CESifo Working Papers

www.cesifo.org/wp

# Panel Cointegration Tests on the Fiscal Sustainability of German Laender

Heiko T. Burret
Lars P. Feld
Ekkehard A. Köhler

CESifo Working Paper No. 4929
Category 1: Public Finance
August 2014

CESifo *Working Paper No. 4929*

# Panel Cointegration Tests on the Fiscal Sustainability of German Laender

## Abstract

This paper provides new evidence on the sustainability of public finances in German states (Laender) by exploiting a newly compiled database covering the years 1950-2011. Unlike previous studies on Germany, we analyze fiscal sustainability by applying "second generation" panel cointegration techniques. A unique identification strategy for the selection of sub-panels improves the robustness of panel cointegration tests and reveals that Laender finances are hardly sustainable.

JEL-Code: H620, H770, H720.

Keywords: fiscal sustainability, federalism, unit root, panel cointegration.

*Heiko T. Burret*
*Walter Eucken Institute*
*Goethestr. 10*
*Germany – 79100 Freiburg*
*burret@eucken.de*

*Lars P. Feld*
*Walter Eucken Institute and*
*Albert-Ludwigs-University Freiburg*
*Goethestr. 10*
*Germany – 79100 Freiburg*
*feld@eucken.de*

*Ekkehard A. Köhler*
*Walter Eucken Institute*
*Goethestr. 10*
*Germany – 79100 Freiburg*
*koehler@eucken.de*

We would like to thank Konstantin Klemmer, Daniel Nientiedt and Leonardo Palhuca for valuable research assistance and Peter Hatzmann from the Federal Statistical Office for providing us the best data available.

*Things that can't go on forever -*
*eventually come to an end.*

*Herbert Stein (1916-1999)*

## 1. Introduction

The 16 German states (Laender) cover a major share of the increase in public debt in Germany between 1850 and 2010 (Burret et al. 2013). The future range for fiscal policy is limited on both levels of government: The German constitution prohibits (almost) all structural deficits on the federal level by 2016 and on the Laender level by 2020. While public finances need to be consolidated until then, appropriate measures are put at risk by the upcoming renegotiations of the German fiscal equalization scheme before its expiration in 2019. Unsound Laender finances might eventually attract larger equalization payments with the result that consolidation costs are spread across all jurisdictions and incentives for sustainable public finances erode.

Against this background this paper aims at clarifying whether German Laender finances have been sustainable in the long-run. Reviewing the existing literature, we observe two shortcomings: First, most empirical studies on sub-national fiscal policy use datasets of limited time range – excluding the 1950's and 1960's when economic growth rates were advantageous for public finances, especially in the German case (e.g., Kitterer 2007; Claeys et al. 2008; Herzog 2010). Second, most frequently empirical evidence is based on time series tests. Potrafke and Reischmann (2014) have submitted a study with panel data – however with a limited time period and without allowing for cross-dependence of German Laender. Thus, for Germany, previous research has not yet adopted multivariate econometric methods such as panel unit root tests and panel cointegration tests. Especially with respect to federal states, "second generation" panel tests are meaningful for the analysis of sub-national fiscal sustainability since they allow for cross-dependence in panel data (e.g., Westerlund and Prohl 2010; Mahdavi and Westerlund 2011).

These shortcomings motivate us to re-examine the issue and to apply most recent econometric test procedures with the following key features: First, we exploit a unique database of public finances of the German Laender covering the years 1950-2011. Second, we conduct "second generation" panel tests that have not yet been applied to fiscal data. In particular, we try to increase the robustness of panel cointegration tests by identifying and estimating sub-panels of German Laender which share similar time series characteristics.

The paper is organized as follows: In section 2, we briefly review the empirical literature on fiscal sustainability and related theoretical assumptions, respectively. The dataset and the empirical test strategy are presented in section 3. Empirical evidence is provided in section 4 for a panel of all West German Laender, and in section 5 for West German sub-panels that are identified conditional on their time series characteristics. Section 6 sums up our key findings. Section 7 draws some conclusions.

## 2. Literature Review

Most empirical studies on fiscal sustainability of German public finances focus on the general and federal government, respectively (e.g., Afonso 2005; Artis and Marcellino 1998, 1999; Bravo and Silvestre 2002; Fève and Hénin 2000; Garcia and Hénin 1999; Greiner and Kauermann 2007, 2008; Greiner and Semmler 1999; Greiner et al. 2005, 2006; Grilli 1988; Payne 1997; Polito and Wickens 2011; Vanhorebeek and van Rompuy 1995). While the results are ambiguous, fiscal sustainability seems to be rejected when Wagner's Law is taken into account (Koester and Priesmeier 2013) or longer time series are considered (Burret et al. 2013).

By contrast, fiscal sustainability of German Laender has, to the best of our knowledge, only been examined by five studies so far (Table 1). Kitterer (2007) estimates unit root tests of the debt to GDP ratio in West German Laender (1971-2004) and East German Laender (1992-2004), respectively. He rejects the hypothesis of fiscal sustainability in all Laender but Hesse, North Rhine-Westphalia and Saxony.[1] Claeys et al. (2008) conduct Model-Based Sustainability (MBS[2]) tests on the budget surplus response to debt developments for the periods 1970-2005 (West German Laender) and 1991-2005 (East German Laender). Claeys et al. (2008) find evidence that politicians do not significantly react to an increase of public debt by increasing budget surplus and therefore conclude that public finances on the German state level are not sustainable. While Herzog (2010) conducts MBS-tests and unit root tests, the study covers only two Laender. The results reject fiscal sustainability in the case of Berlin and provide mixed evidence for Baden-Wuerttemberg. Fincke and Greiner (2011) follow a twofold approach. By estimating the Bohn-model (MBS), they conclude that the reaction of primary surpluses to changes in public

[1] The study builds upon an earlier working paper by Kitterer and Finken (2006).
[2] The MBS test is also known as the "Bohn-model". See Bohn (2008).

debt was positive over the period 1975-2006 in all West German Laender except for Berlin. However, unit root tests reveal that public deficits are non-stationary in Berlin, Bremen and Saarland. Moreover, the time series of Hesse, Lower Saxony, North Rhine-Westphalia, Rhineland-Palatinate and Schleswig Holstein are only stationary with significance levels of 10%. A recent study by Potrafke and Reischmann (2014) estimates the Bohn-model for the West German Laender during the period 1980-2010 using OLS. Their results suggest that fiscal policy is unsustainable if fiscal transfers are excluded and sustainable otherwise.

Table 1 Studies on the Sustainability of German Laender Finances

| | Laender and period | Econometrics | Empirical test | Key findings | Fiscal sustainability?* |
|---|---|---|---|---|---|
| Kitterer (2007) | West (1971-2004)<br>East (1992-2004) | Time series<br>(Univariate) | Unit root tests | Fiscal sustainability not met in most Laender. | YES [HE, NW, SY]<br>NO [all other Laender] |
| Claeys et al. (2008) | West (1970-2005)<br>East (1991-2005) | Time series<br>(Univariate) | MBS-tests | Laender governments do not sufficiently react to increasing debt levels and curb consolidation requirements. | Rather NO [depends on time period under consideration] |
| Herzog (2010) | BE and BW<br>(1970-2005) | Time series<br>(Univariate) | Unit root tests,<br>MBS-tests | For BE sustainability is rejected by both tests.<br>For BW sustainability is rejected by unit root tests but not by MBS-tests. | NO [BE]<br>Mixed [BW] |
| Fincke and Greiner (2011) | West (1975-2006) | Time series<br>(Univariate) | Unit root tests,<br>MBS-tests | All but Bavaria account for rising "debt to GDP ratios which is not compatible with sustainability in the long run" (p. 248)<br>For HE, RP, NI, NW and SH sustainability is rejected by neither test, but budget deficits are only stationary at the 10% level in each state. | YES [BW, BY, HH]<br>Rather YES [HE, RP, NI, NW, SH]<br>NO [SL, HB, BE] |
| Potrafke and Reischmann (2014) | West (1980-2010) | Panel analysis<br>(Multivariate) | MBS-tests applied to panel using OLS | Including/excluding fiscal equalization transfers in the primary surplus changes the results | YES [if transfers included ]<br>NO [if transfers are excluded] |

* "YES" indicates that the empirical results suggest that fiscal sustainability is detected. Studies listed above may focus on further issues. East German Laender include Brandenburg (BB), Berlin (BE), Saxony (SN), Saxony-Anhalt (ST), Thuringia (TH), Mecklenburg Western-Pomerania (MW) and West German Laender include Baden-Wuerttemberg (BW), Bavaria (BY), Bremen (HB), Hesse (HE), Hamburg (HH), North Rhine-Westphalia (NW), Lower Saxony (NI), Rhineland-Palatinate (RP), Schleswig Holstein (SH) and Saarland (SL).

To sum up, studies on the sustainability of German Laender finances are rare and provide ambiguous evidence. While most papers conduct univariate unit-root tests, cointegration tests are not estimated. In particular ADF and PP test statistics are presented. All results are based on datasets of a limited time range with starting points hardly reaching back to the beginning of the 1970s. This is surely a shortcoming since the economically vibrant years are dismissed from the analysis. The recent economic crisis is not included in all but one study (Potrafke and Reischmann 2014). Moreover, multivariate panel cointegration methods have so far not been applied. Especially the recently developed "second generation" panel cointegration tests which consider cross-sectional dependence could provide additional insights on fiscal sustainability in federal systems.

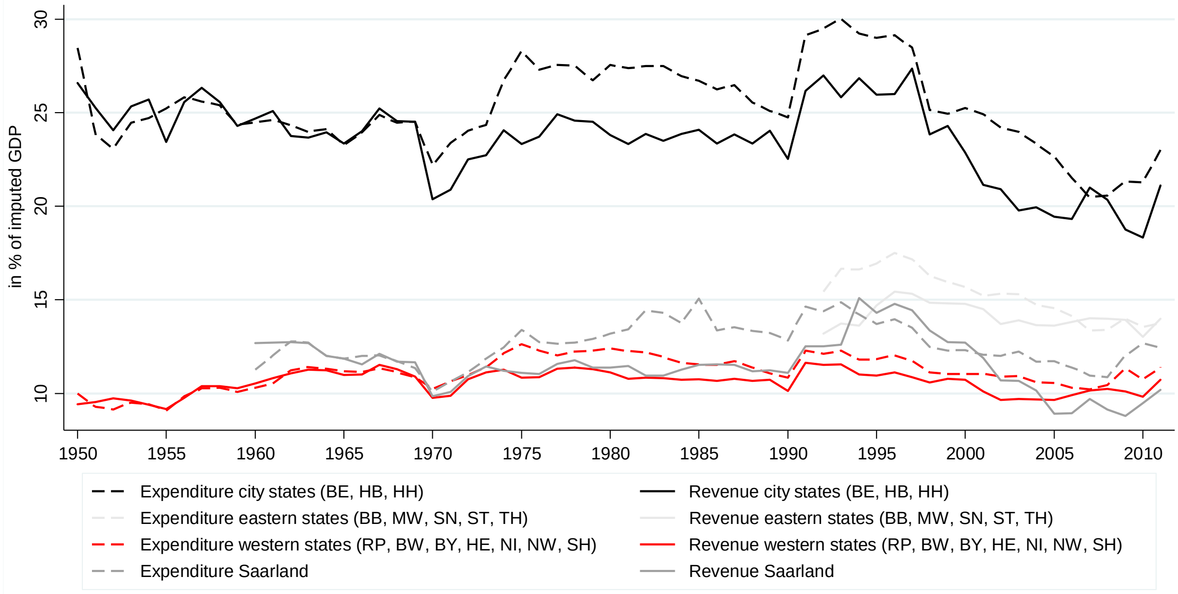
## 3. Data and Empirical Test Strategy

### 3.1. Data

The dataset comprises annual expenditures and revenues of each German Land for the years 1950-2011. Since Germany's territorial delineation changed in the past 60 years, data for the Saarland, the five East German Laender and Berlin is not available before 1960 or 1992, respectively. Instead of applying levels, the variables are measured in relation to GDP[3] to obtain a more natural definition of sustainability that keeps pace with economic development (Afonso 2005; Kirchgässner and Prohl 2008) and to achieve similarly scaled series that offer easily and intuitively interpretable information (Bohn 2008). Assuming rational choice in multi-level jurisdictions and rationality of the individuals involved in the budget process, we are urged to assess fiscal sustainability after the horizontal and vertical German fiscal equalization scheme that harmonizes tax revenues across the Laender (for a recent overview see: Burret and Feld 2013; Feld and Schnellenbach 2013). Figure 1 reveals that Laender spending exceeds Laender revenues in most of the years. While the gap seems to have diminished in the East German Laender on average, the three city states (Hamburg, Bremen and Berlin) show notably large fiscal deficits. In addition, the spending and revenue ratios of the city states are substantially larger than in any other state. Descriptive statistics, definitions and sources of the variables are provided in Table A.1 and A.2.

[3] Since GDP data on the German state level is not reliable we use imputed GDP. This is derived by multiplication of national GDP per capita with the population of the respective Land in the same year.

Figure 1 Development of Laender Expenditures and Revenues




### 3.2. Empirical Test and Panel Identification Strategy

The core idea is to discuss fiscal sustainability among German Laender rather than within single Laender. Unlike previous studies on fiscal sustainability of German Laender, we estimate "second generation" panel cointegration tests. Panel tests are known to be more powerful than time series tests since the cross-sectional dimension is additionally exploited. Moreover, "second generation" tests allow for cross-sectional variation among Laender. As an innovative element, we contribute an identification strategy for the selection of sub-panels.

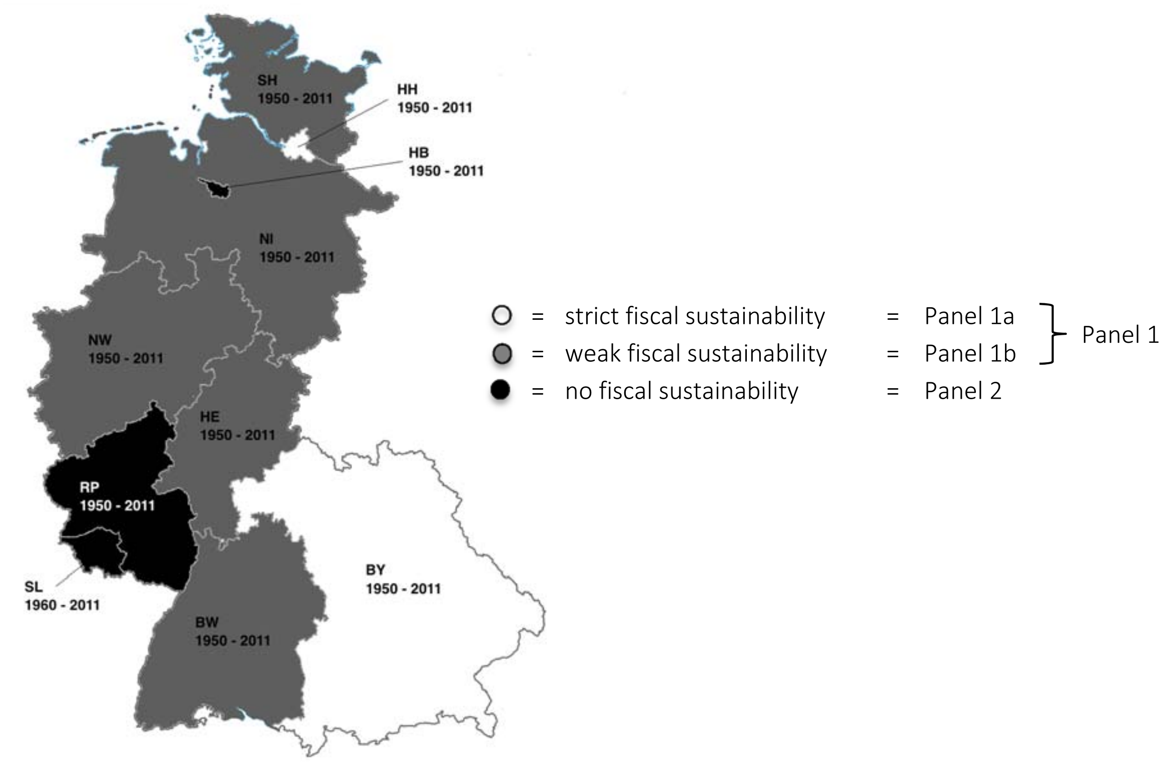
We follow two methods to **identify the panels**. First, the panels are selected in accordance to their territorial delineation: This method leads to three panels: (I) all Laender of the Federal Republic of Germany, (II) East German Laender and (III) West German Laender. The second method identifies the cross-sections conditional on the results of the time series analysis of a companion study that explores univariate and cointegration time series evidence of German Laender finances (Burret et al. 2014).[4] To be precise, we select Laender which share similar time series properties and, thus, sustainability characteristics. Since the limited time dimension leads to ambiguous time series evidence we exclude the East German Laender from conditional selection.[5] This leaves us with the West German Laender, which are grouped into four panels (see Figure 2): Panel 1 comprises each West German Land with a significant cointegration of

[4] For a summary of the results of Burret et al. (2014) and further explanatory notes on the identification of sub-panels see Box A.1.

[5] The results for East Germany reveal that cointegration is rejected for all East German states except for Brandenburg. The stationarity properties of the time series are also not without ambiguity.

expenditures and revenues. Panel 2 comprises each West German Land with no significant cointegration. Panel 1a and Panel 1b are sub-panels of Panel 1. Panel 1a includes each West German Land that does not fail the Chi-square test of the vector [1,-1] and is therefore "strictly" sustainable. Panel 1b includes each West German Land that fails the Chi-square test and is, thus, "weakly" sustainable.

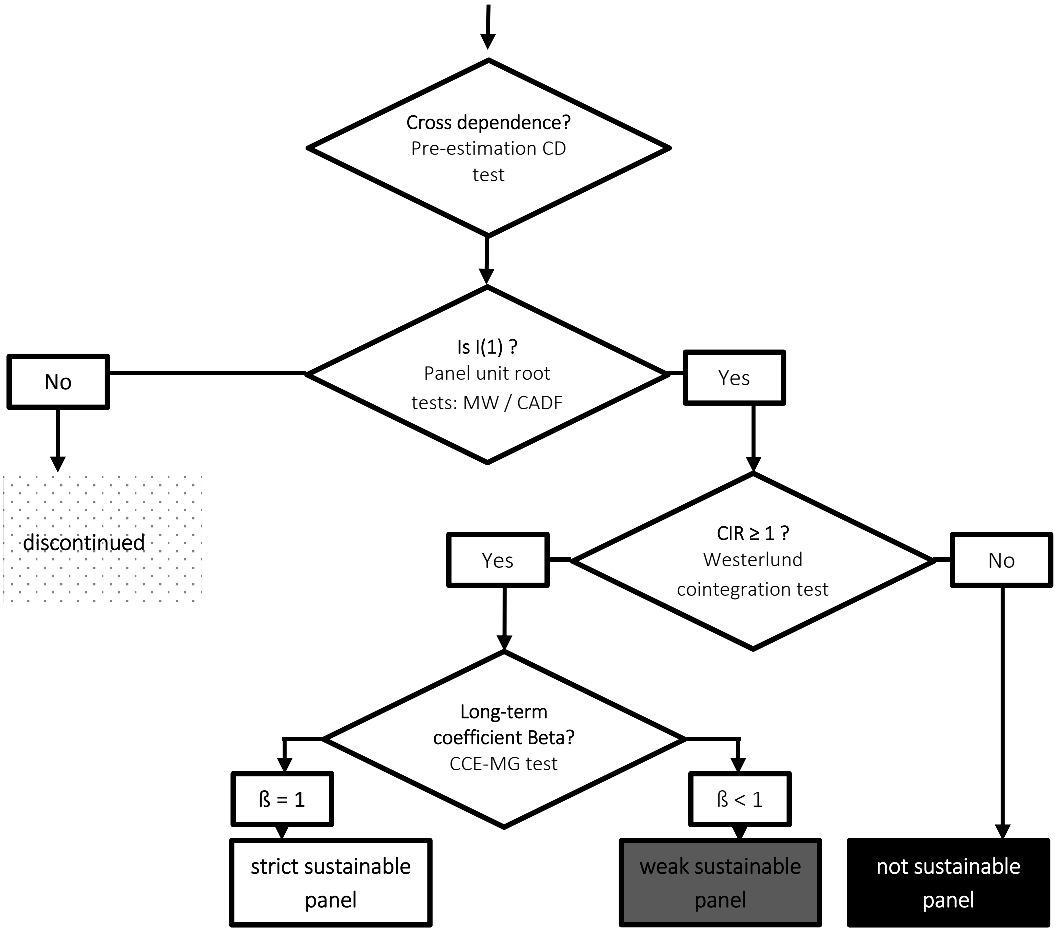
Figure 2 Graphical Representation of Identified West German Laender Sub-panels




Note: The years indicate the start and end date of the time series.

Each panel analysis is conducted in four steps (Figure 3). First, we test for cross-dependence (CD) in each panel. The results are used, second, in order to employ the appropriate panel unit root test. If we find evidence for CD, we apply the CADF test as suggested by Pesaran (2007) that controls for cross-dependence. In case CD is rejected, we use the MW (Maddala and Wu 1999) panel unit root test. If the panel is I(0) the panel analysis is discontinued because we do not have indication for cointegration. Otherwise we continue with panel cointegration tests as a third step testing for a unit slope in the (panel) regression of at least two cointegrated time series. If the panel is not cointegrated we have evidence for fiscal unsustainability because there is no significant long-term relation between expenditures and revenues. If the panel is cointegrated we estimate the magnitude of the cointegration coefficient $\beta$ in cross-section cointegration regressions for each Land and for different panels as a fourth step. Strict sustainability is obtained if β is equal to one. A smaller β still provides evidence for weak fiscal sustainability since expenditures and revenues are cointegrated. The empirical tests applied in each step are briefly explained when discussing the results in section 4.

Figure 3 Empirical Test Procedure: Four Step Panel Analysis for Each Panel




## 4. Empirical Evidence for Panels Identified by Geographical Patterns

While three panels are identified by geographical patterns, we solely report the findings for the West German panel (excluding Berlin) as the few (20) observations of the East German Laender restrict a meaningful interpretation of the full and East German panel results.[6]

### 4.1. Cross-dependence Tests in West German Panel

To assess whether the cross-section independence assumption of the "first generation" tests is valid, we start with a test for error cross-sectional dependence (CD) as suggested by Pesaran (2004).[7] This test is meaningful in our case since German Laender are economically, fiscally and

[6] The results for East Germany (including Berlin) and all German states are available upon request.
[7] CD tests have received great attention in macro- and microeconomic panel analysis (Moscone and Tosetti 2009; Sarafidis and Wansbeek 2012).

politically integrated. Econometrically speaking, the CD test is based on an average of pairwise correlation coefficients of OLS residuals from individual regressions, i.e. for each panel member (see Pesaran 2012; Baltagi 2013, 287ff.). The test works with unbalanced panels and is robust to single and multiple structural breaks in the slope coefficients and the error variances of the individual regressions. If the dataset contains N units (in our case N=10) the test estimates N*(N-1) correlations between state i=1 and all other states (N-1). Table 2 indicates that the null hypothesis of cross-section independence is strongly rejected for both time series. Moreover, the correlation coefficients are rather high.

Table 2 Pre-estimation Test on Cross-section Correlation

| | CD test | p-value | Average correlation coefficient | Absolute correlation coefficient |
|---|---|---|---|---|
| Revenue | 23.00 | 0.00 | 0.444 | 0.461 |
| Expenditure | 31.43 | 0.00 | 0.606 | 0.606 |

Note: We report the average and absolute correlation coefficient across N x (N-1) pairs of correlation. CD presents the Pesaran (2004) cross-section dependence statistic which is distributed standard normal and tests the null hypothesis of cross-section independence. We use the Stata routine xtcd.

### 4.2. Panel Unit Root Tests in West German Panel

In the presence of cross-section dependence, "first generation" panel unit root tests tend to reject the null hypothesis of a unit root excessively. Therefore we apply the CADF test suggested by Pesaran (2007). The test is based on the mean of individual ADF t-statistics of each unit in the panel. It eliminates cross-sectional dependence by augmenting the ADF regression with the lagged cross-sectional mean and its first differences of the individual series (CADF statistics) to capture CD by a single factor model. Since the lag length frequently influences the test results we carefully determine the number of lags using two approaches: First, the "ideal" lag length is separately selected for each Land using the Akaike Information Criterion (AIC). The resulting average number of lags is then used in the CADF test. Second, we alternatively report evidence for the lag bandwidth 0-4. Thereby we try to address the issue that too few lags fail to capture the system's dynamics leading to omitted variable bias, and that too many lags cause a loss of degrees of freedom resulting in over-parameterization. Following Hoechle (2007), we select the ideal lag length by using Newey and West's (1994) plug-in procedure at ($4*(T/100)^{2/9}\approx 3$). The results for the lag bandwidth [0, 4] are reported in the appendix in each case. All tests are estimated in levels and in first differences, with and without a trend, respectively.

The unit root tests for the panel of West German Laender depend on the number of lags included. If we follow Pesaran's procedure and apply VAR estimated state-specific lags that are averaged for the panel unit root test, the number of lags is non-integer, i.e. 1.7 for revenues and 1.5 for expenditures (Table 3). While the test results suggest that the series are I(0) at least at the 10% level with and without trend, it seems reasonable to assume a lag length of two given the non-integer number of lags. If we allow for two or more lags, we have evidence that the series are I(1) (Table A.3). In compliance with this finding, univariate time series properties indicate I(1) in seven out of the ten West German Laender (see Box A.1). Thus, we conclude that expenditure and revenues in the West German Laender are I(1).

Table 3 Pesaran Panel Unit Root Test of the West German Laender

| | Levels | | | | First differences | | | |
|---|---|---|---|---|---|---|---|---|
| | without trend | | with trend | | without trend | | with trend | |
| | Z[t-bar] | p-value | Z[t-bar] | p-value | Z[t-bar] | p-value | Z[t-bar] | p-value |
| Revenues | | | | | | | | |
| Lag average: 1.7 | -2.150** | 0.016 | -1.859** | 0.031 | -10.889*** | 0.000 | -10.342*** | 0.000 |
| Expenditures | | | | | | | | |
| Lag average: 1.5 | -1.845** | 0.033 | -1.501* | 0.067 | -12.226*** | 0.000 | -11.821*** | 0.000 |

Note: The null hypothesis for all tests is that the variables are I(1). We use Stata routine pescadf.

### 4.3. Panel Cointegration Tests in West German Panel

Panel cointegration tests reveal whether there is a linear combination of expenditures and revenues in our panel. If the variables share a conjoint long-run relation within the corresponding group we have first evidence for weak sustainability. We apply the error correction based cointegration test for (unbalanced) panels developed by Westerlund (2007). The test is meaningful for application in our case for the following reasons: First, it is general enough to allow for a large degree of heterogeneity, both in the long-run cointegration relation and in the short-run dynamics (Persyn and Westerlund 2008). Second, it is developed to cope with cross-sectionally dependent data. Third, the test comes along with an optional bootstrap procedure that allows for multiple repetitions of the cointegration tests which is meaningful since we have indications for cointegration in the panel. The Westerlund test has the null hypothesis of no cointegration by "inferring whether the error-correction term in a conditional panel error-correction model is equal to zero" (Persyn and Westerlund 2008: 232). The alternative hypothesis depends on the specific test. While, the group-mean tests (Gt and Ga) examine the alternative hypothesis that at least one unit is cointegrated, the panel tests (Pt and Pa) have the alternative hypothesis that the panel is cointegrated as a whole.[8]

Since the test results may be sensitive to the choice of lags, leads and kernel width, we estimate different specifications of each cointegration test: In the unrestricted case we use AIC to

[8] 'a' refers to the estimation of the error correction estimate, while 't' refers to the estimation for the standard error of 'a'. For further information see Persyn and Westerlund (2008: 233-235).

determine the optimal lag and lead length with 3 at most and with the Bartlett kernel window width set in accordance to the plug-in procedure $4*(T/100)^{2/9} \approx 3$. In the restricted case we assume the same short-run dynamics for all series (with a single lag and lead) and, thus, hold the short-term dynamics fixed. Both cases are estimated with a constant and with a constant and trend in the error correction relation. Bootstrap resampling procedures are applied at 800 re-estimations for each Westerlund panel cointegration test and provide us with robust-p-values. This is required to avoid misleading inference in case of cross-member correlation.

Table 4 Westerlund Panel Cointegration Test of West German Laender Panel

| | Constant | | | | Constant and trend | | | |
|---|---|---|---|---|---|---|---|---|
| | Value | Z-value | p-value | Robust p-value | Value | Z-value | p-value | Robust p-value |
| | Unrestricted (average lag length 0.1) | | | | Unrestricted (average lag length 0) | | | |
| Gt | -3.420 | -5.780 | 0.000 | 0.000 | -4.439 | -8.066 | 0.000 | 0.000 |
| Ga | -17.458 | -5.993 | 0.000 | 0.000 | -25.807 | -6.415 | 0.000 | 0.000 |
| Pt | -9.965 | -5.419 | 0.000 | 0.003 | -12.465 | -6.614 | 0.000 | 0.000 |
| Pa | -14.372 | -7.224 | 0.000 | 0.001 | -20.335 | -5.881 | 0.000 | 0.000 |
| | Fixed shot-term dynamics | | | | Fixed shot-term dynamics | | | |
| Gt | -3.137 | -4.786 | 0.000 | 0.000 | -3.959 | -6.197 | 0.000 | 0.000 |
| Ga | -17.552 | -6.047 | 0.000 | 0.000 | -26.230 | -6.612 | 0.000 | 0.000 |
| Pt | -6.740 | -2.175 | 0.015 | 0.094 | -8.756 | -2.374 | 0.009 | 0.088 |
| Pa | -9.551 | -3.790 | 0.000 | 0.035 | -14.561 | -2.903 | 0.002 | 0.037 |

Note: We use Stata routines xtwest written by Persyn and Westerlund (2008). Gt and Ga are group mean tests, while Pt and Pa are panel mean tests. See footnote 9.

In the unrestricted case the null hypothesis of no cointegration is rejected at any meaningful significant level by the simple and by the robust p-values (Table 4). This also holds if a deterministic trend is included additionally to a constant in the cointegration relation and to the inclusion of a lead. If we restrict the short term dynamics, the robust p-values still reject the null hypothesis of the group-mean tests (Ga and Gt). However, the null hypothesis of the panel tests (Pt and Pa) can only be rejected at a 10% significance level. This provides evidence that the panel is rather not cointegrated as a whole, but expenditures and revenues of at least some Laender are cointegrated. These results match with state-specific time series evidence which rejects cointegration in the case of Bremen (HB), Rhineland-Palatinate (RP) and Saarland (SL) (see Burret et al. 2014; Box A.1). Thus, we conclude that expenditures and revenues are not cointegrated in the West German panel and refrain from estimating the magnitude of the cointegration coefficient ß. The overall findings indicate that it might be meaningful to test sub-panels with similar sustainability patterns based on time series results.

## 5. Empirical Evidence for Panels Identified by Time Series Test Results

Due to the lacking evidence for cointegration in the West German panel, we proceed with the examination of West German sub-panels that share similar sustainability patterns based on time series results. In section 3.2 we have identified the following groups: **panel 1** comprises Laender with cointegrated public finances in the sense of a significant long-term relation between expenditures and revenues (BY, BW, HE, HH, NI, NW, SH), **panel 1a** comprises Laender that conjointly pass the test of a cointegration vector [1,-1] (BY, HH), **panel 1b** comprises Laender with at least one cointegration relation and a trend and constant in the cointegration relation (BW, HE, NI, NW, SH) and **panel 2** comprises Laender that have none or more than one cointegration relation (HB, RP, SL). Note that all tests are applied in the consecutive steps as explained in 3.1 and demonstrated in the section above.

### 5.1. Cross-dependence Tests in West German Sub-panels

Table 5 shows the results for the CD test. The estimations yield CD test statistics that allow for rejecting the null hypothesis of no cross-dependence in all panels except for panel 1a. The rejection of cross-dependence in this panel is econometrically not surprising, since N*(N-1) dimensions are tested. We estimate stationarity patterns of panel 1a using the Maddala and Wu (1999) panel unit root test and refrain from estimating the long-run equilibrium with second generation panel cointegration tests. In all other panels we subsequently apply "second generation" panel unit root and cointegration tests.

Table 5 Cross-dependence Tests of West German Sub-Panels

| | Panel 1<br>BW,BY, HH, HE, NI, NW, SH | | Panel 1a<br>BY, HH | | Panel 1b<br>BW, HE, NI, NW, SH | | Panel 2<br>HB, RP, SL | |
|---|---|---|---|---|---|---|---|---|
| | CD test | p-value | CD test | p-value | CD test | p-value | CD test | p-value |
| Revenues | 17.12 | 0.000 | -1.59 | 0.111 | 13.73 | 0.00 | 5.77 | 0.00 |
| Expenditures | 22.63 | 0.000 | 1.01 | 0.313 | 17.33 | 0.00 | 7.42 | 0.00 |

Note: For reasons of clarity we do not report the average and absolute correlation coefficient across N x (N-1) pairs of correlation. CD presents the Pesaran (2004) cross-section dependence statistic which is distributed standard normal and tests the null hypothesis of cross-section independence. We use the Stata routine xtcd written by Markus Eberhardt.

### 5.2. Panel Unit Root Tests in West German Sub-panels

In Table 6 we report the results of the Pesaran panel unit root test for **panel 1**, when the non-integer average number of lags is applied (1.6 for revenues and 1.7 for expenditures). The null of non-stationarity of revenues is retained without a trend and rejected if we allow for a trend and consider first differences, respectively. The results of expenditures indicate I(0) at the 10% level without and I(1) with trends. If we use a length of two and more lags expenditures and revenues are clearly non-stationary in levels and stationary in first differences (Table A.4).

Table 6 Pesaran Panel Unit Root Test of Panel 1 (BY, BW, HE, HH, NI, NW, SH)

| | Levels | | | | First differences | | | |
|---|---|---|---|---|---|---|---|---|
| | without trend | | with trend | | without trend | | with trend | |
| | Z[t-bar] | p-value | Z[t-bar] | p-value | Z[t-bar] | p-value | Z[t-bar] | p-value |
| Revenues | | | | | | | | |
| Lag average: 1.6 | -0.485 | 0.314 | -1.679** | 0.047 | -10.126*** | 0.000 | -9.550*** | 0.000 |
| Expenditures | | | | | | | | |
| Lag average: 1.7 | -1.331* | 0.092 | -2.028 | 0.021 | -10.242*** | 0.000 | -9.774*** | 0.000 |

Note: The null hypothesis for all tests is that the variables are I(1). We use Stata routine pescadf.

The CADF test for **panel 2** applies an average lag length of 2 in case of revenues and 1 in case of expenditures (Table 7). While the results indicate that the revenues are stationary without a trend and non-stationary with a trend, expenditures seem to be stationary in both cases. If we consider the ideal number of lags determined by the plug-in procedure, i.e. three lags, I(0) for expenditures is confirmed and revenues seem to be I(1) (Table A.5). Thus, the results for panel 2 are ambiguous. Nevertheless, revenues and expenditures can still be cointegrated.

Table 7 Pesaran Panel Unit Root Test of Panel 2 (HB, RP, SL)

| | Levels | | | | First differences | | | |
|---|---|---|---|---|---|---|---|---|
| | without trend | | with trend | | without trend | | with trend | |
| | Z[t-bar] | p-value | Z[t-bar] | p-value | Z[t-bar] | p-value | Z[t-bar] | p-value |
| Revenues | | | | | | | | |
| Lag average: 2.0 | -1.293* | 0.098 | -1.249 | 0.106 | -5.175*** | 0.000 | -5.172*** | 0.000 |
| Expenditures | | | | | | | | |
| Lag average: 1.0 | -2.415*** | 0.008 | -1.526* | 0.064 | -6.387*** | 0.000 | -6.366*** | 0.000 |

Note: The null hypothesis for all tests is that the variables are I(1). We use Stata routine pescadf.

For **panel 1a** we present the test results of the Maddala and Wu (1999) unit root test since cross-dependence is rejected in the CD-test. Non-stationarity is only rejected for revenues in case of zero and one lag if a trend is included (Table 8). If we follow the rule of thumb instead and determine an ideal lag length of three, we find strong evidence that revenues and expenditures are I(1).

Table 8 Maddala and Wu Panel Unit Root Test of Panel 1a (BY, HH)

| | Levels | | | | First differences | | | |
|---|---|---|---|---|---|---|---|---|
| | without trend | | with trend | | without trend | | with trend | |
| | Chi-square | p-value | Chi-square | p-value | Chi-square | p-value | Chi-square | p-value |
| Revenues | | | | | | | | |
| Lag 0 | 7.332 | 0.119 | 16.103*** | 0.003 | 144.217*** | 0.000 | 124.966*** | 0.000 |
| Lag 1 | 5.395 | 0.249 | 10.214** | 0.037 | 79.380*** | 0.000 | 66.971*** | 0.000 |
| Lag 2 | 2.900 | 0.575 | 7.111 | 0.130 | 71.967*** | 0.000 | 60.789*** | 0.000 |
| Lag 3 | 2.608 | 0.625 | 4.471 | 0.346 | 43.048*** | 0.000 | 35.002*** | 0.000 |
| Lag 4 | 3.634 | 0.458 | 4.294 | 0.368 | 25.088*** | 0.000 | 19.757*** | 0.000 |
| Expenditures | | | | | | | | |
| Lag 0 | 6.277 | 0.179 | 6.403 | 0.171 | 115.170*** | 0.000 | 100.827*** | 0.000 |
| Lag 1 | 6.943 | 0.139 | 7.909 | 0.095 | 87.698*** | 0.000 | 75.662*** | 0.000 |
| Lag 2 | 3.371 | 0.498 | 3.446 | 0.486 | 45.513*** | 0.000 | 37.157*** | 0.000 |
| Lag 3 | 3.621 | 0.460 | 3.708 | 0.447 | 34.280*** | 0.000 | 28.859*** | 0.000 |
| Lag 4 | 3.152 | 0.533 | 3.108 | 0.528 | 25.276*** | 0.000 | 20.930*** | 0.000 |

Note: The null hypothesis for all tests is that the variables are I(1). We use Stata routine multipurt.

Laender specific lag lengths suggest to use an average lag length of 1.2 for revenues and 1.6 for expenditures in panel 1b. The results of the CADF test with Laender specific lags are reported in Table 9. We cannot reject I(1) in revenues and expenditures at the 5% level if we allow for a trend. Note however that expenditures are I(0) at the 10% significance level with a trend. At the ideal lag length of three determined by the plug-in procedure, the series are I(1) with a trend (Table A.6). The same holds for revenues if no trend is included. Thus, we conclude that both series are I(1) with a trend in the cointegration relation (at the ideal lag length of three).

Table 9 Pesaran Panel Unit Root Test of Panel 1b (BW, HE, NI, NW, SH)

| | Levels | | | | First differences | | | |
|---|---|---|---|---|---|---|---|---|
| | without trend | | with trend | | without trend | | with trend | |
| | Z[t-bar] | p-value | Z[t-bar] | p-value | Z[t-bar] | p-value | Z[t-bar] | p-value |
| Revenues | | | | | | | | |
| Lag average: 1.2 | -1.180 | 0.119 | -0.741 | 0.229 | -9.684*** | 0.000 | -9.526*** | 0.000 |
| Expenditures | | | | | | | | |
| Lag average: 1.6 | -2.952*** | 0.002 | -1.570* | 0.058 | -8.685*** | 0.000 | -8.408*** | 0.000 |

Note: The null hypothesis for all tests is that the variables are I(1). We use Stata routine pescadf.

To sum up, the results of the panel unit root tests indicate that sub-panels 1, 1a and 1b are I(1). This holds in particular if we estimate with trends and assume an "ideal" lag length of three as determined by the plug-in procedure and the rule of thumb, respectively. Similar findings are obtained for the full panel of West German Laender in 4.2. However, since our results are, at least partially, sensitive to the number of lags included and evidence of I(1) is ambiguous in panel 2, we apply cointegration tests for every sub-panel to further explore cointegration characteristics.

### 5.3. Panel Cointegration Tests in West German Sub-panels

The Westerlund error correction based cointegration tests for panel 1 indicate that the null hypothesis of no cointegration is rejected at least at the 1% significance level in each specification, even if we hold the short-term dynamics fixed (Table 10). Similar results are obtained if we consider the robust p-values. This is clear evidence that expenditures and revenues are cointegrated in panel 1 as a whole and that these Laender share a conjoint long-run relation.

The results for panel 2 are somewhat contrary to the findings above. In the unrestricted case the null hypothesis of no cointegration within the cross-section can be rejected at least at the 5% significance level if we do not allow for a trend (Table 11). Adding a trend leads to retaining the null hypothesis of no cointegration - particularly if the robust p-values of the panel tests (Pt and Pa) are considered. The finding of no cointegration is confirmed in the case of fixed short-term dynamics: Here the null hypothesis cannot be rejected in any test specification, neither by the simple p-values nor by the robust p-values.

Table 10 Westerlund Panel Cointegration Test of Panel 1 (BY, BW, HE, HH, NI, NW, SH)

| | Constant | | | | Constant and trend | | | |
|---|---|---|---|---|---|---|---|---|
| | Value | Z-value | p-value | Robust p-value | Value | Z-value | p-value | Robust p-value |
| Unrestricted (average lag length 0.14) | | | | | Unrestricted (average lag length 0) | | | |
| Gt | -3.701 | -5.664 | 0.000 | 0.000 | -4.933 | -8.357 | 0.000 | 0.000 |
| Ga | -19.518 | -6.015 | 0.000 | 0.000 | -29.683 | -6.875 | 0.000 | 0.000 |
| Pt | -10.031 | -6.237 | 0.000 | 0.000 | -15.195 | -10.983 | 0.000 | 0.000 |
| Pa | -18.351 | -8.416 | 0.000 | 0.000 | -35.483 | -11.457 | 0.000 | 0.000 |
| Fixed shot-term dynamics | | | | | Fixed shot-term dynamics | | | |
| Gt | -3.564 | -5.262 | 0.000 | 0.000 | -4.501 | -6.950 | 0.000 | 0.000 |
| Ga | -20.377 | -6.432 | 0.000 | 0.000 | -30.650 | -7.252 | 0.000 | 0.000 |
| Pt | -8.082 | -4.277 | 0.000 | 0.006 | -11.707 | -6.996 | 0.000 | 0.001 |
| Pa | -15.484 | -6.707 | 0.000 | 0.001 | -32.671 | -10.243 | 0.000 | 0.000 |

Note: Null: No cointegration. Average AIC selected lag length for the unrestricted test. We apply xtwest command by Joakim Westerlund.

Thus, the restricted case largely suggests that there is neither a cointegration vector in panel 2 as a whole (Pt and Pa tests) nor between single Laender (Gt and Ga tests), respectively. The finding is supported by time series evidence derived in a companion study (Burret et al. 2014): Expenditures and revenues are only cointegrated in Bremen, while no cointegration is found in Saarland and Rhineland-Palatinate (Box A.1). In sum, our results for panel 2 depend on the trend assumption and the restriction of short-term dynamics but evidence tends to retain the null hypothesis of no cointegration.

Table 11 Westerlund Panel Cointegration Test of Panel 2 (HB, RP, SL)

| | Constant | | | | Constant and trend | | | |
|---|---|---|---|---|---|---|---|---|
| | Value | Z-value | p-value | Robust p-value | Value | Z-value | p-value | Robust p-value |
| Unrestricted (average lag length: 0.14) | | | | | Unrestricted (average lag length: 0) | | | |
| Gt | -2.764 | -1.901 | 0.029 | 0.037 | -3.286 | -1.960 | 0.025 | 0.045 |
| Ga | -12.379 | -1.666 | 0.048 | 0.031 | -16.248 | -1.079 | 0.140 | 0.036 |
| Pt | -4.815 | -2.321 | 0.010 | 0.059 | -5.342 | -1.924 | 0.027 | 0.108 |
| Pa | -12.105 | -3.072 | 0.001 | 0.039 | -14.366 | -1.535 | 0.062 | 0.108 |
| Fixed shot-term dynamics | | | | | Fixed shot-term dynamics | | | |
| Gt | -2.141 | -0.700 | 0.242 | 0.238 | -2.694 | -0.698 | 0.243 | 0.251 |
| Ga | -10.588 | -1.096 | 0.137 | 0.106 | -15.332 | -0.846 | 0.199 | 0.136 |
| Pt | -2.803 | -0.298 | 0.383 | 0.478 | -3.643 | 0.018 | 0.507 | 0.599 |
| Pa | -6.670 | -0.952 | 0.171 | 0.306 | -9.327 | -0.111 | 0.456 | 0.558 |

Note: Null: No cointegration. Average AIC selected lag length for the unrestricted test. We apply xtwest command by Joakim Westerlund.

The Westerlund cointegration test results for **panel 1b** are similar to the findings for panel 1 (Table 12). No cointegration can be rejected for the panel as a whole (Pt and Pa test) as well as for cointegration of at least one cross-section on the 1% significance level (Gt and Ga test). If short-term dynamics are fixed, the null hypothesis can still be rejected at the 1% level. Thus, we have conclusive evidence that expenditures and revenues are cointegrated in panel 1b as a whole and between single Laender. A similar conclusion is drawn with respect to sample 1.

Table 12 Westerlund Panel Cointegration Test of Panel 1b (BW, HE, NI, NW, SH)

| | Constant | | | | Constant and trend | | | |
|---|---|---|---|---|---|---|---|---|
| | Value | Z-value | p-value | Robust p-value | Value | Z-value | p-value | Robust p-value |
| | Unrestricted (average lag length: 0.14) | | | | Unrestricted (average lag length: 0) | | | |
| Gt | -3.693 | -4.767 | 0.000 | 0.000 | -4.643 | -6.265 | 0.000 | 0.000 |
| Ga | -18.998 | -4.870 | 0.000 | 0.000 | -27.442 | -5.074 | 0.000 | 0.000 |
| Pt | -7.768 | -4.557 | 0.000 | 0.000 | -10.541 | -6.652 | 0.000 | 0.000 |
| Pa | -16.727 | -6.294 | 0.000 | 0.000 | -27.689 | -6.840 | 0.000 | 0.000 |
| | Fixed shot-term dynamics | | | | Fixed shot-term dynamics | | | |
| Gt | -3.412 | -4.068 | 0.000 | 0.000 | -4.358 | -5.481 | 0.000 | 0.000 |
| Ga | -19.361 | -5.019 | 0.000 | 0.000 | -28.742 | -5.501 | 0.000 | 0.000 |
| Pt | -6.914 | -3.698 | 0.000 | 0.003 | -9.603 | -5.579 | 0.000 | 0.000 |
| Pa | -16.130 | -5.994 | 0.000 | 0.001 | -28.427 | -7.109 | 0.000 | 0.000 |

Note: Null: No cointegration. Average AIC selected lag length for the unrestricted test. We apply xtwest command by Joakim Westerlund.

## 5.4. Magnitude of Cointegration Coefficient ß

Empirical evidence suggests that expenditures and revenues are cointegrated in the sub-panels 1 and 1a and less likely in sub-panel 2. To further explore the sustainability condition, we estimate the magnitude of the cross-section ß coefficient in the cointegration relation of each panel using the Cross Correlated Effects (CEE) and the Common Correlated Effects Mean Group (CCE-MG) estimation procedures developed by Pesaran (2006) CCE-MG estimations. In econometric terms, we are interested in the magnitude of the ß coefficient in the cointegration relations in the panel and in each cross-section. We choose the CEE-MG approach for the following reasons: First, it allows for cross-section dependence which is required according to our CD-test results. Second, it is robust to the presence of a limited number of 'strong' factors and an infinite number of 'weak' factors. According to Eberhardt (2012: 65), the latter can be "associated with local spillovers effects", whereas the former can represent global shocks "such as the recent financial crisis". Third, the CCE estimator accounts for the presence of unobserved heterogeneity (Eberhardt and Presbitero 2013: 10). Since we examine post fiscal equalization data, we have indication for spillovers as well as shocks that affect the panel as a whole. Therefore this test seems to be more appropriate as compared to other "first generation" panel cointegration tests that neither allow for unobserved common factors with heterogeneous impact nor cross-sectional dependence.

The cointegration regression is augmented with cross-section means of the dependent variable and observed regressors. In analogy to Afonso and Rault (2013), we estimate cross-section averages of the dependent variable for revenues and expenditures.

(5) $$Rev_{it} = \alpha_i + \beta_i Exp_{it} + \mu_1 \overline{Rev}_t + \mu_2 \overline{Exp}_t + \mu_{it}$$

Let $Rev_{it}$ and $Exp_{it}$ be the revenues and expenditures in state i at time t, respectively, while $\overline{Rev_t}$ and $\overline{Exp}_t$ denote the cross-section averages of expenditures and revenues in time t, respectively.

Table 13 depicts the results for each sub-panel and the full West German panel. However, the estimates for the full panel need to be taken with a great deal of caution since panel cointegration cannot be assumed for the panel as a whole (see section 4). Thus, these results are only reported for reasons of comprehensibility and completeness and are not discussed in detail.

Table 13 Panel CCE-MG and Laender CCE Estimates, West German Panel and Sub-panels

| | ß | t-Stat | $\mu_1$ | t-Stat | $\mu^2$ | t-Stat | α | t-Stat |
|---|---|---|---|---|---|---|---|---|
| West German panel, excluding Berlin (N=10) | 0.818 | 13.02 | 0.991 | 3.70 | -0.852 | -3.72 | 0.002 | 0.14 |
| Baden-Wuerttemberg | 0.636 | 6.23 | 0.353 | 6.98 | -0.216 | -2.34 | 0.022 | 4.15 |
| Bavaria | 0.815 | 9.76 | 0.303 | 3.33 | -0.270 | -2.49 | 0.015 | 1.87 |
| Bremen | 1.061 | 15.66 | 3.147 | 14.98 | -2.701 | -11.58 | -0.074 | -3.45 |
| Hamburg | 0.927 | 10.54 | 1.148 | 4.78 | -1.370 | -8.12 | 0.047 | 2.37 |
| Hesse | 0.866 | 18.87 | 0.611 | 10.44 | -0.558 | -9.27 | 0.008 | 1.24 |
| Lower Saxony | 0.954 | 10.98 | 0.553 | 8.69 | -0.646 | -5.50 | 0.016 | 2.65 |
| North Rhine-Westphalia | 0.739 | 8.06 | 0.718 | 11.55 | -0.651 | -9.08 | 0.018 | 2.59 |
| Rhineland-Palatinate | 0.933 | 25.01 | 0.646 | 11.73 | -0.602 | -8.93 | -0.001 | -0.08 |
| Saarland | 0.363 | 2.17 | 1.581 | 9.88 | -0.689 | -2.92 | -0.043 | -2.74 |
| Schleswig-Holstein | 0.887 | 10.27 | 0.850 | 13.71 | -0.809 | -10.54 | 0.006 | 1.05 |
| Panel 1: BY, BW, HE, HH, NI, NW, SH (N=7) | 0.741 | 15.87 | 0.999 | 4.13 | -0.746 | -3.80 | 0.000 | 0.06 |
| Baden-Wuerttemberg | 0.517 | 5.81 | 0.485 | 8.67 | -0.185 | -1.97 | 0.016 | 3.42 |
| Bavaria | 0.849 | 10.44 | 0.448 | 3.84 | -0.401 | -3.12 | 0.011 | 1.24 |
| Hamburg | 0.774 | 11.18 | 2.328 | 11.10 | -1.774 | -11.64 | -0.013 | -0.85 |
| Hesse | 0.863 | 20.55 | 0.873 | 14.42 | -0.720 | -11.29 | -0.004 | -0.87 |
| Lower Saxony | 0.698 | 8.52 | 0.721 | 10.24 | -0.441 | -3.44 | -0.006 | -1.08 |
| North Rhine-Westphalia | 0.668 | 8.03 | 0.959 | 12.46 | -0.755 | -9.40 | 0.008 | 1.16 |
| Schleswig-Holstein | 0.821 | 8.86 | 1.177 | 17.23 | -0.949 | -10.00 | -0.010 | -1.74 |
| Panel 2: HB, SL, RP (N=3) | 0.664 | 3.61 | 0.991 | 2.29 | -0.703 | -2.18 | 0.002 | 0.16 |
| Bremen | 0.820 | 10.77 | 1.753 | 38.13 | -1.345 | -8.49 | -0.021 | -2.53 |
| Rhine-Palatinate | 0.875 | 25.97 | 0.252 | 7.03 | -0.328 | -7.56 | 0.023 | 4.23 |
| Saarland | 0.280 | 3.39 | 0.967 | 25.59 | -0.437 | -6.03 | 0.005 | 0.091 |
| Panel 1b: BW, HE, NI, NW, SH (N= 5) | 0.757 | 12.71 | 0.991 | 8.65 | -0.756 | -5.18 | 0.001 | 0.02 |
| Baden-Wuerttemberg | 0.525 | 6.13 | 0.628 | 11.09 | -0.257 | -2.71 | 0.013 | 3.14 |
| Hesse | 0.826 | 18.06 | 0.997 | 14.71 | -0.770 | -11.04 | -0.005 | -1.04 |
| Lower Saxony | 0.790 | 7.34 | 0.904 | 13.49 | -0.681 | -3.97 | -0.003 | -0.54 |
| North Rhine-Westphalia | 0.783 | 10.20 | 1.101 | 13.21 | -0.952 | -11.71 | 0.007 | 0.98 |
| Schleswig-Holstein | 0.860 | 11.56 | 1.324 | 17.23 | -1.118 | -13.80 | -0.007 | -1.30 |

Note: We use the Stata routine xtmg.

The results for **panel 1** indicate a panel cointegration coefficient below one which is smaller compared to the full West German panel. This is further evidence that the West German panel cannot be interpreted in a meaningful way. Cross-section ß coefficients are in a bandwidth between 0.5 and 0.9. This provides evidence, that the long-run relation is smaller than one in the cross-sections. Hence, strict fiscal sustainability can be rejected – however a significant and stable long-run relation exists. The results for Hamburg show a negative constant (α) and high magnitude of the means for revenues ($\mu_1$) and expenditures ($\mu_2$). The finding suggests to exclude Hamburg from panel 1.

While we do not have evidence for I(1) in panel 2, we present the corresponding ß estimates in order to compare the results with other panel estimations. The cointegration coefficient for panel 2 (0.664) is smaller than in panel 1 (0.741). We refrain from interpreting the results because of evidence against panel cointegration in Table 11. Excluding them from the full West German panel was however a meaningful step.

Results in panel 1b reaffirm the choice of the identification strategy that allows for estimating sub-panels with similar time series test results: Significance levels are increased in most cross-sections while the coefficients' magnitudes are slightly changed. We refrain from interpreting these results by ranking them. Instead, we conclude that the West German Laender can be divided into two panel groups: Panel 1 includes a group of at least "weakly sustainable" Laender such as Bavaria, Baden-Wuerttemberg, Hesse, Hamburg, Lower Saxony, North Rhine-Westphalia and Schleswig-Holstein. Panel 2 comprises Laender (HB, RP and SL) that are not sustainable since they do not share a long-term equilibrium relation.

## 6. Summary of Empirical Findings

The results of the panel time series analysis are briefly summarized in Table 14. We have analyzed fiscal sustainability of West German Laender with a post fiscal equalization database that comprises expenditures and revenues. In a first step, we have found evidence for cross-sectional dependence (CD) in the West German Laender and in the sub-panels that are selected conditional on their time series properties. This evidence suggests to apply "second generation" tests. In a second step, we have estimated panel unit root tests and panel cointegration tests in the West German Laender panel and corresponding sub-panels. Third, we have estimated panel and cross-section cointegration coefficients for each panel and explored common correlation effects.

We have evidence that expenditures and revenues of the West German Laender as a whole (excluding Berlin) are not cointegrated, while cointegration can be assumed for sub-panel 1 (BY, BW, HE, HH, NI, NW, SH) and sub-panel 1b (BW, HE, NI, NW, SH). Panel 2 (HB, SL, RP) fails in the panel cointegration analysis and is, thus, considered to have unsustainable public finances. Our estimation results indicate that the cointegration coefficient is between 0.5 and

0.9 in the panels and in the corresponding cross-section. This is evidence that the strict condition for fiscal sustainability (i.e. the coefficient is equal to one) is not met in any panel. Instead we have abundant evidence for weak sustainability in panel 1 and panel 1b, respectively. These two panels pass every step of the panel cointegration analysis.

Table 14 Summary of "Second Generation" Panel Time Series Analysis

| | CD | I(1) | Panel cointegration of expenditures and revenues with fixed short term dynamics of… | | Cointegration coefficient **β** of… | | Verdict |
|---|---|---|---|---|---|---|---|
| | | | …the whole panel | …at least one cross-section | …the whole panel | …each cross-section | |
| Panel 1 | Yes | Yes | Yes | Yes | 0.741 | β < 1 | Weak sustainability |
| Panel 1a | No | Yes | n.a. | n.a. | n.a. | n.a. | n.a. |
| Panel 1b | Yes | Yes | Yes | Yes | 0.75 | β < 1 | Weak sustainability |
| Panel 2 | Yes | (Yes) | No | No | (0.664) | (β < 1) | No sustainability |
| West German panel | Yes | Yes | No | Yes | (0.818) | (β < 1)* | Ambiguous results |

Note: 'n.a.' indicates that no cointegration test is applied due to lack of indication for cointegration. Results in parentheses have to be taken with a great deal of caution and are, thus, reported for reason of comparison, only. *For Bremen ß is 1.061.

While Potrafke and Reischmann (2014) also use post fiscal equalization data, they estimate the Bohn-Model using OLS regression and find evidence for "fiscal sustainability" in ten West German Laender between 1975 and 2010.[9] We, however, find evidence for systematically overshooting expenditures in (panel) time series from 1950 until 2011, allowing for cross-dependence and cross-section heterogeneity. In fact, we do not have evidence for a long-term relation among all (West German) Laender. This evidence questions the efficacy of the German fiscal equalization scheme: It has not significantly contributed to the harmonization of Laender finances such that a conjoint fiscal equilibrium is significant among (West German) Laender.

## 7. Conclusion

The study contributes to the existing literature in two ways: First, we introduce an identification strategy for panel cointegration tests that connects evidence from cross-section specific time series with panel cointegration analysis. We conclude that panel estimations conditional on Laender specific cointegration tests increase the robustness of the evidence of panel cointegration tests. This test strategy is meaningful for application to cross-dependent panels such as federal systems. Second, we use a unique dataset that covers a period of up to 62 years to provide new evidence for the fiscal sustainability of German Laender. Since we find evidence

[9] In addition, we cast doubt on the robustness of their OLS regression since we have found abundant evidence for I(1) of public debt in all West German states (See also Burret et al. 2014), cross-dependence among German Laender finances, and evidence for the application of Laender specific lag lengths in panel regressions. Potrafke and Reischmann (2014) do not control for any of these panel characteristics.

of cross-dependence among Laender in almost all panels, it is required to apply “second generation” panel techniques. This, however, has not been applied to sub-national public finance datasets so far. The existing empirical literature on fiscal sustainability in multi-level jurisdictions has to be reviewed in this regard.

The economic upshot is that all West German Laender fail to obtain strict fiscal sustainability in the panel cointegration analysis. In particular, we provide empirical evidence that public finances in Bremen, Rhineland-Palatinate and Saarland are not sustainable. The other West German Laender (BY, BW, HE, NI, NW, SH) meet some requirements for weak fiscal sustainability.

## References


Afonso, A. (2005), ‘Fiscal Sustainability: The Unpleasant European Case’, *Public Finance Analysis* 61, 19–44.

Afonso, A. and C. Rault (2013), ‘Multi-Step Analysis of Public Finances Sustainability’, CESifo Area Conference on Macro, Money and International Finance, 22-23 February, Munich, Last access April 1st 2014. http://www.cesifo-group.de/portal/page/portal/CFP_CONF/CFP_CONF_2013/Conf-mmi13-De%20Grauwe/Paper/mmi13_Rault.pdf.

Artis, M. and M. Marcellino (1998), ‘Fiscal Solvency and Fiscal Forecasting in Europe’, *CEPR Discussion Papers* 1836.

Artis, M. and M. Marcellino (1999), ‘Fiscal Forecasting: the Track Record of the IMF, *OECD, and EC’, CEPR Discussion Papers* 3098.

Baltagi, B.H. (2013), *Econometric Analysis of Panel Data*, John Wiley and Sons, Chichester.

Bohn, H. (2008), ‘The Sustainability of Fiscal Policy in the United States’, in: R. Neck and J. Sturm (eds.), *Sustainability of Public Debt*, MIT Press, Cambridge/London, pp. 15–49.

Bravo, A. and A. Silvestre (2002), ‘Intertemporal Sustainability of Fiscal Policies: Some Tests for European Countries’, *European Journal of Political Economy* 18, 517–528.

Burret, H. T. and L. P. Feld (2013), ‘Fiscal Institutions in Germany’, *Swiss Journal of Economics and Statistics* 149, 249–290.

Burret, H. T., L. P. Feld and E. A. Köhler (2013), ‘Sustainability of Public Debt in Germany - Historical Considerations and Time Series Evidence’, *Journal of Economics and Statistics* 233, 291–335.

Burret, H. T., L. P. Feld and E. A. Köhler (2014), ‘Fiscal Sustainability of the German Laender in Time Series Analysis’, Unpublished Manuscript, Walter Eucken Institute, Freiburg.

Claeys, P., R. Ramos and J. Suriňach (2008), 'Fiscal Sustainability across Government Tiers', *International Economics and Economic Policy* 5, 139–163.

Eberhardt, M. (2012), 'Estimating Panel Time-Series Models with Heterogeneous Slopes', *The Stata Journal* 12, 61–71.

Eberhardt, M. and A. Presbitero (2013), ’This Time They Are Different: Heterogeneity and Nonlinearity in the Relationship Between Debt and Growth’, *IMF Working Paper* 13/248.

Feld, L. P. and J. Schnellenbach (2013), ‘Verzerrungen im bundesstaatlichen Finanzausgleich‘, *Gutachten im Auftrag des Freistaats Bayern und des Landes Hessen*, Freiburg.

Fève, P. and P. Hénin (2000), ‘Assessing Effective Sustainability of Fiscal Policy within the G-7’, *Oxford Bulletin of Economic Research* 62, 175–195.

Fincke, B. and A. Greiner (2011), 'Debt Sustainability in Germany: Empirical Evidence for Federal States', *International Journal of Sustainable Economy* 3, 235–254.

Garcia S. und P. Hénin (1999): 'Balancing Budget Through Tax Increases or Expenditure Cuts: Is It Neutral?’, *Economic Modelling* 16, 591–612.

Greiner, A. and G. Kauermann (2007), ‘Sustainability of US Public Debt: Estimating Smoothing Spline Regressions’, *Economic Modelling* 24, 350–364.

Greiner, A. and G. Kauermann (2008), ‘Evidence for Germany and Italy Using Penalized Spline Smoothing’, *Economic Modelling* 25, 1144–1154.

Greiner, A. and W. Semmler (1999), ‘An Endogenous Growth Model with Public Capital and Government Borrowing’, *Annals of Operations Research* 88, 65–79.

Greiner, A., W. Semmler and W. Zhang (2005), *Monetary and Fiscal Policies in the Euro-Area: Macro Modelling, Learning and Empirics*, Elsevier, Amsterdam,

Greiner A., U. Köller and W. Semmler (2006), ‘Testing the Sustainability of German Fiscal Policy: Evidence for the Period 1960-2003’, *Empirica* 33, 127–140.

Grilli, V. (1988), ‘Seigniorage in Europe‘, *National Bureau of Economic Research (NBER) Working Paper* 2778.

Herzog, B. (2010), 'Anwendung des Nachhaltigkeitsansatzes von Bohn zur Etablierung eines Frühindikators in den öffentlichen Finanzen', *Kredit und Kapital* 42, 183–206.

Hoechle, D. (2007), ‘Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence’, *The Stata Journal* 7, 281–312.

Kirchgässner, G. and S. Prohl (2008), ‘Sustainability of Swiss Fiscal Policy’, *Swiss Journal of Economics and Statistics* 144, 57–83.

Kitterer, W. (2007), 'Nachhaltige Finanz- und Investitionspolitik der Bundesländer', *Perspektiven der Wirtschaftspolitik* 8, 53–83.

Kitterer, W. and J. Finken (2006), 'Zur Nachhaltigkeit der Länderhaushalte – Eine empirische Analyse', *Finanzwissenschaftliche Diskussionsbeiträge* 7, Finanzwissenschaftliches Forschungsinstitut an der Universität zu Köln.

Koester, G. B. and C. Priesmeier (2013), ‘Does Wagner’s Law Ruin the Sustainability of German Public Finance?’, *Public Finance Analysis* 69, 256–288.

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

Mahdavi, S. and J. Westerlund (2011), 'Fiscal Stringency and Fiscal Sustainability: Panel Evidence from the American State and Local Governments’, *Journal of Policy Modeling* 33, 953–969.

Moscone, F. and E. Tosetti (2009), ‘A Review and Comparisons of Tests of Cross-Section Independence in Panels’, *Journal of Economic Surveys* 23, 528–561.

Newey, W. and K. West (1994), ‘Automatic Lag Selection in Covariance Matrix Estimation’, *Review of Economic Studies* 61, 631–653.

Payne, J. (1997), ‘International Evidence on the Sustainability of Budget Deficits’, *Applied Economics Letters* 12, 775–779.

Persyn, D. and J. Westerlund (2008), ‘Error-Correction–Based Cointegration Tests for Panel Data’, *The Stata Journal* 8, 232–241.

Pesaran, M. H. (2004), ‘General Diagnostic Tests for Cross Section Dependence in Panels', *CESifo Working Paper* 1229; *IZA Discussion Paper No.* 1240.

Pesaran, M. H. (2006), 'Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure', *Econometrica* 74, 967–1012.

Pesaran, M. H. (2007), ‘A Simple Panel Unit Root Test in the Preference of Cross Section Dependence’, *Journal of Applied Econometrics* 22, 265–312.

Pesaran, M. H. (2012), ‘Testing Weak Cross-Sectional Dependence in Large Panels’, *IZA Discussion paper No.* 6432.

Polito, V. and M. Wickens (2011), ‘Assessing the Fiscal Stance in the European Union and the United States 1970-2011’, *Economic Policy* 26, 599–647.

Potrafke, N. and M. Reischmann (2014), 'Fiscal Transfers and Fiscal Sustainability’, *CESifo Working Paper* 4716, Munich.

Sarafidis, V. and T. Wansbeek (2012), ‘Cross-Sectional Dependence in Panel Data Analysis’, *Econometric Reviews* 31, 483–531.

Vanhorebeek, F. and P. van Rompuy (1995), ‘Solvency and Sustainability of Fiscal Policies in the EU’, *De Economist* 143, 457–473.

Westerlund, J. (2007), ‘Testing for Error Correction in Panel Data’, *Oxford Bulletin of Economics and Statistics* 69, 709–748.

Westerlund, J. and S. Prohl (2010), 'Panel Cointegration Tests of the Sustainability Hypothesis in Rich OECD Countries', *Applied Economics* 42, 1355–1364.

## Appendix

### Box A.1 Summary of Time Series Results of the Companion Paper (Burret et al. 2014)

Public debt is not sustainable in most German Laender according to time series results. Exceptions regarding debt are Bavaria, Hesse and Saxony: The KPSS test with trend retains the hypothesis of trend stationarity which is why Bavaria is the first exemption. This evidence is also retrieved from the same test for Hesse. Hesse has never exceeded its starting level in 1956 which contributes to this observation. Saxony has a unique bell shaped debt time series since it has successfully managed to reduce initial debt levels over the course of the last decade. We are, however, reluctant to overestimate evidence from the East German Laender including Berlin. With regard to the three city states (BE, HB, HH) we conclude that they are significantly different with regard to the ADF and PP unit root tests on public debt. Hamburg could be assumed to be sustainable while Bremen is I(1) and Berlin is I(2) if we refer to univariate unit root and stationarity tests on debt. The general observation in all German Laender is that debt has been increasing across time. Revenues and expenditures are, too, not stationary and I(1) in most Laender. Expenditures exceed revenues in most years. In order to further explore the relation between these variables, a VECM is estimated. The results of the cointegration analysis are summarized in the Table below.

Summary of "First Generation" Time Series Analysis of Revenues and Expenditures

| | Stationarity of | | | Cointegration of expenditure and revenue | | | Verdict |
|---|---|---|---|---|---|---|---|
| | debt | expenditure | revenue | Cointegration relation | Cointegration vector [1,-1] | Significant trend | Sustainability |
| | *A* | *B* | *C* | *D* | *E* | *F* | *G* |
| Baden-Wuerttemberg | No | ~ | ~ | ✓ | No | No | Weak |
| Bavaria | ~ | No | No | ✓ | ✓ | No | Strict |
| Bremen | ~ | No | ~ | No | n.a. | n.a. | No |
| Hamburg | ~ | No | ~ | ✓ | ✓ | ✓ | Strict |
| Hesse | ~ | No | ~ | ✓ | No | ✓ | Weak |
| Lower Saxony | ~ | No | No | ✓ | No | ✓ | Weak |
| North Rhine-Westphalia | No | ~ | ~ | ✓ | No | ✓ | Weak |
| Rhine-Palatinate | No | No | No | No | No | n.a. | No |
| Saarland | No | No | ~ | No | n.a. | n.a. | No |
| Schleswig-Holstein | No | No | ~ | ✓ | No | ✓ | Weak |
| Brandenburg | ~ | ~ | No | ✓ | No | ✓ | Weak |
| Mecklenburg-Western Pomerania | No | ~ | ~ | No | n.a. | n.a. | No |
| Saxony | No | ~ | ~ | n.a. | n.a. | n.a. | ~ |
| Saxony-Anhalt | ~ | ~ | ~ | No | n.a. | n.a. | No |
| Thuringia | ~ | ~ | ~ | No | n.a. | n.a. | No |
| Berlin | No | ~ | No | n.a. | n.a. | n.a. | No |

#### Panel identification (see Figure 2 and 3)

Dependent on the existence of one significant cointegration relation (cointegration rank equals one), we have identified two sub-panels: Panel 1 (strictly or weakly sustainable) includes each Land with cointegrated revenues and expenditures. Panel 2 (not sustainable) consists of Laender that have no cointegration relation. The second possibility to be grouped in panel 2 exists if we have evidence for one long-term relation but cannot find any long-term components such as a constant or a trend.

In line with the empirical test procedure we have further subdivided panel 1 into a sub-panel (1a) that includes all Laender that pass the test of this cointegration vector of [1,-1] that is commonly associated with "strict fiscal sustainability". The other sub-panel (1b) includes the Laender that fail to realize the cointegration vector [1,-1] in the corresponding VECM but have at least a constant or a trend in the cointegration relationship. If we find a significant trend in the cointegration relation, be it [1,-1] or not, the

Land is likely not to sustain an ever growing wedge between revenues and expenditures. Thus, the group of "weak sustainable" German Laender has to consolidate public finances.

## Table A.1 Descriptive Statistics of Various Sub-Panels

| | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| All German Laender | | | | | |
| Expenditures | 730 | 0.1458 | 0.0573 | | 0.3226 |
| Revenues | 730 | 0.1367 | 0.0523 | | 0.3144 |
| West German Laender | | | | | |
| Expenditures | 610 | 0.1411 | 0.0585 | 0.0680 | 0.3044 |
| Revenues | 610 | 0.1330 | 0.0544 | 0.0705 | 0.3144 |
| East German Laender | | | | | |
| Expenditures | 120 | 0.1701 | 0.0443 | 0.1245 | 0.3226 |
| Revenues | 120 | 0.1555 | 0.0350 | 0.1265 | 0.2842 |

## Table A.2 Definition and Source of Data

| Variable | Level | Period* | Definition | Source |
|---|---|---|---|---|
| Expenditures and revenues** | Federal Laender (without municipalities) | 1950-1969<br>1970-2011 | Total revenues and total expenditures<br>Total revenues and total expenditures adjusted for payments from the same level. Data in accordance with cash statistics for 2011 and in accordance with final annual accounting otherwise. | Federal Statistical Office |
| Population | Federal Laender | 1950-2011 | End of each year | Federal Statistical Office |
| GDP per capita | Federal level | 1950-2011 | GDP in current prices | Federal Statistical Office |

Note: *Data for Saarland is not available before 1960. Data for East German Laender and whole of Berlin starts in 1992. **1960 is a short fiscal year spanning from April to December. Therefore data has been derived by interpolation and in the case of Saarland by extrapolation. Data is derived by a search request at Germany's Federal Statistical Office.

## Table A.3 Pesaran Panel Unit Root Test of West German Laender with Lag Bandwidth [0,4]

| | Levels | | | | First differences | | | |
|---|---|---|---|---|---|---|---|---|
| | without trend | | with trend | | without trend | | with trend | |
| | Chi-square | p-value | Chi-square | p-value | Chi-square | p-value | Chi-square | p-value |
| Revenues | | | | | | | | |
| Lag 0 | -4.081*** | 0.000 | -4.391*** | 0.000 | -15.158*** | 0.000 | -14.966*** | 0.000 |
| Lag 1 | -2.286** | 0.011 | -2.006 | 0.022 | -13.774*** | 0.000 | -13.420*** | 0.000 |
| Lag 2 | -1.765 | 0.039 | -1.703 | 0.044 | -9.584*** | 0.000 | -8.738*** | 0.000 |
| Lag 3 | -1.219 | 0.111 | -1.237 | 0.108 | -7.420*** | 0.000 | -6.369*** | 0.000 |
| Lag 4 | -1.531 | 0.063 | -1.621 | 0.053 | -6.070*** | 0.000 | -4.921*** | 0.000 |
| Expenditures | | | | | | | | |
| Lag 0 | -3.926*** | 0.000 | -3.881*** | 0.000 | -15.158*** | 0.000 | -14.910*** | 0.000 |
| Lag 1 | -2.676*** | 0.004 | -2.615*** | 0.040 | -13.788*** | 0.000 | -13.458*** | 0.000 |
| Lag 2 | -1.059 | 0.145 | -0.795 | 0.213 | -10.076*** | 0.000 | -9.275*** | 0.000 |
| Lag 3 | -1.264 | 0.103 | -0.995 | 0.160 | -7.950*** | 0.000 | -6.991*** | 0.000 |
| Lag 4 | -0.669 | 0.252 | -0.731 | 0.232 | -5.030*** | 0.000 | -3.889*** | 0.000 |

Note: The null hypothesis of all test is I(1).

Table A.4 Pesaran Panel Unit Root Test of Panel 1 with Lag Bandwidth [0,4]

| | Levels | | | | First differences | | | |
|---|---|---|---|---|---|---|---|---|
| | without trend | | with trend | | without trend | | with trend | |
| | Chi-square | p-value | Chi-square | p-value | Chi-square | p-value | Chi-square | p-value |
| Revenues | | | | | | | | |
| Lag 0 | -3.168*** | 0.001 | -3.645*** | 0.000 | -12.682*** | 0.000 | -12.521*** | 0.000 |
| Lag 1 | -0.757 | 0.225 | -2.184** | 0.011 | -11.888*** | 0.000 | -11.542*** | 0.000 |
| Lag 2 | 0.459 | 0.677 | -0.694 | 0.187 | -8.999*** | 0.000 | -8.194*** | 0.000 |
| Lag 3 | -0.290 | 0.386 | -0.679 | 0.111 | -8.193*** | 0.000 | -7.273*** | 0.000 |
| Lag 4 | -0.113 | 0.455 | 0.495 | 0.463 | -6.279*** | 0.000 | -5.191*** | 0.000 |
| Expenditures | | | | | | | | |
| Lag 0 | -6.880*** | 0.000 | -5.015*** | 0.000 | -12.682*** | 0.000 | -12.521*** | 0.000 |
| Lag 1 | -2.291** | 0.014 | -3.456*** | 0.000 | -11.963*** | 0.000 | -11.546*** | 0.000 |
| Lag 2 | -0.888 | 0.244 | -1.724* | 0.042 | -9.028*** | 0.000 | -8.288*** | 0.000 |
| Lag 3 | -1.221 | 0.249 | -1.565 | 0.059 | -7.984*** | 0.000 | -7.175*** | 0.000 |
| Lag 4 | -0.094 | 0.690 | -0.201 | 0.420 | -5.901*** | 0.000 | -4.929*** | 0.000 |

Note: The null hypothesis of all test is I(1).

Table A.5 Pesaran Panel Unit Root Test of Panel 2 with Lag Bandwidth [0,4]

| | Levels | | | | First differences | | | |
|---|---|---|---|---|---|---|---|---|
| | without trend | | with trend | | without trend | | with trend | |
| | Chi-square | p-value | Chi-square | p-value | Chi-square | p-value | Chi-square | p-value |
| Revenues | | | | | | | | |
| Lag 0 | -2.762*** | 0.003 | -2.618*** | 0.004 | -8.303*** | 0.000 | -8.197*** | 0.000 |
| Lag 1 | -1.840** | 0.033 | -1.717** | 0.043 | -7.451*** | 0.000 | -7.560*** | 0.000 |
| Lag 2 | -1.650* | 0.049 | -1.417* | 0.078 | -5.259*** | 0.000 | -5.451*** | 0.000 |
| Lag 3 | -1.489 | 0.068 | -0.840 | 0.201 | -2.232** | 0.013 | -2.417*** | 0.008 |
| Lag 4 | -2.344 | 0.010 | -2.164 | 0.015 | -2.024** | 0.021 | -2.256** | 0.012 |
| Expenditures | | | | | | | | |
| Lag 0 | -2.946*** | 0.002 | -1.892** | 0.029 | -8.303*** | 0.000 | -8.197*** | 0.000 |
| Lag 1 | -2.415*** | 0.008 | -1.526* | 0.064 | -6.387*** | 0.000 | -6.366*** | 0.000 |
| Lag 2 | -1.798** | 0.036 | -0.893 | 0.186 | -3.349*** | 0.000 | -3.114*** | 0.001 |
| Lag 3 | -2.202** | 0.014 | -1.534* | 0.062 | -1.485* | 0.069 | -0.906 | 0.182 |
| Lag 4 | -3.429*** | 0.000 | -3.180*** | 0.001 | -1.014 | 0.155 | -0.642 | 0.260 |

Note: The null hypothesis of all test is I(1).

Table A.6 Pesaran Panel Unit Root Test of Panel 1b with Lag Bandwidth [0,4]

| | Levels | | | | First differences | | | |
|---|---|---|---|---|---|---|---|---|
| | without trend | | with trend | | without trend | | with trend | |
| | Chi-square | p-value | Chi-square | p-value | Chi -square | p-value | Chi-square | p-value |
| Revenues | | | | | | | | |
| Lag 0 | -3.666*** | 0.000 | -5.002*** | 0.000 | -10.719*** | 0.000 | -10.582*** | 0.000 |
| Lag 1 | -1.332* | 0.091 | -0.887 | 0.188 | -10.423*** | 0.000 | -10.312*** | 0.000 |
| Lag 2 | -0.160 | 0.436 | 1.055 | 0.854 | -7.224*** | 0.000 | -6.673*** | 0.000 |
| Lag 3 | -0.686 | 0.246 | 0.401 | 0.656 | -5.505*** | 0.000 | -4.892*** | 0.000 |
| Lag 4 | -0.945 | 0.172 | 0.497 | 0.689 | -4.517*** | 0.000 | -3.765*** | 0.000 |
| Expenditures | | | | | | | | |
| Lag 0 | -4.365*** | 0.000 | -3.905*** | 0.000 | -10.719*** | 0.000 | -10.582*** | 0.000 |
| Lag 1 | -3.058*** | 0.001 | -1.877** | 0.030 | -9.967*** | 0.000 | -9.595*** | 0.000 |
| Lag 2 | -1.756** | 0.036 | -0.480 | 0.316 | -7.337*** | 0.000 | -6.930*** | 0.000 |
| Lag 3 | -2.110** | 0.017 | -0.526 | 0.299 | -6.065*** | 0.000 | -5.672*** | 0.000 |
| Lag 4 | -1.351* | 0.088 | 0.235 | 0.593 | -4.601*** | 0.000 | -4.163*** | 0.000 |

Note: The null hypothesis of all test is I(1).