where, according to the simulation results reviewed above, the power of cointegration tests is small in most cases. Thus, there is a real risk to miss substantively interesting long-run relationships. On the other hand, the implications of wrongly rejecting the null hypothesis are less serious than in the case of unit-root testing: the long-run parameter estimated in the next step will not reach significance if the model estimated is correctly specified, leading to the correct conclusion that the long-run effect is nil. One might

therefore consider being more liberal with respect to alpha errors in cointegration testing than in unit-root testing, and putting more weight on power properties. Therefore, Pedroni's parametric tests, which often perform best in terms of power (but less so in terms of size), might be a choice if there are no indications of strong cross-section dependence (recall that the effects of cross-section correlation on cointegration tests are generally moderate). If the latter is the case, one might apply the error-correction tests (but not P_a, in view of the simulation results) with the bootstrapping procedure proposed by Westerlund. If there are breaks in the series, there is no satisfying solution yet, because appropriate methods need larger samples than available in most cases to achieve power. Besides that, none of the procedures mentioned above is implemented in standard statistical software. An ad-hoc approach might be to first adjust the series to the breaks (by regressing them on appropriate dummy-variables, for example) and then to apply conventional tests, although the results will only be roughly indicative then.

Regarding the interpretation of cointegration tests, finally, a caveat analogous to that with respect to panel unit-root tests applies here: The rejection of the hypothesis of no cointegration does *not* imply that *all* units are cointegrated. This has to be kept in mind in view of the fact that estimators for cointegration parameters assume that, indeed, cointegration holds for all units.

4. Estimating Cointegration Relationships in Non-Stationary PTCS Data
4.1. Fully Modified OLS and Dynamic OLS

First, there are several approaches to "fully modifying" OLS (FM-OLS) for cointegrated panels by using non-parametric corrections to (5) (Kao and Chiang 2000; Chiang, Kao, and Lo 2007; Pedroni 2000) for serial correlation and endogeneity.[14] They differ in the degree of homogeneity of variances and serial correlation properties assumed, while they presume generally homogeneous cointegration parameters. For each of them, analogues based on adding

---

14 In the case of cointegration it is also possible that several cointegration relationships ("cointegration vectors") exist if the estimation equation has more than one variable on its right-hand side. But since system approaches allowing the estimation of

multiple cointegration vectors need huge data sets to perform well, I only cover the case of a single cointegration relationship. The interested reader is referred to the discussion in Wagner and Hlouskova (2007) and the references given there. I also do not

treat the estimation of the "long-run average relationship" in the absence of cointegration, an issue that has received little attention so far (but see Sun 2004, Fuertes 2008, and Trapani 2012).

leads and lags of the first differences of the regressors to (5) have been proposed. This so-called Dynamic OLS approach (DOLS) is asymptotically equivalent to FM-OLS (Kao and Chiang 2000).

All these estimators converge to a normal distribution, the variance of which depends on the deterministic specification of the equation and the long-run covariance. To conduct statistical inference, kernel-density estimates of the long-run variance-covariance matrix are needed.

4.2. Error-Correction Models

An alternative way to deal with serial correlation is to model short-run dynamics explicitly using an error-correction model like (6). The disadvantage of this strategy in the panel context, however, is that it will be subject to the so-called Nickell bias due to the presence of the lagged dependent variable on the right-hand side of the equation if individual-specific intercepts are allowed for. But this bias usually vanishes fast with an increasing time-dimension (Judson and Owen 1999). Therefore, the error-correction model might nonetheless be useful for panels with a large enough time dimension. This approach has been favored by Pesaran, Shin, and Smith (1999), who consider three variants of the error correction model, varying in the degree of homogeneity assumed for long- and short-run parameters:

The "Dynamic Fixed Effects" (DFE) model is intended for situations in which it is reasonable to assume that short- and long-run parameters are identical for all cross-sections. Therefore, the following equation is estimated:[15]

$$(8), \quad \Delta\mathbf{y}_i = \varphi\,\mathbf{y}_{i,-1} + \mathbf{X}_i\boldsymbol{\beta} + \sum_{j=1}^{p-1} \lambda_j^\star \Delta\mathbf{y}_{i,-j} + \sum_{j=0}^{q-1} \Delta\mathbf{X}_{i,-j}\boldsymbol{\delta}_j^\star + \alpha_i\mathbf{d}_i + \zeta_i$$

where $\mathbf{X}_i$ is a matrix with the observations of the independent variables and $\mathbf{d}_i$ a matrix containing a unit-specific intercept and – if specified – a time trend. The cointegration parameters $\boldsymbol{\theta}$ can be computed as

$$\frac{-\boldsymbol{\beta}}{\varphi} = \vartheta.$$

If the cointegration coefficients are, in fact, heterogeneous, DFE will produce an inconsistent estimate of the average cointegration parameter.

For the case that it is plausible that the cointegration parameters $\boldsymbol{\theta}$ are homogeneous, but the short-run dynamics differ across units, Pesaran et al. propose the "Pooled Mean Group" (PMG) estimator (note the two subscripts of the $\boldsymbol{\lambda}^\star$- and $\boldsymbol{\delta}^\star$-parameter vectors):

$$(9) \quad \Delta\mathbf{y}_i = \varphi_i\mathbf{y}_{i,-1} + \mathbf{X}_i\boldsymbol{\beta}_i + \sum_{j=1}^{p-1} \lambda_{ij}^\star\Delta\mathbf{y}_{i,-1} + \sum_{j=0}^{q-1} \Delta\mathbf{X}_{i,-j}\boldsymbol{\delta}_{ij}^\star + \alpha_i\mathbf{d}_i + \zeta_i$$

with the restriction

$$\frac{-\boldsymbol{\beta}_i}{\varphi_i} = \vartheta_i = \vartheta, i = 1,2,\dots,N.$$

For the estimation of (9), a maximum likelihood algorithm is described by Pesaran, Shin, and Smith.

In the Mean Group (MG) estimator the assumption of homogeneous cointegration parameters is also relaxed. Here, an estimate for the average cointegration parameter θ_{MG} is computed as the arithmetic mean of the individual θ_i after running (9) for each unit via OLS without restrictions on $\boldsymbol{\beta}_i$ and φ_i. A consistent estimate of the variance of θ_{MG} can be obtained as follows (Pesaran, Smith, and Im Kyong So 1996, 157):

$$(10) \quad V_{\vartheta MG} = \frac{1}{N(N-1)} \sum_{i=0}^{N} (\hat\vartheta_i - \hat\vartheta_{MG})^2$$

Pesaran et al. suggest choosing among the three estimators based on a Hausman test of the null hypothesis that the difference between the MG and the PMG (or DFE) estimator is zero. If this hypothesis is rejected, the MG estimator is appropriate; otherwise, the more efficient pooled estimator is to be preferred.

---

15 The equation contains the so-called "Bårdsen transformation" of the error-correction model (6).

4.3. Estimation of Cointegration Parameters under Cross-Section Dependence

Bai and Kao suggest adapting the FM-OLS estimator to the case when there is cross-section correlation due to common stationary factors (Bai and Kao 2006). Here it is assumed that the cointegration parameters are the same for all units. In this approach, principal component analysis is used to extract the common factors from the residuals of a first-step OLS regression and to estimate the factor loadings. These are used to construct correction terms in a modified estimation formula, which is used to produce second-round parameters. The residuals of this second estimation step are used to construct new correction terms, and so on. This procedure is continued until convergence is achieved, resulting in the "continuously-updated and fully-modified estimator" (CUP-FM). In a later paper, the approach is extended to the case of common non-stationary factors. Furthermore, Bai, Kao, and Ng also consider an estimator where the correction is applied only in the last iteration, the so-called "continuously-updated and bias-corrected" estimator (CupBC) (Bai, Kao, and Ng 2009). Their approach is also valid when there is a mixture of stationary and non-stationary common factors, or a mixture of stationary and non-stationary regressors.

The computation of the "continuously-updated" estimators of Bai, Kao, and Ng is fairly involved. A simpler approach has been suggested by Pesaran and others, who transfer the logic of the "cross-sectionally augmented" unit-root tests to the estimation of cointegration regressions (Kapetanios, Pesaran, and Yamagata 2011): Common factors are simply accounted for by adding the cross-section averages of the dependent variable and the regressors to the left-hand side variables of the estimation equation. They consider static "common correlated effects" mean group estimators,[16] called CCEMG, which allow the cointegration parameters to vary across units, as well as a static pooled FE estimator, where the cointegration coefficients (but not the parameters for the cross-section averages) are assumed to be homogeneous, which they designate CCEP (Common Correlated Effects Pooled). Hypothesis tests on the average cointegration parameter can be conducted analogous to tests on MG. The estimation of the variance of coefficients estimated using CCEP, however, is more complicated and requires the computation of the CCEMG estimator.[17] The CCE estimators are consistent under various types of cross-section dependence, including single or multiple common factors, which might be stationary or non-stationary, and even when the idiosyncratic errors are cross-sectionally correlated (Kapetanios, Pesaran, and Yamagata 2011; Pesaran and Tosetti 2011).

4.4. Small Sample Properties

From several simulation studies (Breitung 2005; Pedroni 2000; Kao and Chiang 2000; Wagner and Hlouskova 2007; Bai and Kao 2006; Eberhardt and Bond 2009; Kapetanios, Pesaran, and Yamagata 2011), the following picture emerges: FM-OLS does not work very well in the reduction of bias, especially with respect to t-tests. DOLS performs better than FM-OLS, but even here estimates are biased and significance tests oversized in very small samples. This is especially a feature of an estimator proposed by Kao and Chiang (2000) which allows for heterogeneous variances and dynamics. Correlation of the residuals between cross-sections induces small increases in the bias of estimates and size distortions in tests for significance – but these are negative for DOLS. Cross-unit cointegration makes the biases of conventional estimates a bit worse. Unfortunately, there is no simulation evidence on the performance of the error correction models discussed above (MG, DFE, PMG).

Among the estimators for cointegrated PTCS data with cross-section correlation, the CUP-FM and especially the CupBC estimators perform well if there is one common factor (stationary or non-stationary), but there are large size distortions in significance tests if there are several common factors. The CCE estimators, on the other hand, seem to have good properties, although large samples (~ N = T = 100) are needed to achieve good power for significance tests.

---

16 With "static" I refer to estimation equations which contain no lagged values of the dependent variable or the regressors (as error-correction models do, for example) to model dynamic relationships.

17 The appropriate formula can be found in Kapetanios, Pesaran, and Yamagata (2011, 330f.).

4.5. Comments on the Estimation of Long-Run Relationships in Applied Research

If cross-section residual correlation is not a concern, and cross-unit cointegration is not plausible, the choice of the estimator depends on the degree of cross-section homogeneity with respect to parameters, residual variances, and (in case of static estimators) residual serial correlation one is willing to assume. If homogeneity of parameters is a reasonable assumption, DOLS is an option (although not in the variant due to Kao mentioned above), because it outperforms FM-OLS and avoids possible problems due to the Nickell bias which might occur with ECM estimators (such as DFE) with small T. The latter risk might be outweighed by the higher costs (in terms of bias) arising from wrongly imposing homogeneous parameters. Thus, if parameter heterogeneity is to be expected, one might turn to the MG or PMG estimators. The tenability of assumptions regarding parameter homogeneity can be tested within the ECM approach.[18] If cross-section dependence is an issue, one should turn to the CCE approach, which can be implemented in Stata and R (see below). The drawback of this method is the low power of significance tests when the CCE approach is applied to

data sets of the size usually encountered in macro-level criminological research. One would have to accept this, due to the costs of ignoring cross-section dependence. Besides that, one should keep in mind that in this way, one can only estimate relationships within the idiosyncratic part of the data (that is, the portion which is not driven by common factors); thus, if a parameter turns out to be non-significant, this does not preclude the possibility that there are long-run relationships between factors common to all units.

### 5. An Example: Robbery Rates in the West German Federal States

To illustrate the general approach to the analysis of non-stationary PTCS data described here, I use a data set with (completed) robbery rates for the eleven West German federal states for 1971–2004.[19] It includes also data for several plausible explanatory variables, namely per capita real disposable income, per capita consumption, and the clearance rate for robbery. It also contains the percentage of inhabitants aged 65 or older, as well as the percentage of foreigners (in the legal sense); these latter two variables serve as controls for changes in the demographic composition of the population.[20]

---

18 Although there will be cases where it is not possible to conduct the appropriate Hausman test, because there is no guarantee that the matrix in the denominator of the test statistic will be – as required – positive-definite.

19 Data for the five federal states on the territory of the former GDR are not used, because robbery rates for them are only available from 1993.

20 For data sources and the motivation for the selection of the variables, see Birkel (2015).

Figure 1: Robbery rates for the West German Federal States, 1971–2004
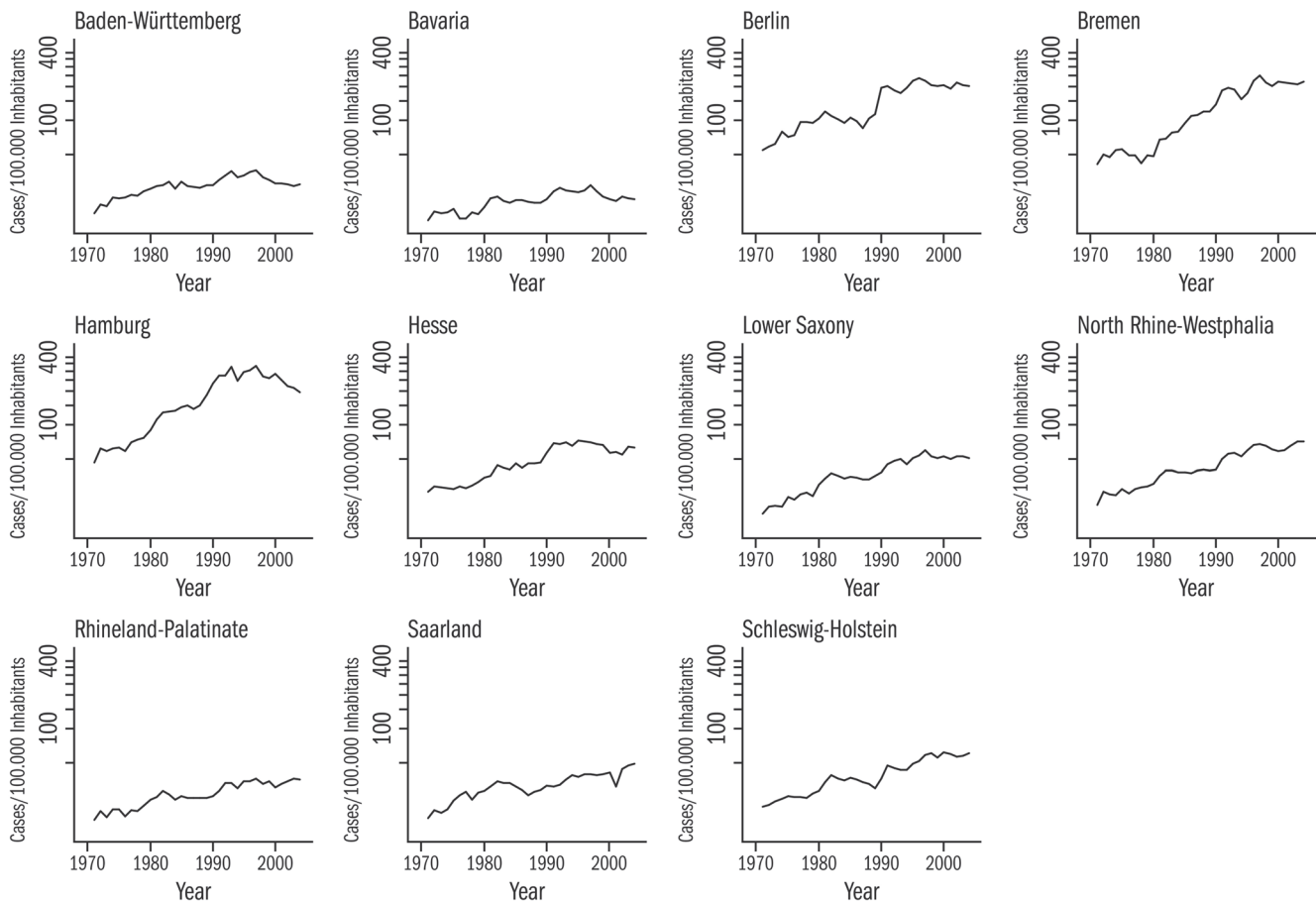
Source: Federal Criminal Police Office, Criminal Police Offices of the Federal States.

5.1. Unit-Root Tests

A panel of the robbery rates is shown in Figure 1 (note the logarithmic scale of the y-axes). The rates for most states exhibit a clear upward tendency until around 2000. Thus, the robbery rates may contain a unit-root. The graphs for the other variables – which are not shown here – suggest also stochastic trends. Therefore, the hypothesis of a unit-root was tested formally, using the IPS test as well as the

Maddala-Wu Test based on individual ADF tests. These were conducted with unit-specific intercepts and a unit-specific time trend in the test equation, because at least for some federal states stationarity around a deterministic trend seemed to be a plausible alternative hypothesis. For lag-length selection, the modified AIC (MAIC) of Ng and Perron (see above) was used. The results are shown in Table 1.[21]

---

21 The unit-root test for the percentage of inhabitants aged 65 or older was applied to the first-differenced series, because preliminary analyses suggested that it might be a second-order integrated process.

Table 1: Unit-Root Tests

Variable	Period	Test	Test statistic	p
Robbery rate	1971–2004	IPS	0.79	0.788
		Maddal-Wu-ADF	16.42	0.795
Clearance rate robbery	1971–2004	IPS	–1.26	0.104
		Maddal-Wu-ADF	25.55	0.272
Real per capita disposable income	1971–2004	IPS	2.129	0.983
		Maddal-Wu-ADF	10.452	0.982
Real per capita consumption	1971–2004	IPS	2.417	0.992
		Maddal-Wu-ADF	8.237	0.997
Percentage foreigners	1971–2004	IPS	1.670	0.953
		Maddal-Wu-ADF	10.185	0.985
Δ Percentage 65+	1968–2004	IPS	5.013	1.000
		Maddal-Wu-ADF	0.827	1.000

For none of the variables can the null hypothesis of a unit-root be rejected. To check roughly if the results might be affected by cross-section correlation, the averages of the absolute values of the pairwise correlation coefficients between the first-differenced series for the individual federal states (|r|) were computed (Table 2).

Table 2: Average absolute value of the pairwise correlation coefficient and CIPS*-Unit-Root Tests

Variable	|r|	CIPS* statistic
Robbery rate	0.382	-
Clearance rate robbery	0.162	-
Real per capita disposable income	0.634	–2.019
Real per capita consumption	0.623	–1.598
Percentage foreigners	0.692	–1.207
Δ Percentage 65+	0.743	–3.003**

* p < 0.10 ** p < 0.05 (CIPS*-statistic only)

For all explanatory variables except the clearance rate, |r| exceeds the threshold of 0.6, above which – according to Hlouskova and Wagner (see above) – conventional unit-root tests are affected. Therefore, for these variables the CIPS* test by Pesaran was also computed (Table 2, third column). Pesaran's test was selected because the assumption of one common factor seemed a plausible possibility, while there was no reason to suspect that there might be several common factors. The CIPS* version was chosen because it seems to perform slightly better than CIPS in terms of size if T is small, according to the simulation results reported by Pesaran (2007). The results of the CIPS* test corroborate the findings of the conventional tests, with the exception of the first differences of the percentage of older people, for which the unit-root hypothesis is rejected. But it has to be kept in mind that this result applies only to the unit-specific component; it might nonetheless be that there is a common component with a unit-root, inducing non-stationarity in the observed series. Therefore, this variable was treated as non-stationary in further analyses despite the result of the CIPS* test.[22]

---

22 This should not lead to erroneous findings regarding long-term effects if this variable is, in fact, stationary: the residuals of a regression of the robbery rates on it will be non-stationary, and the cointegration test in the second step of the analysis will not reject the null hypothesis of no cointegration. If the latter happens, one could, in principle, miss a long-run effect of the (not-differenced) percentage of older people on the robbery rates, however. But if the null hypothesis of no cointegration is rejected (as here), it is fairly safe to conclude that the changes in the percentage of people aged 65 or older are, indeed, random walks (otherwise the regression residuals would not be stationary) and that they exert a long-run effect on the robbery rates.

5.2. Cointegration Tests

In a next step, it was assessed whether the supposed explanatory variables are cointegrated with the robbery rates. This was done using Pedroni's parametric t-statistics, which have comparatively good power in small samples. These were computed in the "group" as well as in the "panel" version, because none of the two seems to outperform the other under all possible circumstances. For lag selection the AIC was used, and besides a unit-specific intercept also a unit-specific time trend was specified. The tests were applied to the natural logs of the series, because a non-linear relationship is theoretically plausible.

Table 3: Parametric Pedroni cointegration tests

Independent variable (logged)	p-value panel-t-statistic	p-value group-t-statistic
Clearance rate robbery	0.066	**0.015**
Real per capita disposable income	0.339	**0.001**
Real per capita consumption	0.465	**0.012**
Percentage foreigners	**0.043**	**0.017**
Δ Percentage 65+	**0.002**	**0.000**

p-values < 0.05 are printed bold

The results are not totally clear-cut, because for three of the series the panel t statistics are not significant at the 5 percent level. I suspect that the lack of significance of some of the panel tests is due to the fact that they are constructed under the assumption that the residuals of the cointegration regression follow identical autoregressive processes under the alternative hypothesis, which might be too restrictive here. Therefore, I retain the group statistic, which allows the autoregressive behavior of the residuals to be heterogeneous across units, and conclude that all variables are cointegrated with the robbery rates. I did not apply a cointegration test constructed for the case of cross-section dependence because the effects of cross-section correlation on cointegration tests seem to be mild (see 4.4 above).[23]

5.3. Estimation of the Long-Run Parameters

To estimate the cointegration parameters, the MG and PMG estimators were computed and a Hausman test applied to determine if the more efficient PMG estimator is appropriate. Since complete parameter homogeneity seemed not realistic for the data at hand, I did not consider the DFE estimator.[24] All variables were entered in their natural logs. Besides the aforementioned variables, a time trend was entered into the estimation equation. The number of lagged first differences of the independent variables in the equation was determined using the Schwartz-Bayes Information Criterion (SBIC) after setting the maximum number to one (due to the limited number of observation periods). The resulting estimates for the average cointegration coefficients and the average error correction parameters are shown in Table 4.

Table 4: Error correction models for the logged robbery rate

Independent variable (logged)	PMG	MG
Real per capita disposable income	1.41**	2.74**
Real per capita consumption	–1.27**	–2.78**
Clearance rate robbery	–0.49**	–0.41*
Percentage foreigners	0.36**	0.29
Δ Percentage 65+	–4.70**	–5.05**
Lag specification	**1,1,1,1,1**	**0,0,0,0,0**
Average error correction parameter	–0.45	–0.56
Hausman-χ^2-statistic	3.23	
|r|	0.23	0.30
Number of observations	363	363

* $p < 0.10$ ** $p < 0.05$

---

23 There were no indications of breaks in the robbery rates. Therefore, there was no need to apply one of the cointegration tests (see 3.2 above) that account for breaks.

24 For the same reason, I decided against DOLS.

The Hausman test is not significant, so the hypothesis that the cointegration parameters are homogeneous cannot be rejected. Therefore, the PMG-specification is valid.

Besides that, the results obtained by the MG and the PMG estimators have the same sign, although the absolute values of the parameters differ, sometimes remarkably. Furthermore, the MG estimate for the coefficient of the percentage of foreigners among the population does not reach significance – probably because the MG estimator is less efficient than the PMG specification. It can be concluded that, in the long-run, the robbery rate drops if the clearance rate – an indicator for the probability of apprehension – for this crime rises, which is in line with the economic theory of crime (Becker 1968; Ehrlich 1973). Furthermore, the incidence of robbery also declines if real per capita consumption – which can be interpreted as a proxy for the supply of potential loot (the expected returns of criminal acts) – increases, while it grows with rising disposable income, a measure of legal income opportunities. These two findings are contrary to the predictions of economic theory; a detailed discussion of their implications is beyond the scope of the present article. The latter applies also to the findings regarding the percentage of foreigners and the change in the proportion of elderly people, which primarily served as control variables here.

Finally, there is only moderate cross-section correlation between the residuals for the individual federal states (see |r| in the second last row in Table 4). In view of this finding, and because the consequences of cross-section correlation seem to be mild according to the simulation results mentioned above, I did not apply estimators for cross-sectionally correlated data.

6. Conclusion

In recent years, much effort has been spent on the study of non-stationarity in PTCS data. It emerges that non-stationarity has the potential to invalidate conventional approaches to the analysis of such data. Therefore, unit-root tests are mandatory. A variety of such tests have been proposed, some of which are also appropriate if the data are subject to cross-section dependence and/or structural breaks. Furthermore, if the data are, in fact, non-stationary, the appropriate method of estimation of long-run relationships depends on whether the variables are cointegrated or not. This can be determined using one of the cointegration tests reviewed here. After establishing cointegration, long-run relationships can be estimated; for this purpose, a number of approaches have been developed. For each of the three steps, many of the procedures proposed in the literature can be implemented using standard software packages.[25] The general approach to the analysis of non-stationary PTCS data was illustrated using data for the West German federal states 1971–2004. The results regarding cointegration relationships with robbery rates only partially support the economic theory of crime.

For most situations there are now procedures available that show good performance with sufficient sample size. The latter qualifier (sufficient sample size), however, can make it difficult in applied work to simultaneously account for all possible complications mentioned here, because the data sets actually available often have a modest size. There is no general rule of thumb as to which problem might be ignored without seriously jeopardizing the validity of the results in such a situation. Throughout the paper, advice was given with respect to the issues to be considered when deciding how to proceed. In addition, it is advisable to carefully explore the properties of the spe-

25 EViews 8 offers a full suite of "first-generation" methods, including the LL, IPS and Maddala-Wu panel unit-root tests, Pedronis cointegration tests, and FM-OLS as well as DOLS estimation. Stata 13 offers panel unit-root tests (LL, IPS, Maddala-Wu and others). In addition, there are user-written ado-files which allow the implementation of Pesaran's CIPS* test (-pescadf-), Pedroni's residual-based cointegration tests (-xtpedroni-), Westerlund's ECM-based cointegration test (-xtwest-), DOLS (-xtdolshm-), panel error correction models (DFE, PMG, MG; -xtpmg-), and the CCEMG estimator (-xtmg-). Similarly, there are user-written codes available for RATS (panel unit-root tests (IPS, LL), Pedroni's panel cointegration tests, DOLS and FM-OLS). For R, there is a package plm, by which several first generation unit-root tests (LL, IPS, Maddala-Wu, Hadri), Pesaran's CIPS* test, as well as the CCEMG and CCEP estimators can be implemented; furthermore, the factor-analytic approach to unit-root testing of Bai and Ng (2004) can be applied via the package PANICr. The PANEL procedure in SAS/ETS 13.2 also provides limited facilities for non-stationary PTCS data (unit-root tests: LL, IPS, Maddala-Wu, Hadri, and a further test not discussed here due to Harris and Tzavalis [1999]).

cific data at hand and gather contextual information regarding relevant events which might induce breaks. If it emanates from this investigation that the observations are only moderately correlated cross-sectionally, and that there are no breaks in the series, one might apply "first generation" methods. If the results of the exploration are less reassuring, "second generation" approaches might be used, but one should be aware that there is some uncertainty with respect to the validity of the results, due to their limited performance when applied to samples of moderate size.

Besides that, further problems remain: For example, there is the issue of figuring out for which units the data are stationary if the hypothesis of a unit-root is rejected. Similarly, it would be useful to have a procedure available to find out if there is cointegration for all units if a cointegration test is significant. Regarding the latter problem, it is to be hoped that estimation methods which are valid even if there is a mixture of cointegrated and not-cointegrated units will be developed in the future,[26] which would obviate the need to partition the sample into cointegrated and not-cointegrated units.

References

Bai, Jushan, and Chihwah Kao. 2006. On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence. In *Panel Data Econometrics: Theoretical Contributions and Empirical Applications,* ed. Badi H. Baltagi, 3–30. Amsterdam: Elsevier.

Bai, Jushan, Chihwa Kao, and Serena Ng. 2009. Panel Cointegration with Global Stochastic Trends. *Journal of Econometrics* 149: 82–99.

Bai, Jushan, and Serena Ng. 2004. A PANIC Attack on Unit Roots and Cointegration. *Econometrica* 72 (4): 1127–77.

Baltagi, Badi H., Georges Bresson, and Alain Pirotte. 2007. Panel Unit Root Tests and Spatial Dependence. *Journal of Applied Econometrics* 22: 339–60.

Baltagi, Badi H., Chihwa Kao, and Min-Hsien Chiang. 2000. Nonstationary Panels, Cointegration in Panels and Dynamic Panels: A Survey. In *Nonstationary Panels, Panel Cointegration, and Dynamic Panels,* ed. Badi H. Baltagi, 7–51. Amsterdam: Elsevier.

Banerjee, Anindya, and Josep Lluis Carrion-i-Silvestre. 2006. *Cointegration in Panel Data with Breaks and Cross-Section Dependence,* Working Paper Series 591. Frankfurt am Main: Europäische Zentralbank.

Banerjee, Anindya, Massimiliano Marcellino, and Chiara Osbat. 2004. Some Cautions on the Use of Panel Methods for Integrated Series of Macro-Economic Data. *Econometrics Journal* 7:322–40.

Banerjee, Anindya, Massimiliano Marcellino, and Chiara Osbat. n.d. *Testing for PPP: Should We Use Panel Methods?* Florence: European University Institute.

Becker, Gary S. 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76:169–217.

Binder, Michael, Cheng Hsiao, and M. Hashem Pesaran. 2003. *Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration.* Frankfurt am Main and Los Angeles: Johann Wolfgang Goethe-Universität and University of Southern California.

Birkel, Christoph. 2015. *Die Entwicklung der Gewaltkriminalität in Deutschland: Theoretische Erklärungsansätze im empirischen Vergleich.* Wiesbaden: Springer.

Blomquist, Johan, and Joakim Westerlund. 2014. A Non-Stationary Panel Data Investigation of the Unemployment-Crime Relationship. *Social Science Research* 44:114–25.

Blundell, Richard, and Stephen R. Bond. 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics* 87:115–43.

Bond, Stephen R., Céline Nauges, and Frank Windmeijer. 2005. *Unit Roots: Identification and Testing in Micro Panels,* Cemmap Working Paper CWP07/05. Oxford: Institute of Fiscal Studies.

Breitung, Jörg. 2000. The Local Power of Some Unit Root Tests for Panel Data. In *Nonstationary Panels, Panel Cointegration, and Dynamic Panels,* ed. Badi H. Baltagi, 161–77. Amsterdam: Elsevier Science.

Breitung, Jörg. 2005. A Parametric Approach to the Estimation of Cointegration Vectors in Panel Data. *Econometric Reviews* 24 (2): 151–73.

Breitung, Jörg, and Bertrand Candelon. 2005. Purchasing Power Parity during Currency Crises: A Panel Unit Root Test under Structural Breaks. *Review of World Economics* 141 (1): 124–40.

Breitung, Jörg, and Hashem Pesaran. 2008. Unit Roots and Cointegration in Panels. In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice,* ed. László Mátyás and Patrick Sevestre, 279–322. Dordrecht: Kluwer.

Carrion-i-Silvestre, Josep Lluís, Tomás del Barrio-Castro, and Enrique López-Bazo. 2005. Breaking the Panels: An Application to the GDP per Capita. *Econometrics Journal* 8:159–75.

Chang, Yoosoon. 2004. Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependency. *Journal of Econometrics* 120 (2): 263–93.

Chen, Bangtian, Suzanne K. McCoskey, and Chiwah Kao. 1999. Estimation and Inference of a Cointegrated Regression in Panel Data: A Monte Carlo Study. *American Journal of Mathematical and Management Sciences* 19 (1–2): 75–114.

Chiang, Min-Hsien, Chihwa Kao, and Chih-Hsien Lo. 2007. The Estimation and Inference of a Panel Cointegration Model with a Time Trend. *Communications in Statistics – Theory and Methods* 36 (5–8): 1233–50.

Choi, In. 2001. Unit Root Tests for Panel Data. *Journal of International Money and Finance* 20:249–72.

Coakley, Jerry, Ana-Maria Fuertes, and Ron Smith. 2001. *Small Sample Properties of Panel Time-Series Estimators with I(1) Errors.* Cambridge: Birkbeck College.

Di Iorio, Francesca, and Stefano Fachin. 2007. *Cointegration Testing in Dependent Panels with Breaks.* University of Rome La Sapienza.

Eberhardt, Markus, and Stephen R. Bond. 2009. *Cross-section Dependence in Nonstationary Panel Models: A Novel Estimator.* Oxford: Department of Economics, University of Oxford and Institute of Fiscal Studies.

---

26 See the proposal of Hu (2006) for cointegration testing and estimation in this situation.

Ehrlich, Isaac. 1973. Participation in Illegitimate Activities: A Theoretical and Empirical Investigation. *Journal of Political Economy* 81: 521–65.

Enders, Walter. 2004. *Applied Econometric Time Series,* 2d ed. Hoboken: Wiley.

Entorf, Horst. 1997. Random Walks with Drifts: Nonsense Regression and Spurious Fixed-Effect Estimation. *Journal of Econometrics* 80:287–96.

Entorf, Horst, and Hannes Spengler. 2000. Socioeconomic and Demographic Factors of Crime in Germany: Evidence From Panel Data of the German States. *International Review of Law and Economics* 75:75–106.

Entorf, Horst, and Peter Winker. 2005. *Investigating the Drugs-Crime Channel in Economics of Crime Models: Empirical Evidence from Panel Data of the German States,* Darmstadt Discussion Papers in Economics 165. Technische Universität Darmstadt.

Fuertes, Ana-Maria. 2008. Sieve Bootstrap t-Tests on Long-Run Average Parameters. *Computational Statistics and Data Analysis* 52:3354–70.

Gengenbach, Christian, Franz C. Palm, and Jean-Pierre Urbain. 2004. *Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications for Modelling.* Maastricht: Department of Quantitative Economics, Universiteit Maastricht.

Gengenbach, Christian, Franz C. Palm, and Jean-Pierre Urbain. 2005. *Panel Cointegration Testing in the Presence of Common Factors.* Maastricht: Department of Quantitative Economics, Universiteit Maastricht.

Gregory, Allan W., and Bruce E. Hansen. 1996. Residual-Based Tests for Cointegration in Models with Regime Shifts. *Journal of Econometrics* 70:99–126.

Gutierrez, Luciano. 2003. On the Power Of Panel Cointegration Tests: A Monte Carlo Comparison. *Economics Letters* 80:105–11.

Gutierrez, Luciano. 2005. *Tests for Cointegration in Panels with Regime Shifts.* Sassari: Department of Agricultural Economics, University of Sassari.

Gutierrez, Luciano. n.d. Panel Unit Root Tests for Cross-Sectionally Correlated Panels: A Monte Carlo Comparison. Sassari: Department of Agricultural Economics, University of Sassari.

Hadri, Kaddour. 2000. Testing for Stationarity in Heterogenous Panel Data. *Econometrics Journal* 3:148–61.

Hadri, Kaddour, and Yao Rao. 2008. Panel Stationarity Test with Structural Breaks. *Oxford Bulletin of Economics and Statistics* 70 (2): 245–69.

Hale, Chris. 1998. Crime and the Business Cycle in Post-War Britain Revisited. *British Journal of Criminology* 38 (4): 681–98.

Han, Chirok, and Peter C.B. Phillips. 2010. *GMM Estimation for Dynamic Panels with Fixed Effects and Strong Instruments at Unity,* Cowles Foundation Paper 1290. New Haven: Cowles Foundation for Research in Economics.

Hanck, Christoph. 2009. For Which Countries Did PPP Hold? A Multiple Testing Approach. *Empirical Economics* 37:93–103.

Harris, David, Stephen Leybourne, and Brendan McCabe. 2005. Panel Stationarity Tests for Purchasing Power Parity with Cross-Sectional Dependence. *Journal of Business and Economic Statistics* 23 (4): 395–409.

Harris, Richard D. F., and Elias Tzavalis. 1999. Inference for Unit Roots in Dynamic Panels Where the Time Dimension is Fixed. *Journal of Econometrics* 91 (2):201–26.

Hayashi, Fumio. 2000. *Econometrics.* Princeton: Princeton University Press.

Hlouskova, Jaroslava, and Martin Wagner. 2006. The Performance of Panel Unit Root and Stationarity Tests: Results from a Large Scale Simulation Study. *Econometric Reviews* 25 (1): 85–116.

Hu, Ling. 2006. *A Simple Panel Cointegration Test and Estimation in I(0) / I(1) Mixed Panel Data.* Columbus: Department of Economics, The Ohio State University.

Hurlin, Christophe, and Valérie Mignon. 2004. *Second Generation Panel Unit Root Tests.* Paris: THEMA-CNRS, University of Paris X.

Im, Kyong-So, Junsoo Lee, and Margie Tieslau. 2005. Panel LM Unit-root Tests with Level Shifts. *Oxford Bulletin of Economics and Statistics* 67 (3): 393–419.

Im, Kyong So, M. Hashem Pesaran, and Yongcheol Shin. 2003. Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* 115 (1): 53–74.

Jacobs, David, and Amber M. Richardson. 2008. Economic Inequality and Homicide in the Developed Nations From 1975 to 1995. *Homicide Studies* 12 (1): 28–45.

Jönsson, Kristian. 2005. Using Panel Data to Increase the Power of Modified Unit Root Tests in the Presence of Structural Breaks. *Applied Mathematics and Computation* 171:832–42.

Judson, Ruth A., and Ann L. Owen. 1999. Estimating Dynamic Panel Data Models: A Guide for Macroeconomists. *Economics Letters* 65 (1): 9–15.

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

Kao, Chihwa, and Min-Hsien Chiang. 2000. On the Estimation and Inference of a Cointegrated Regression in Panel Data. In *Nonstationary Panels, Panel Cointegration, and Dynamic Panels,* ed. Badi H. Baltagi, 179–222. Amsterdam: Elsevier.

Kapetanios, George, M. Hashem Pesaran, and Takashi Yamagata. 2011. Panels With Nonstationary Multifactor Error Structures. *Journal of Econometrics* 160 (2): 326–48.

Levin, Andrew, Chie-Fu Lin, and Chia-Shang James Chu. 2002. Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties. *Journal of Econometrics* 108:1–24.

Maddala, Gangadharrao Soundalyarao, and Shaowen Wu. 1999. A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test. In *Testing for Unit Roots and Cointegration Using Panel Data: Theory and Applications,* ed. Amindya Banerjee, 631–52. Oxford: Blackwell.

McCoskey, Suzanne, and Chihwa Kao. 1998. A Residual-Based Test of the Null of Cointegration in Panel Data. *Econometric Reviews* 17:57–84.

Moon, Hyungsik Roger, and Benoit Perron. 2004. Testing for a Unit Root in Panels with Dynamic Factors. *Journal of Econometrics* 122 (1): 81–126.

Moon, Hyungsik Roger, and Benoit Perron. 2012. Beyond Panel Unit Root Tests: Using Multiple Testing to Determine the Nonstationarity Properties of Individual Series in a Panel. *Journal of Econometrics* 169 (1): 29–33.

Murray, Christian J., and David H. Papell. 2000. Testing for Unit Roots in Panels in the Presence of Structural Change with an Application to OECD Unemployment. In *Nonstationary Panels, Panel Cointegration, and Dynamic Panels,* ed. Badi H. Baltagi, 223–38. Amsterdam: Elsevier.

Ng, Serena, and Pierre Perron. 2001. Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. *Econometrica* 69 (6): 1519–54.

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

Örsal, Deniz Dilan Karaman. 2007. *Comparison of Panel Cointegration Tests,* SFB 649 Discussion Paper 2007–029. Humboldt-Universität zu Berlin.

Palm, Franz C., Stephan Smeekes, and Jean-Pierre Urbain. 2011. Cross-Sectional Dependence Robust Block Bootstrap Panel Unit Root Tests. *Journal of Econometrics* 163:85–104.

Pedroni, Peter Louis. 1999. Critical Values for Cointegration Tests in Heterogenous Panels with Multiple Regressors. In *Testing for Unit Roots and Cointegration Using Panel Data: Theory and Applications,* ed. Amindya Banerjee, 653–70. Oxford: Blackwell.

Pedroni, Peter Louis. 2000. Fully Modified OLS for Heterogenous Cointegrated Panels. In *Nonstationary Panels, Panel Cointegration, and Dynamic Panels,* ed. Badi H. Baltagi, 93–130. Amsterdam: Elsevier.

Pedroni, Peter Louis. 2004. Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the Purchasing Power Parity Hypothesis. *Econometric Theory* 20: 597-625.

Pesaran, M. Hashem. 2007. A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence. *Journal of Applied Econometrics* 22:265–312.

Pesaran, Hashem, Ron Smith, and Im Kyong So. 1996. Dynamic Linear Models for Heterogenous Panels. In *The Econometrics of Panel Data: A Handbook of the Theory with Applications,* ed. László Mátyás, and Sevestre, Patrick, 145–95. Dordrecht: Kluwer.

Pesaran, M. Hashem, and Elisa Tosetti. 2011. Large Panels with Common Factors and Spatial Correlation. *Journal of Econometrics* 161:182–202.

Pesaran, M. Hashem, L. Vanessa Smith, and Takashi Yamagata. 2008. *Panel Unit Root Tests in the Presence of a Multifactor Structure.* Discussion Papers in Economics No. 2008/03. Heslington: Department of Economics and Related Studies, University of York.

Pesaran, M. Hashem, Yongcheol Shin, and Ron Smith. 1999. Pooled Mean Group Estimation of Dynamic Heterogenous Panels. *Journal of the American Statistical Association* 94 (446): 621–34.

Phillips, Peter C. B., and Hyungsik Roger Moon. 1999. Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica* 67 (5): 1057–1111.

Phillips, Peter C. B., and Hyungsik Roger Moon. 2000. Nonstationary Panel Data Analysis: An Overview of Some Recent Developments. *Econometric Reviews* 19 (3): 263–86.

Phillips, Peter C. B., and Sam Ouliaris. 1990. Asymptotic Properties of Residual Based Tests for Cointegration. *Econometrica* 58: 165–93.

Sethrapramote, Yuthana. 2004. *Panel Unit Root Tests with Structural Breaks.* Coventry: Department of Economics, University of Warwick.

Smeekes, Stephan. 2010. *Bootstrap Sequential Tests to Determine the Stationary Units in a Panel,* METEOR Working paper RM/11/03. Maastricht: Department of Economics, Maastricht University.

Sun, Yixiao. 2004. Estimation of the Long-Run Average Relationship in Nonstationary Panel Time Series. *Econometric Theory* 20 (6): 1227–60.

Tam, Pui Sun. 2006. *Breaking Trend Panel Unit Root Tests.* Faculty of Business Administration, University of Macau.

Trapani, Lorenzo. 2012. On the Asymptotic t-Test for Large Nonstationary Panel Models. *Computational Statistics and Data Analysis* 56: 3286–3306.

Urbain, Jean-Pierre, and Joakim Westerlund. 2011. Least Squares Asymptotics in Spurious and Cointegrated Panel Regressions with Common and Idiosyncratic Stochastic Trends. *Oxford Bulletin of Economics and Statistics* 73:119–39.

Vieraitis, Lynne M., Tomislav V. Kovandzic, and Thomas B. Marvell. 2007. The Criminogenic Effects of Imprisonment: Evidence from State Panel Data, 1974–2002. *Criminology and Public Policy* 6 (3): 589–622.

Wagner, Martin, and Jaroslava Hlouskova. 2007. *The Performance of Panel Cointegration Methods: Results from a Large Scale Simulation Study,* Reihe Ökonomie/Economic Series 210. Vienna: Institut für Höhere Studien.

Westerlund, Joakim. 2006a. *Simple Unit Root Tests with Multiple Breaks.* Department of Economics, Lund University.

Westerlund, Joakim. 2006b. Testing for Panel Cointegration with a Level Break. *Economics Letters* 91:27–33.

Westerlund, Joakim. 2007. Testing for Error Correction in Panel Data. *Oxford Bulletin of Economics and Statistics* 69 (6): 709–48.

Westerlund, Joakim, and Syed A. Basher. 2008. Mixed Signals Among Tests for Panel Cointegration. *Economic Modelling* 25 (1): 128–36.

Westerlund, Joakim, and Jörg Breitung. 2013. Lessons from a Decade of IPS and LLC. *Econometric Review* 32 (5–6): 547–91.

Westerlund, Joakim, and David Edgerton. 2008. A Simple Test for Cointegration in Dependent Panels with Structural Breaks. Oxford *Bulletin of Economics and Statistics* 70 (5): 665–704.

Christoph Birkel
christoph.birkel@bka.bund.de