

Jewell, T., Lee, J., Tieslau, M., & Strazicich, M.C. (2003). Stationarity of Health Expenditures and GDP: Evidence from Panel Unit Root Tests with Heterogeneous Structural Breaks. *Journal of Health Economics*, 22(2): 313-323 (March 2003). Published by Elsevier (ISSN: 0167-6296). [http://dx.doi.org/10.1016/S0167-6296\(02\)00122-4](http://dx.doi.org/10.1016/S0167-6296(02)00122-4)

Stationarity of health expenditures and GDP: evidence from panel unit root tests with heterogeneous structural breaks

Todd Jewell, Junsoo Lee, Margie Tieslau, and Mark C. Strazicich

ABSTRACT

This paper re-examines the stationarity of national health care expenditures and GDP in a panel setting utilizing data from 20 OECD countries over the period from 1960 to 1997. Previous research in this area has recognized the drawback of not allowing for structural breaks in their unit root tests and noted that their empirical results may not be robust. We advance the literature by utilizing a recently developed panel LM unit root test that allows for heterogeneous level shifts. In contrast to previous analyses that did not consider breaks, our results reject the unit root null hypothesis for both series.

1. INTRODUCTION

This paper analyzes the time series properties of per-capita real health care expenditures and per-capita real GDP (hereafter HE and GDP, respectively) in a panel setting using recently advanced econometric techniques in panel unit root tests. Much of the early research on this subject found a strong positive correlation between HE and GDP in developed countries. In general, earlier studies found that the overwhelming majority of the variation in HE can be explained by variation in per-capita GDP (Newhouse, 1977, Parkin et al., 1987 and Gerdtham and Jonsson, 1991). Recently, health care economists have devoted a great deal of attention to analyzing the time series pattern of HE and GDP.

When examining the relationship between HE and GDP it is important to determine whether or not these variables are stationary. Empirical tests that ignore this issue can lead to spurious regressions and meaningless results. Most previous research suggests that both HE and GDP are non-stationary. For example, Hansen and King (1996) and Blomqvist and Carter (1997) find that HE and GDP are non-stationary.¹ Recently, researchers have begun to favor panel-based unit root tests due to the increase in power that panel tests offer. This increase in power stems from greater degrees of freedom and the inclusion of heterogeneous cross-country information not available in univariate tests. In particular, McCoskey and Selden (1998) use the panel unit root test of Im et al. (2003) and determine that both HE and GDP are stationary. Hansen and King (1998) question McCoskey and Selden's preference for omitting a time trend and note that this may lead to biased results in unit root tests. This outcome may be especially important given the positive trend in HE. Gerdtham and Lothgren (2000) challenge the findings of McCoskey and Selden using a battery of tests, which include a time trend, and conclude that both HE and GDP are non-stationary. The contradiction in these results may be due to the inclusion or exclusion of a time trend.²

Another, perhaps more important, shortcoming of the previous research is failure to allow for structural breaks. McCoskey and Selden (1998) and Gerdtham and Lothgren (2000) both recognize this drawback, but note that a suitable panel unit root test allowing for structural breaks was not available at the time of their analysis. These authors note that further research that allows for structural breaks should be undertaken in order to properly investigate the stationarity of HE and GDP. We respond to this call by employing a robust test that allows for heterogeneous structural breaks in both univariate and panel frameworks. The remainder of the paper is organized as follows. In Section 2, we discuss the issue of allowing for structural change in panel unit root tests. In Section 3, we describe the panel LM unit root test methodology. Section 4 reports our empirical findings and Section 5 concludes.

2. STRUCTURAL BREAKS IN PANEL UNIT ROOT TESTS

It is now well known that a structural break can be mistaken for non-stationarity (Perron, 1989). Therefore, it is possible that previous findings of non-stationarity in HE and GDP may be due to failure to allow for structural breaks. For example, Gerdtham and Lothgren (2000) do not consider structural breaks in their panel unit root tests and recognize the drawback.

Nonetheless, the task of allowing for structural breaks in the existing panel unit root tests, such as those proposed by Im et al. (2003) and Levin et al. (2002), would be quite difficult to implement. This is due to the fact that the distribution of these panel tests with structural breaks will critically depend on nuisance parameters indicating their location, as noted by Im et al. (2002). As such, it would be extremely difficult, if not impossible, to control for the numerous possible combinations of heterogeneous structural breaks that might occur when using these panel unit root tests.

To provide a remedy, we apply the panel LM unit root test recently developed by Im et al. (2002), which is derived from the univariate one-break LM unit root test developed by Amsler and Lee (1995). Unlike conventional panel unit root tests, the distribution of the panel LM test statistic is invariant to break point nuisance parameters.³ As such, the critical values that apply to the panel LM unit root test *without* breaks are also valid for the panel LM unit root test *with* breaks, regardless of the location and number of breaks in each cross-section.

In applying the Im et al. (2002) test, several important features are utilized and worth noting. First, we allow for each country to have unique fixed effects, differing time trend coefficients, and varying persistence parameters. Second, we allow for the number of structural breaks to vary by country. Third, we allow for heterogeneous break points, which are endogenously determined for each country. Fourth, we allow for time-specific fixed effects to capture any common year structural breaks. Fifth, we allow for the optimal number of augmentation terms in the unit root tests to be heterogeneous and determined jointly with the breaks.

3. THE PANEL LM UNIT ROOT TESTS WITH BREAKS

3.1. The panel data model with breaks

The question of interest in our analysis is whether or not a panel data series, Y_{it} , is stationary. To address this question, we consider the following model:

(1)

$$Y_{it} = \delta_i' Z_{it} + e_{it}, \quad e_{it} = \beta_i e_{i,t-1} + e_{it},$$

where “ i ” indexes cross-section units (i.e. countries); “ t ” indexes time periods (i.e. years); Z_{it} is a vector of exogenous variables; δ_i is the corresponding parameter vector; and e_{it} is the error term of the process. The parameter β_i is used to test the unit root null hypothesis, and e_{it} is a zero-mean error term that allows for heterogeneous variance structure across cross-section units, but assumes no cross-correlations.⁴ Similar to the Im et al. (2003) panel test, β_i allows for heterogeneous measures of persistence. As noted previously, when the data generating process follows Eq. (1), the resulting critical values of the panel unit root test will be invariant to δ_i .

Now consider the case where the underlying model for Y_{it} allows for a structural break in levels.⁵ In this case, the vector of exogenous variables Z_{it} takes the form $[1, t, D_{it}]'$, where D_{it} is a dummy

variable that denotes the time when a structural break occurs. That is, if we define the location at which the break occurs for country “ i ” as “ T_{Bi} ”, then the dummy variable takes the form $D_{it}=1$ if $t > T_{Bi}$, and 0 otherwise. In addition, this framework can easily be extended to allow for two structural breaks by defining Z_{it} as $[1, t, D_{1it}, D_{2it}]'$, where D_{1it} is a dummy variable that captures the first break and D_{2it} the second break, respectively. Note that this model has the advantage of allowing for heterogeneous structural breaks that vary by country, as distinguishable from common shocks that apply uniformly to all countries.

Another important feature of this model is that it allows for structural breaks under both the null and alternative hypotheses. To see this, suppose that $\beta_i=1$ for all “ i ” so that:

(2)

$$Y_{it} = \delta_{1i} + \delta_{2i}t + \delta_{3i}D_{it} + e_{it}, \quad e_{it} = e_{it,t-1} + e_{it}.$$

Solving this equation for e_{it} and applying the differencing transformation, we obtain:

(3)

$$\Delta e_{it} = Y_{it} - Y_{i,t-1} - \delta_{2i} - \delta_{3i}[D_{it} - D_{i,t-1}].$$

Suppose we define $B_{it} \equiv [D_{it} - D_{i,t-1}]$ (so that B_{it} takes on a value of 1 for the time period $T_{Bi}+1$, and 0 at all other times) and let $\Delta e_{it} \equiv v_{it}$. In this case, Eq. (3) can be re-written so that the model under the null becomes:

(4)

$$Y_{it} = \delta_{2i} + \delta_{3i}B_{it} + Y_{i,t-1} + v_{it}.$$

Allowing for the one-time break term “ B_{it} ” under the null is important since omitting this term may lead to invalid results and spurious rejections (see Perron, 1989, Nunes et al., 1997 and Lee and Strazicich, 2001).

3.2. The panel LM unit root statistic

The first step in computing the panel LM unit root test statistic is to compute univariate LM unit root test statistics for each country. To that end, we begin with:

(5)

$$\Delta Y_{it} = \delta_i' \Delta Z_{it} + \phi_i \tilde{S}_{i,t-1} + u_{it}$$

where ΔY_{it} and ΔZ_{it} are the first-differenced values of Y_{it} and Z_{it} , respectively; $\tilde{S}_{i,t-1}$ is the detrended value of $Y_{i,t-1}$ (following Schmidt and Phillips, 1992); and u_{it} is the model’s stochastic disturbance term. The presence of a unit root in Y_{it} for country “ i ” implies that $\phi_i=0$. It follows that the univariate LM test statistics can be computed using the t -test that tests $H_0: \phi_i=0$ (implying a

unit root and non-stationary) versus $H_1: \varphi < 0$ (implying no unit root and stationary) for each country. The t -test for each country is denoted by $\tilde{\tau}_i$.

Following Im et al. (2002), the null and alternative hypotheses, respectively, in the panel test are given by $H_0: \varphi = 0$ (implying a unit root for all countries) versus $H_1: \varphi < 0$ (implying that one or more countries reject the unit root). The panel LM test statistic is computed by averaging the optimal univariate LM unit root t -test statistics estimated for each country, denoted as LM_i^T , as follows:

(6)

$$\overline{LM}_{NT} = \frac{1}{N} \sum_{i=1}^N LM_i^T.$$

A standardized panel LM unit root test statistic is then constructed by letting $E(L_T)$ and $V(L_T)$ denote the expected value and variance of LM_i^T , respectively, under H_0 , and computing:

(7)

$$\Gamma_{LM} = \frac{\sqrt{N} [\overline{LM}_{NT} - E(L_T)]}{\sqrt{V(L_T)}}.$$

Numerical values for $E(L_T)$ and $V(L_T)$ are provided in Table 1 of Im et al. (2002). In the presence of heterogeneous autocorrelated errors, we use a weighted average of $E(L_T(k_i))$ and $V(L_T(k_i))$ to be consistent with the country-specific optimal values of k_i .

Table 1
LM unit root tests on real per-capita health care expenditures, 1960–1997

Country	Univariate LM unit root test statistic	Optimal number of breaks	Optimal lag length (k)	Break location(s)
Australia	-3.658*	2	0	1976, 1989
Austria	-2.996	1	8	1977
Belgium	-2.211	0	7	–
Canada	-2.758	1	1	1980
Denmark	-2.557	0	7	–
Finland	-3.529**	1	1	1990
France	-1.517	0	1	–
Germany	-3.920**	1	8	1990
Greece	-3.613**	1	8	1982
Iceland	-2.240	0	7	–
Ireland	-2.472	1	1	1975
Italy	-3.361*	1	6	1975
Japan	-3.099**	0	5	–
Netherlands	-1.976	0	0	–
Norway	-2.356	0	1	–
Spain	-0.957	2	1	1978, 1982
Sweden	-3.499**	0	7	–
Switzerland	-3.921**	1	3	1983
United Kingdom	-1.990	1	7	1979
United States	-2.433	1	1	1973
Panel LM test statistic	-17.333***			

All tests allow for time fixed effects and all regressions include an intercept and time trend. The 1, 5, and 10% critical values for the LM unit root test with no break are: -3.63, -3.06, and -2.77. The 1, 5, and 10% critical values for the minimum LM test with one break are: -4.239, -3.566, and -3.211. The 1, 5, and 10% critical values for the minimum LM test with two breaks are: -4.545, -3.842, and -3.504, respectively. The 1, 5, and 10% critical values for the panel LM unit root test (with or without breaks) are: -2.326, -1.645, and -1.282.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

The panel LM unit root test has several attractive features. First, the distribution of the test statistic depends on “ N ” and “ T ” but does not depend on any other parameters under the null hypothesis. Similar to the Im et al. (2003) panel unit root test, the asymptotic distribution of the panel LM unit root test is standard normal. Most importantly, the distribution of the panel LM test is unaffected by the presence of break(s). This so called “invariance result” holds for any finite number of breaks. As such, it is unnecessary to simulate new critical values for $E(L_T)$ and $V(L_T)$ as the number and location of breaks varies in each country. In contrast, this feature is absent from conventional panel unit root tests, such as those of Im et al. (2003) and Levin et al. (2002).

As noted above, the panel LM test assumes no cross-correlations in the error terms across countries. However, controlling for time-specific effects common to all countries will mitigate any cross-sectional dependencies in the data (Hakkio, 1984).⁶ Allowing for common time effects may be especially important in the context of this paper, since it is likely that health expenditures in each country are contemporaneously correlated. While Im et al. (2002) does not consider

time-fixed effects in their panel test, it is unnecessary to simulate new critical values for $E(L_T)$ and $V(L_T)$ when including these effects.⁷

4. EMPIRICAL RESULTS

4.1. Data and implementation

Annual data from 1960 to 1997 on logged per-capita real health care expenditures and per-capita GDP (in 1990 dollars), for 20 OECD countries, were obtained from the OECD Health Data File (1998). The data is consistent with that used in Gerdtham and Lothgren (2000).⁸

To implement our unit root tests, we must determine both the number and location of structural breaks in each country. At the same time, we will determine the optimal number of augmentations to correct for autocorrelated errors. We begin by employing the univariate two-break minimum LM unit root test of Lee and Strazicich (2003), the one-break minimum LM test of Lee and Strazicich (1999), and the no-break LM test of Schmidt and Phillips (1992). From the outcome of these results, we determine the optimal number of breaks in each country. For each of these tests, we wish to determine the optimal break location(s) (T_{Bi}) and the optimal number of augmentations jointly and endogenously from the data. The augmentation terms $\Delta \tilde{S}_{it-j}$ to be included in (5), where $j = \{1, 2, \dots, k_i\}$, correct for the presence of autocorrelated disturbances. The implementation procedure is as follows. First, for each possible combination of two, one, or no breaks, the optimal value of k_i is determined using the general-to-specific methodology suggested by Perron (1989) and Ng and Perron (1995).⁹ Second, to endogenously determine the location of each break, we employ the univariate two-break and one-break minimum LM unit root tests as suggested in Lee and Strazicich, 2003 and Lee and Strazicich, 1999, using the optimal value of k_i given for each location of breaks. They propose a grid search procedure that determines the location of breaks in each country (T_{Bi}), where $\tilde{\tau}_i$ is at a minimum (i.e. the most negative); hence, the test is referred to as a “minimum LM test”.¹⁰ The minimized value of $\tilde{\tau}_i$ is denoted as “LM_{*i*}”.¹¹

Third, after endogenously determining the location of the break(s), we next determine the optimal number of breaks, using a general-to-specific search approach. To do so, the t -statistic of each estimated break coefficient is examined for significance at the 10% level in an asymptotic normal distribution (i.e. absolute value greater than 1.645). If less than two breaks is significant, the procedure is repeated using the one-break minimum LM unit root test. If no break is significant, then the no-break LM unit root test is employed. In this manner, the location of breaks, the number of breaks, and the number of lagged augmentation terms (k_i) are jointly and endogenously determined for each country.¹²

One final consideration is to allow for the possibility of country-specific and time-specific fixed effects in the panel test. Country-specific fixed effects are incorporated quite naturally within the framework of the panel LM test. This is due to the fact that the t -statistics used to calculate the panel test statistic are estimated with heterogeneous intercept terms in the univariate LM tests. To allow for time-specific fixed effects, we “demean” the data by subtracting the average value

of Y_t from each Y_{it} in each year " t ". We note that the inclusion of time-specific effects in the panel setting is equivalent to allowing for a structural break in each year that is common (homogeneous) to all OECD countries. After including common time effects, the remaining structural breaks are those specific (heterogeneous) to individual countries.

4.2. Test results

The results of the LM unit root tests on HE and GDP, allowing for possible time-fixed effects, are shown in Tables 1 and 2, respectively. To allow for the possibility of a trend stationary alternative, all unit root tests include both an intercept and time trend. The univariate LM unit root test statistics appear in the second column of each table. The optimal number of breaks is shown in the third column of each table. The optimal number of lagged differenced terms that correct for serial correlation is given in the fourth column, and the location of the breaks appears in the last column. The last row of each table presents the panel LM test statistic.

Table 1 displays the unit root test results on health expenditures. Upon examination of the univariate test results, we find eight rejections of the null at the 10% level of significance or better. A structural break is found to exist in 12 out of 20 countries, with all but three occurring during the period from 1973 to 1983. This preponderance of break points during the 1970s may reflect technological advances during this period that lead to large increases in the cost of medical care.¹³ In addition to technology, the business cycle might significantly impact health expenditures, since the government plays a major role in the financing of HE in most OECD countries. As such, the structural breaks in HE can be potentially influenced by significant fluctuations in the business cycle that directly impact government revenues. Close examination of the break points seems to support this conjecture, since many of the break points in HE are consistent with recessionary periods. For example, Australia has a break in 1989 preceding a recession in 1990; Austria has a break in 1977 following a recession in 1975; Canada has a break in 1980 preceding a recession in 1982; Finland has a break in 1990 during a year of zero economic growth and preceding a 3-year recession that began in 1991; Italy has a break and recession in 1975; Spain has one break in 1978 preceding a recession in 1979, and a second break in 1982 following a recession in 1981; Switzerland has a break in 1983 following a recession in 1982; the United Kingdom has a break in 1979 preceding a recession in 1980–1981; and the US has a break in 1973 preceding a recession in 1974–1975. With regard to the panel unit root test, we find a strong rejection of the unit root null hypothesis. It stands to reason that this strong rejection result is obtained from the increased power of the panel test.

Table 2 displays the unit root test results on GDP. In 11 of 20 cases, the univariate tests reject the unit root at the 10% level of significance or better. In addition, structural breaks are found to exist in seven countries, three of which contain two breaks. Most breaks are again associated with recessions. For example, Australia has a break and recession in 1982; Belgium has a break and recession in 1993; and Iceland has a break and recession in 1990. In other cases the break year is close to a recession. For example, Greece has a break in 1974 and recession in 1975; Italy has a break in 1981 and recession in 1983; and Spain has a break in 1982 and recession in 1981. With regard to the panel unit root test, we again find a strong rejection of the

unit root null. This result lends further support to the notion that the power of the unit root test is significantly increased in the panel setting.¹⁴

Table 2
LM unit root tests on real per-capita GDP, 1960–1997

Country	Univariate LM unit root test statistic	Optimal number of breaks	Optimal lag length (<i>k</i>)	Break location(s)
Australia	-3.031	2	8	1978, 1982
Austria	-2.173	0	3	–
Belgium	-2.725	2	8	1991, 1993
Canada	-6.559***	0	0	–
Denmark	-2.820*	0	8	–
Finland	-2.681	0	8	–
France	-6.438***	0	0	–
Germany	-3.488	2	8	1974, 1986
Greece	-3.951**	1	0	1990
Iceland	-6.657***	0	0	–
Ireland	-1.439	1	3	1981
Italy	-1.471	0	3	–
Japan	-5.303***	1	0	1988
Netherlands	-2.897*	0	8	–
Norway	-6.941***	0	0	–
Spain	-2.163	1	8	1982
Sweden	-6.508***	0	0	–
Switzerland	-6.823***	0	0	–
United Kingdom	-2.681	0	3	–
United States	-5.471***	0	0	–
Panel LM test statistic	-15.821***			

All tests allow for time fixed effects and all regressions include an intercept and time trend. The 1, 5, and 10% critical values for the LM unit root test with no break are: -3.63, -3.06, and -2.77. The 1, 5, and 10% critical values for the minimum LM test with one break are: -4.239, -3.566, and -3.211. The 1, 5, and 10% critical values for the minimum LM test with two breaks are: -4.545, -3.842, and -3.504, respectively. The 1, 5, and 10% critical values for the panel LM unit root test (with or without breaks) are: -2.326, -1.645, and -1.282.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Overall, our findings support the notion that both HE and GDP are stationary processes with, in many cases, one or two permanent shifts in levels. These findings differ sharply from those of Hansen and King (1996), Blomqvist and Carter (1997), and Gerdtham and Lothgren (2000), who did not consider structural breaks in their tests. In addition, since our unit root tests allow for trending data, our findings of stationary HE and GDP is unlikely due to inclusion or exclusion of a time trend.

5. CONCLUSION

The issue of whether or not national HE is stationary has serious implications for researchers and policy makers. Previous research has, in general, concluded that HE is non-stationary. These findings may be called into question, however, since no previous empirical research has properly allowed for the possibility of structural breaks in unit root tests on HE. We address this shortcoming by analyzing HE (and GDP) using a panel LM unit root test that allows for heterogeneous structural breaks, thus significantly increasing the power of the test.

Our results strongly support the notion that HE and GDP are stationary around one or two breaks. The structural breaks are identified most often around recessions. Our findings are in sharp contrast to those of Hansen and King (1996), Blomqvist and Carter (1997), and Gerdtham and Lothgren (2000) who conclude that HE and GDP are non-stationary, but do not consider structural breaks in their unit root tests. In addition, while our findings are consistent with those of McCoskey and Selden (1998), their model may not be correctly specified as they do not allow for either a time trend or structural breaks in their unit root tests. An important implication of our findings is that researchers and policy makers modeling health expenditures and GDP in a panel regression framework can get meaningful results that are not spurious, if structural changes are allowed.

NOTES

1. Hansen and King use ADF unit root tests while Blomqvist and Carter employ Phillips and Perron (1988) tests as well as the panel unit root test of Levin et al. (2002). These findings indicate that previous analyses of HE and GDP that assume stationarity may be invalid.
2. Since the data sets used in the above mentioned papers are nearly identical, any differences in unit root test results is not likely due to differences in the data. In particular, Hansen and King (1996) and McCoskey and Selden (1998) use data from 20 OECD countries for the period 1960–1987. Blomqvist and Carter (1997) use data from the same 20 OECD countries for the period 1960–1991. Gerdtham and Lothgren (2000) use data from 21 OECD countries for the period 1960–1997.
3. Both the Im et al. (2003) and the Levin et al. (2002) tests are Dickey–Fuller-type tests, and as such their distributions will depend on break point nuisance parameters, when allowing for breaks.
4. Allowing for time-fixed effects can mitigate possible cross-correlations in the error terms. We address this issue in Section 3.2.
5. As in Im et al. (2002), we consider only the “crash” model that allows for a permanent change in level.

6. O'Connell (1998) suggests that allowing for time-fixed effects may not control for all cross-correlations in the error terms. Adopting his suggested GLS-based test would be more complicated in our case due to the incidental parameter problem of having to estimate too many parameters of cross-covariances when " N " is big. Given that asymptotic properties of O'Connell's test are not known, the effect of including structural breaks in this setting would add even more uncertainty to our results and, therefore, this approach will not be considered in our paper.

7. Im et al. (2003) showed that allowing for time-fixed effects does not change the critical values of terms similar to $E(L_T)$ and $V(L_T)$. This same result holds for the panel LM test with time-fixed effects, with or without structural breaks.

8. To maintain a balanced panel, we omit New Zealand and Portugal, which have missing observations, and we include Japan.

9. Beginning with a maximum number of eight lagged terms, $\max k=8$, the last augmented term with $k=8$ is examined to see if it is significant at the 10% level in an asymptotic normal distribution (critical value is 1.645). If insignificant, the maximum lagged term is dropped and the model re-estimated using $k=7$ terms, etc. until the coefficient of the last lagged term is significant, at which point the procedure stops. If no significant lagged terms are discovered, then the search selects $k=0$.

10. The grid search is performed over the interval $[0.1 T, 0.9 T]$ to eliminate end points. In addition, this methodology can easily be extended to the case of allowing for two breaks, T_{1B_i} and T_{2B_i} . The grid search is repeated at each possible combination of two break points to locate the minimum test statistic.

11. An important feature of the endogenous break LM unit root test is that unlike the competing Dickey–Fuller-type endogenous break tests, the minimum LM test is not subject to spurious rejections in the presence of a break under the null (Nunes et al., 1997 and Lee and Strazicich, 2001). Once the break locations are identified, we consider them as given in the panel tests. Thus, we do not consider a minimum test version of the panel LM test and our procedure is recursive in nature. Since critical values are invariant to the location of the break(s), it is unnecessary to simulate new critical values for different combinations of the break points.

12. One might suggest extending the above test procedure to allow for more than two structural breaks in each cross-section. Although the invariance results for the LM unit root test should hold for any finite number of breaks, the finite sample properties with more than two breaks have not been examined. In addition, the computational burden would significantly increase when searching all combinations of more than two break points in conjunction with identifying the number of lagged augmented terms.

13. One might have reasonably expected to find a structural break in US health expenditures during the mid 1960s, reflecting the introduction of Medicare. However, since our data set begins in 1960, and our methodology tests for breaks over only the middle 80% of the sample, this likely did not leave enough "pre-Medicare" data for such a break to be discovered.

14. To examine the robustness of our results, we additionally performed all of the univariate and panel unit root tests without demeaning for common time effects. The inferences drawn from these results were not qualitatively different from those displayed in Table 1 and Table 2. These results have been omitted to conserve space, but are available from the authors upon request.

REFERENCES

- Amsler, C., Lee, J., 1995. An LM test for a unit-root in the presence of a structural change. *Econometric Theory* 11, 359–368.
- Blomqvist, A.G., Carter, R.A.L., 1997. Is health care really a luxury? *Journal of Health Economics* 16, 207–229.
- Gerdtham, U.-G., Jonsson, B., 1991. Conversion factor instability in international comparisons of health care expenditure. *Journal of Health Economics* 10, 227–234.
- Gerdtham, U.-G., Lothgren, M., 2000. On stationarity and cointegration of international health expenditure and GDP. *Journal of Health Economics* 19, 461–475.
- Hansen, P., King, A., 1996. The determinants of health care expenditure: a cointegration approach. *Journal of Health Economics* 15, 127–137.
- Hansen, P., King, A., 1998. Health care expenditure and GDP: panel data unit root test results—comment. *Journal of Health Economics* 17, 377–381.
- Hakkio, C., 1984. A re-examination of purchasing power parity. *Journal of International Economics* 17, 265–277.
- Im, K.S., Lee, J., Tieslau, M., 2002. Panel LM Unit Root Tests with Level Shifts. Mimeo, Department of Economics, University of Central Florida.
- Im, K.S., Pesaran, M., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, in press.
- Lee, J., Strazicich, M., 1999. Minimum LM Unit Root Test. Mimeo, Department of Economics, University of Central Florida.
- Lee, J., Strazicich, M., 2001. Break point estimation and spurious rejections with endogenous unit root tests. *Oxford Bulletin of Economics and Statistics* 63, 535–558.
- Lee, J., Strazicich, M., 2003. Minimum LM unit root test with two structural breaks. *Review of Economics and Statistics*, in press.
- Levin, A., Lin, C.-F., Chu, C.J., 2002. Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics* 108, 1–24.
- McCoskey, S.K., Selden, T.M., 1998. Health care expenditures and GDP: panel data unit root test results. *Journal of Health Economics* 17, 369–376.

Newhouse, J.P., 1977. Medical care expenditures: a cross-national survey. *Journal of Human Resources* 12, 115–125.

Ng, S., Perron, P., 1995. Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90, 269–281.

Nunes, L., Newbold, P., Kuan, C., 1997. Testing for unit roots with breaks: evidence on the great crash and the unit root hypothesis reconsidered. *Oxford Bulletin of Economics and Statistics* 59, 435–448.

O'Connell, P., 1998. The overvaluation of purchasing power parity. *Journal of International Economics* 44, 1–19.

OECD Health Data File, 1998. *OECD Health Care Data: A Software Package for International Comparison of Health Care Systems*. OECD, Paris.

Parkin, D., McGuire, A., Yule, B., 1987. Aggregate health care expenditures and national incomes: is health care a luxury good? *Journal of Health Economics* 6, 109–127.

Perron, P., 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57, 1361–1401.

Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. *Biometrika* 75, 407–436.

Schmidt, P., Phillips, P.C.B., 1992. LM tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics* 54, 257–287.