Health expenditure and gross domestic product: causality analysis by income level

Rezwanul Hasan Rana[1,3] · Khorshed Alam[1] · Jeff Gow[1,2]

Abstract
The empirical findings on the relationship between gross domestic product (GDP) and health expenditure are diverse. The influence of income levels on this causal relationship is unclear. This study examines if the direction of causality and income elasticity of health expenditure varies with income level. It uses the 1995–2014 panel data of 161 countries divided into four income groups. Unit root, cointegration and causality tests were employed to examine the relationship between GDP and health expenditure. Impulse-response functions and forecast-error variance decomposition tests were conducted to measure the responsiveness of health expenditure to changes in GDP. Finally, the common correlated effects mean group method was used to examine the income elasticity of health expenditure. Findings show that no long-term cointegration exists, and the growth in health expenditure and GDP across income levels has a different causal relationship when cross-sectional dependence in the panel is accounted for. About 43% of the variation in global health expenditure growth can be explained by economic growth. Income shocks affect health expenditure of high-income countries more than lower-income countries. Lastly, the income elasticity of health expenditure is less than one for all income levels. Therefore, healthcare is a necessity. In comparison with markets, governments have greater obligation to provide essential health care services. Such results have noticeable policy implications, especially for low-income countries where GDP growth does not cause increased health expenditure.

✉ Rezwanul Hasan Rana
   rezwanul_54@yahoo.com; rezwanul.rana@usq.edu.au

   Khorshed Alam
   Khorshed.alam@usq.edu.au

   Jeff Gow
   gowj@usq.edu.au

[1] School of Commerce, University of Southern Queensland, Toowoomba, Australia

[2] School of Accounting, Economics and Finance, University of KwaZulu-Natal, Durban, South Africa

[3] School of Commerce, University of Southern Queensland, Room: T230, West Street, Toowoomba, QLD 4350, Australia

JEL Classification C55 · I10 · I15 · O1

Introduction

Health expenditure is increasing substantially relative to gross domestic product (GDP) growth in almost all countries at all income levels (Mladenović et al. 2016). This increase in expenditure has become a major concern for governments and policymakers (Panopoulou and Pantelidis 2012). Health expenditure has increased from 3% of global GDP in 1948 to 7.9% in 1997 (Self and Grabowski 2003) to approximately 10% in 2014 (World Health Organization 2016). Health expenditure is growing faster than GDP in numerous countries.

Therefore, the primary objective of this study is to examine if the causal relationship between GDP and health expenditure varies with income level. From a policy perspective, understanding the causal relationship between these two variables is pivotal in helping governments and policymakers implement policies that ensure sustainable health systems (Atilgan et al. 2016).

Baltagi et al. (2017) investigated this relationship to examine the size of income elasticity across low- and high-income countries as well as geopolitical regions.

'Despite the extensive literature on the topic, few studies so far have offered a global perspective, using a very large sample of countries. One limitation of these studies is that they have treated countries under the strong assumptions of homogeneity and cross-section independence'.

The current study addresses these limitations by conducting appropriate tests and estimation techniques to improve the robustness of the results. It contributes to the existing literature by extending the findings of Baltagi et al. (2017), with the use of heterogeneous panel and Toda–Yamamoto causality tests as well as by investigating the impact of negative shocks in the model with impulse-response functions (IRF) and forecast-error variance decomposition (FEVD) tests. Furthermore, it investigates the income elasticity of health expenditure using GDP and health expenditure data expressed at constant 2011 PPP prices via the common correlated effects (CCE) mean group approach. The findings indicate that the statistical outcomes fluctuate significantly as the definition of the variables changes (whether measured at constant 2005 PPP versus constant 2011 PPP) even if identical estimation methods are used.

This study utilises a comprehensive dataset and the latest econometric techniques to examine the causal relationship at different levels of income with a large panel data comprising 161 countries. It aims to answer the following questions: (i) Does a long-term cointegration relationship exists between health expenditure and GDP at all income levels? (ii) Does the direction of causality vary with changes in income levels? (iii) To what extent can future variations in the growth rate of health expenditure be explained by economic growth rates? (iv) Does significant heterogeneity exist in the income elasticity of health expenditure for countries at different income levels?

Previous studies have not agreed that increasing income is a major determinant of rising health expenditure across all income levels and that a causal relationship exists between the two (Acemoglu and Johnson 2007; Amiri and Ventelou 2012). Previous literature also provides mixed conclusions on whether health care is a necessity or a luxury (Baltagi et al. 2017). Past research has focused on OECD countries (Baltagi and Moscone 2010; Hartwig 2008). A few exceptions examined causality for low-income and developing countries (Chen et al. 2013; Ke et al. 2011). Some investigated numerous countries from all income groups using a cross-sectional design (Farag et al. 2013; van der Gaag and Stimac 2008). Noticeably,

these cross-sectional studies did not examine causality and overlooked the unit root problem in the GDP and health expenditure data. Thus, Villaverde et al. (2014) suggested the grouping of countries based on their distinct characteristics, such as per capita income, to examine the association between GDP and health expenditure.

High economic growth rates are broadly associated with high health expenditure rates (Farag et al. 2013; Hansen and King 1996; Hartwig 2008; Wang 2009). However, the direction of causality has been subject to debates. Previous studies provided mixed results for using diverse datasets, assumptions and estimation techniques for countries across income groups. Halici-Tuluce et al. (2016) summarised the findings of the selected literature that explores this relationship and showed contradictions in the direction of causality. In addition, existing literature is ambiguous as to the direction of GDP growth and rising health expenditure for countries at different income levels. Chen et al. (2013) and Ke et al. (2011) concluded that the rate of health expenditure growth vary at different levels of economic development. However, little is known as to how this causal relationship changes with income levels. Long-range, macro-level databases enable comparisons among various income groups (Carrion-i-Silvestre 2005).

To the best of the authors' knowledge, this study is one of the few works to address these issues by pooling data from an extensive group of countries over time. It has the following contributions to the literature: (i) comparing the variations in the causal relationship between health expenditure and GDP for different income groups; (ii) using appropriate statistical analysis techniques, such as Westerlund's cointegration test and Dumitrescu–Hurlin's approach to Granger noncausality test; (iii) employing IRF and FEVD to understand the causal relationship from a new perspective; and (iv) analysing the variations in findings if cross-section dependence and heterogeneity is taken into account in the panel data whilst conducting unit root, cointegration and causality tests.

Method

Conceptual framework

Firstly, the long-term relationship between health expenditure and national income based on the Keynesian macroeconomic model is represented by

$$\mathrm{Y}_{it} = C_{it} + U_{it}$$

where Y_{it} indicates the national income (GDP) of country i at time t; C_{it} indicates the total expenditure on health of country i at time t; U_{it} represents other macroeconomic factors, including consumption expenditure (except health), investment, government purchase and net exports of country i at time t. Rearranging the equation according to the theory of consumption results in

$$\mathrm{Y}_{it} = \hat{C}_i + \gamma_i C_{it} + U_{it}$$

where $\hat{C}_i$ indicates the autonomous consumption of health expenditure, and γ_i is the income elasticity of health expenditure. A long-term panel cointegration relationship can be formulated with the following panel VECM type model (Wang 2011):

$$\Delta X_{it} = \partial_{0i} + \sum_{j=1}^{n} \partial_{1i} \Delta X_{it-j} + \emptyset_i \left(\mathrm{Y}_{it-1} - \hat{C}_i - \gamma_i C_{it-1} \right) + \in_{it}$$

where $\Delta X_{it} = (\Delta Y_{it}, \Delta C_{it})$; the error correction terms (ECM) $= \left(Y_{it-1} - \hat{C}_i - \gamma_i C_{it-1} \right)$, and parameter $\emptyset_i$ is the speed of adjustment in the variables. Thus, if $\emptyset_i$ is significant, then the short-term disequilibrium may adjust to the long-term equilibrium through the ECM process (Wang 2011).

If the absence of cointegration hypothesis in the preliminary cointegration test is not rejected, then a panel vector autoregressive (VAR) approach will be considered (Clarke and Mirza 2006; Dolado and Lütkepohl 1996). Hence, the following system linear equation will be used with n variate homogenous panel VAR of order m (Abrigo and Love 2016)

$$Y_{it} = Y_{it-1}\theta_1 + Y_{it-2}\theta_2 + \cdots + Y_{it-p}\theta_p + X_{it}\varphi + \in_{it}$$
$$[i = 1, 2, \ldots, N]; \quad [t = 1, 2, \ldots, T_i]$$

where Y_{it} is the $(1 \times n)$ vector of dependent variables; X_{it} is the $(1 \times m)$ vector of exogenous covariates; $\in_{it}$ represents the error-terms. The parameters to be estimated are the $(n \times n)$ matrices $\theta_1, \theta_2, \ldots \theta_p$ and the $(m \times n)$ matrix φ (Abrigo and Love 2016).

Estimation and procedures

Many recent empirical studies have used panel data to investigate the long-term cointegration relationship between health expenditure and GDP with various statistical assumptions and techniques (Hall et al. 2011; Okunade and Karakus 2001; Tamakoshi and Hamori 2015). Although most studies have found a long-term association between GDP and health expenditure, several studies concluded that no cointegration relationship exists (Granados 2012; Hansen and King 1996). Halici-Tuluce et al. (2016) outlined the contradictory findings on the relationship between health expenditure and GDP. To the best of the author's knowledge, no study examining the long-run association among the variables have used the panel cointegration test developed by Westerlund (2007). According to Persyn and Westerlund (2008), the test can efficiently account for heterogeneity, in short- and long-term dynamics of a cointegration relationship as well as for cross-section dependence. The test does not enforce any common-factor condition and examines the null hypothesis of no cointegration by testing if the error-correction term in the panel model is equal to zero.

Panel data analysis is often subject to heterogeneity and cross-section dependence, and the latter can lead to misleading results (McCoskey and Selden 1998; Gengenbach et al. 2006). Nonetheless, many previous studies failed to account for these issues. Therefore, this study used appropriate estimation techniques to control for cross-section dependence and unobserved heterogeneity to come up with robust results on the relationship between health expenditure and GDP.

In addition to the Granger causality and cointegration tests, this study also uses IRF and FEVD tests. IRF measures the effect of a shock to a predictor variable on the predicted variable (Koop et al. 1996), whilst FEVD measures the strength of the causal relationship by investigating variations in the values of one variable that can be explained by the other variable (Shahbaz 2012). FEVD is insensitive to the order of the variables in the VAR system, thus providing a good understanding of the depth of the relationship. No previous study has used IRF and FEVD to analyse the link between rising health expenditure and rising income at different national income levels. These new approaches will also enhance knowledge on the relationship between health expenditure and GDP. The robustness of the results is again buttressed by employing three unit root and Granger causality tests with varying assumptions.

Panel unit root test

Pesaran ([2007](#)) introduced the cross-sectional augmented IPS (CIPS) unit root test which is applicable for dynamic panels with cross-section dependence as well as serial correlation. CIPS unit root test provides consistent inferences in contrast to other tests that do not consider cross-section dependence (Baltagi and Moscone [2010](#)). The Harris–Tsavalis (HT) ([1999](#)) and Im–Pesaran–Shin (IPS) ([2003](#)) unit root tests were also employed to verify the robustness of the results. These tests provided the option (demean) to mitigate the challenges of cross-section dependence. The results of the HT and IPS tests are presented in Table [6](#) in "Appendix".

Panel cointegration test

An error-correction-based panel cointegration test developed by Westerlund ([2007](#)) was used to measure long-term relationship. The approach analyses the null hypothesis of no cointegration by examining if the error-term in a conditional panel is equal to zero. The cointegration considering the cross-sectional dependency assumes the following form (Persyn and Westerlund [2008](#)):

$$\Delta lHE_{it} = \delta_i d_t + \alpha_i lHE_{i,t-1} + \varphi_i lGDP_{i,t-1} + \sum_{j=1}^{p_i} \alpha_{ij} \Delta lHE_{i,t-j} + \sum_{j=q_i}^{p_i} \alpha_{ij} \Delta lGDP_{i,t-j} + \varepsilon_{it}$$

where $\varphi_i = -\alpha_i \beta_i$. In addition, α_i defines the speed at which the system corrects itself back to the equilibrium relationship of $HE_{i,t-1} - \beta_i GDP_{i,t-1}$ following an unanticipated shock. The bootstrap method was chosen as the p values are robust in case the data have cross-sectional dependence, and this method was performed using the Akaike Information Criterion (AIC) for selecting the optimal lag and lead length. Furthermore, the Bartlett kernel window width was calculated based on the formula $4(T/100)^{2/9}$ (Persyn and Westerlund [2008](#)) which is approximately equal to three.

Granger causality test

Dumitrescu and Hurlin Granger noncausality test A heterogeneous panel causality test proposed by Dumitrescu and Hurlin (DH) ([2012](#)) was employed to determine the robustness of the results. The heterogeneity of the panel data reflects the appropriateness of DH panel noncausality test. The test focused on noncointegrated bi-variate variables of stationary characteristics and fixed-effect panel model (Liddle and Messinis [2015](#)). For Granger causality, this test allows for heterogeneity of the casual relationship and heterogeneity of the regression model (Dumitrescu and Hurlin [2012](#)).

This study investigated the following DH Granger noncausality equations:

$$\Delta lHE_{i,t} = \propto + \sum_{k=1}^{K} \beta_{ik} \Delta lHE_{i,t-k} + \sum_{k=1}^{K} \emptyset_{ik} \Delta lGDP_{i,t-k} + \in_{i,t}$$

$$\Delta lGDP_{i,t} = \propto + \sum_{k=1}^{K} \beta_{ik} \Delta lGDP_{i,t-k} + \sum_{k=1}^{K} \emptyset_{ik} \Delta lHE_{i,t-k} + \in_{i,t}$$

where $lGDP_{i,t}$ indicates the log of per capita gross domestic product and $lHE_{i,t}$ indicates the per capita health expenditure for country i in period t, respectively, and coefficients are allowed to vary across individual countries. However, these coefficients are assumed to be

time-invariant (Dumitrescu and Hurlin 2012). The test also requires a balanced panel dataset and similar lag order K for all countries. The finite sample properties of the test indicate reliable results with small values of T and N. The formula $T > 5 + 2X$ was used to determine the minimum number of lags, where T is the number of time periods and X the number of lags which signified the minimum time needed for each number of lags (Menard and Weill 2016). Only the Z-bar tilde results were presented in the results section because they are appropriate for panel data with small T (Dumitrescu and Hurlin 2012).

Toda–Yamamoto approach to Granger noncausality The Toda and Yamamoto (TY) (1995) test examines causality in the VAR models at levels that reduce the probability of inaccurate identification of the order of integration. TY also avoids the difficulty of investigating Granger causality based on the power and size attributes of stationary and cointegration tests. This causality test uses a modified Wald test to estimate zero restrictions on the parameters of the VAR (n) model, and it has an asymptotic (Chi-sq) distribution with n degrees of freedom (Adriana 2014). Similar to Amiri and Ventelou (2012), this study used the following VAR system to apply the TY version of the Granger noncausality test:

$$\Delta lHE_t = \alpha_0 + \sum_{k=1}^{n} \alpha_{1k} \Delta lHE_{t-k} + \sum_{j=n+1}^{d_{max}} \alpha_{2j} \Delta lHE_{t-j}$$
$$+ \sum_{k=1}^{n} \partial_{1k} \Delta lGDP_{t-k} + \sum_{j=n+1}^{d_{max}} \partial_{2j} \Delta lGDP_{t-j} + \omega_{1t}$$

$$\Delta lGDP_t = \beta_0 + \sum_{k=1}^{n} \beta_{1k} \Delta lGDP_{t-k} + \sum_{j=n+1}^{d_{max}} \beta_{2j} \Delta lGDP_{t-j}$$
$$+ \sum_{k=1}^{n} \emptyset_{1k} \Delta lHE_{t-k} + \sum_{j=n+1}^{d_{max}} \emptyset_{2j} \Delta lHE_{t-j} + \omega_{2t}$$

Moreover, Adriana (2014) indicates that if $\partial_1 = \text{vec}\ (\partial_{11}, \partial_{12}, \partial_{13, \dots} \partial_{1n})$ is the vector of the first n VAR coefficient, then the null hypothesis of *lGDP* does not cause *lHE* which is structured as follows: Ho: $\partial_{1k} = 0, k = 1, \dots, n$ and vice versa. The lag length n was obtained using AIC and BIC lag selection criterion.

Impulse response function and variance decomposition

To measure the evolution of economic shocks, the VAR analysis often leads to the calculation of IRF and FEVD which are the essential parts of the VAR method originally proposed by Sims (1980). The orthogonalised IRFs in the VAR model examine the sensitivity of the dependent variable to shocks to each of the variables (Rafiq et al. 2009). A unit shock is carried out for each variable in the equation, and the effects on the VAR model are presented. Using Choleski decomposition to orthogonalise the covariance matrix in the VAR model reduces the problem of contemporaneous correlation among the variables (Swanson and Granger 1997). The simple impulse function with Choleski decomposition proposed by Sims (1980) is

$$y_t = \sum_{i=0}^{\infty} \emptyset_i v_{t-i} \quad [\emptyset_0 = I_k \text{ is the } (K \ \times \ K) \text{ identity matrix}]$$

$$\text{and } \emptyset_i = \sum_{j=1}^{i} \emptyset_{i-j} A_j \quad [i = 1, 2, \ldots]$$

where $\emptyset_i$ are interpreted as impulse responses of the system; $A_j = 0$ for $j > p$ (for a k dimensional VAR (p) process); v_t represents the orthogonal residuals (Swanson and Granger 1997). In addition, The IRFs has no causal interpretation, but it measures the probability of a shock on one variable to impact on the other variable(s). In addition, the decomposition is not unique but influenced by the ordering of the variables (Abrigo and Love 2016). Variance decomposition explains the fraction of changes in the dependent variable due to their individual shocks (Rafiq et al. 2009).

The g-step ahead forecast error equation used in this study is

$$Y_{it+g} - E\big[Y_{it+g}\big] = \sum_{k=0}^{g-1} e_{i(t+g-i)} \emptyset_i$$

where Y_{it+g} is the observed vector at time $t + g$; $E\big[Y_{it+g}\big]$ is the g-step ahead predictor vector made at time t; the orthogonalised shocks $e_{it} K^{-1}$ (with K matrix) have a covariance matrix I_k (Abrigo and Love 2016).

CEE mean group approach

The CCE mean group estimator proposed by Pesaran (2006) depends on the heterogeneous slope coefficients assumption (Baltagi et al. 2017) and can account for cross section dependence in the error term (Pesaran and Tosetti 2011; Everaert and De Groote 2016). Moreover, the approach provides consistent estimates in the presence of a restricted number of strong and unlimited weak, common factors (Chudik et al. 2011), irrespective of whether the common factors are stationary or nonstationary (Kapetanios et al. 2011). Pesaran (2006) provides detailed methodological discussion on the CCE mean group approach, and Chudik et al. (2011) and Kapetonios et al. (2011) offer additional extensions of the method.

Data

Data for 161 countries from 1995 to 2014 were collected from the World Development Indicators (World Bank 2016). Per capita GDP and health expenditure were measured at constant 2011 PPP prices. Previous studies used per capita GDP as a proxy for income because of its close relationship with living standards (Lago-Peñas et al. 2013). Noticeably, several countries changed their status from one income level to another during the study period. For simplicity of analysis, a country was considered low income if it mostly had low income during the 20-year period.

This study uses panel data. The key advantages of using panel data include the following: great sample size, multiple observations for each country, controlling a wide range of time-invariant country-specific attributes, being able to relax the assumption of a homogeneous relationship across countries and inclusion of country and time specific effects (Glied and Smith 2011). Moreover, panel data account for any unobserved heterogeneity and allows for estimation of heterogeneous causal effects (Chen et al. 2013). Log-linear functions were used to structure the data as it gives several advantages, including variable parameters representing elasticities; it also assumes diminishing marginal returns for the explanatory variable and

Table 1 Descriptive analysis of the key variables

Variable	Obs	1995	2000	2005	2010	2014
Low-income	39					
HEPC		16.7	15.9	27.6	48.1	58.3
GDPPC		345.8	366.2	585.3	1004.4	1263.6
LnHE		2.62	2.6	3.1	3.6	3.86
LnGDP		5.72	5.7	6.2	6.6	6.88
HEXPU B		1.6	1.7	2.2	2.3	2.3
Lower-middle	51					
HEPC		65.1	79.76	129.4	222.1	295.5
GDPPC		1280.6	1374.7	2168.4	3697.7	4944
LnHE		3.9	4.14	4.7	5.2	3.85.5
LnGDP		6.9	7.1	7.5	8.1	8.3
HEXPUB		2.8	3.0	3.3	3.5	3.9
Upper-middle	31					
HEPC		216.5	249.5	391.5	673.9	768.7
GDPPC		3910	4238.3	6847.6	10,646.1	12,442.7
LnHE		5.1	5.3	5.8	6.4	6.5
LnGDP		8.1	8.2	8.7	9.2	9.4
HEXPUB		3.5	3.4	3.6	4.0	3.9
High-income	40					
HEPC		1516.2	1552.4	2630.9	3384.2	3810.9
GDPPC		21,440	21,987	33,436	39,686	44,605
LnHE		7.1	7.1	7.6	7.8	8.0
LnGDP		9.8	9.88	10.3	10.5	10.6
HEXPUB		4.6	4.7	5.3	5.9	6.1

HEPC and GDPPC values are in USD. HEPC, health expenditure per capita; GDPPC, gross domestic product per capita; LnHE, log of health expenditure per capita; LnGDP, log of gross domestic product per capita; and HEXPUB, total public health expenditure as percentage of total gross domestic product

finally implies an elasticity which is invariant across time and countries at different income levels (Hall and Jones 2007; Shaw et al. 2005).

Per capita health expenditure is the total private and public health expenditure as a ratio of the total population. Per capita GDP is the sum of value added by all resident producers plus any product tax (less subsidies) not included in the valuation of output and net receipts of primary income (compensation of employees and property income) from abroad (World Bank 2016). Table 1 summarises this study's statistical data.

Results

Descriptive statistics on the variables used in the analysis are presented in Table 1 for five different periods of time for four income groups.

Over the years, expenditure on health and the share of public health expenditure to GDP have increased for countries of all income groups. However, the disparity of resources used for healthcare services between rich and poor countries has widened significantly over time.

Table 2 CIPS unit root test with trend and no trend option

Income group	Per capita GDP (trend)		Per capita health expenditure (trend)		Per capita GDP (no trend)		Per capita health expenditure (no trend)	
	Level	1st Diff	Level	1st Diff	Level	1st Diff	Level	1st Diff
Global	− 2.44	− 3.76*	− 2.69	− 4.25*	− 2.46	− 3.50*	− 2.44	− 4.02*
Low-income	− 3.33*	− 4.25*	− 3.20*	− 4.91*	− 2.80*	− 4.28*	− 2.63	− 4.45*
Lower-middle	− 2.49	− 3.83*	− 2.64	− 4.22*	− 2.57	− 3.59*	− 2.69	− 4.09*
Upper-middle	− 2.08	− 3.41*	− 2.45	− 3.94*	− 1.88	− 3.34*	− 2.26	− 3.77*
High income	− 2.10	− 3.55*	− 1.84	− 4.02*	− 1.96	3.23*	− 1.58	− 3.57*

*Means significant at 95% confidence interval to reject the H0 = homogenous non-stationary with critical value − 2.7. AIC lag selection criteria have been used

Unit root test

The CIPS unit root tests were applied with and without trend to test the stationarity of the variables. Table 2 indicates that the panel data of health expenditure and GDP have unit root at levels and stationary at first differences.

Cointegration and causality tests

For the Westerlund cointegration test the asymptotic and the bootstrap *p values* were different for group means and panel statistics at all income levels. The asymptotic *p values* strongly reject the null hypothesis of no cointegration (Table 3). However, the bootstrap *p value* accepted the null hypothesis of the absence of long-term cointegration. The bootstrap method was used to account for cross-section dependence. Hence, this study found no long-term cointegration between per capita health expenditure and per capita GDP when cross-section dependence is accounted for in the panel data.

The Granger causality test results in Table 3 indicate that for low-income countries, no causality exists running from GDP to health expenditure. For lower- and upper-middle income countries unidirectional causality flows from GDP to health expenditure and for high-income countries, the relationship is bidirectional. The results clearly indicate that the direction of causality varies depending on income levels. The DH noncausality test results showed that the direction of causality was consistent for different values of lag (1, 2, 3; Table 8 in "Appendix"). The results contradicted the VAR Granger results in the direction of causality for lower-income countries. The *Z-bar* tilde value was used to measure the direction of causality as the panel data consist of small fixed *T* and relatively large *N*. The test accounted for cross-section dependence in the data. Lastly, the TY causality test results showed complete similarity with the VAR Granger causality test results.

Comprehensive statistical results of all the tests are presented in Tables 7, 8, 9 and 10 in "Appendix". This study predominantly used VAR model for causal estimations after finding no cointegration (Table 3 and "Appendix"), once cross-section dependence in the panel data was accounted for.

Table 4 presents the country-specific analysis for the cointegration and causality tests. The pattern of the results matches with the panel data analysis. The tests found cointegration of 56.25% and causality of 38.6% for all countries. Among the income groups, the association

Table 3 Cointegration and causality tests

Tests	Cointegration test		Granger causality test		
Income groups	Asymptotic results	Bootstrap results	PVAR causality	DH non-causality	TY non-causality
All countries	Cointegrated	Not cointegrated	Bi-directional	Bi-directional	Bi-directional
Low-income	Cointegrated	Not cointegrated	LnHE to LnGDP	Bi-directional	LnHE to LnGDP
Lower-middle	Cointegrated	Not cointegrated	LnGDP to LnHE	LnGDP to LnHE	LnGDP to LnHE
Upper-middle	Cointegrated	Not cointegrated	LnGDP to LnHE	LnGDP to LnHE	LnGDP to LnHE
High-income	Cointegrated	Not cointegrated	Bi-directional	Bi-directional	Bi-directional

All cointegration tests were run with the option trend and bootstrap value = 400. The null hypothesis of the Westerlund cointegration test was H0 = no cointegration. For the Granger causality test lag was selected based on MAIC and MBIC. All results are significant at 95% confidence interval. The null hypothesis of the Granger causality test is H0 = the excluded variable does not Granger cause the equation variable. LnGDP, log of per capita GDP and LnHE, log of per capita health expenditure

Table 4 Country-specific cointegration and VAR granger causality tests

Income groups	Total countries	Cointegration test (% of countries showing a long-run cointegrating relationship)	Granger causality test (% of countries showing causal relation of any direction)
All countries	161	56.25	38.6
Low-income	39	52.5	33.5
Lower-middle	51	53	30
Upper-middle	31	64.5	51
High-income	40	55	40

Only 10.5% of countries showed a bidirectional relationship for the period of 1995–2014

between growth in health expenditure and GDP growth is strongest for upper-middle income countries. Although the Westerlund test proves the absence of long-term cointegration, a country-specific analysis depicts that health expenditure has a long-run relationship with GDP for most countries.

FEVD test

The results in Fig. 1 indicate that approximately 43% of the variation in the growth of health expenditure can be explained by GDP growth. They also show that the percentage of variation in the growth of per capita health expenditure explained by the percentage growth in per capita GDP. Once again, the strongest association is for upper-middle income countries. FEVD standard error and confidence intervals (95%) are based on 200 Monte Carlo simulations (Abrigo and Love 2016).
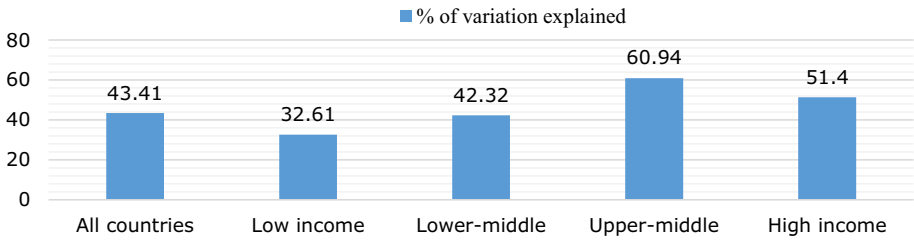
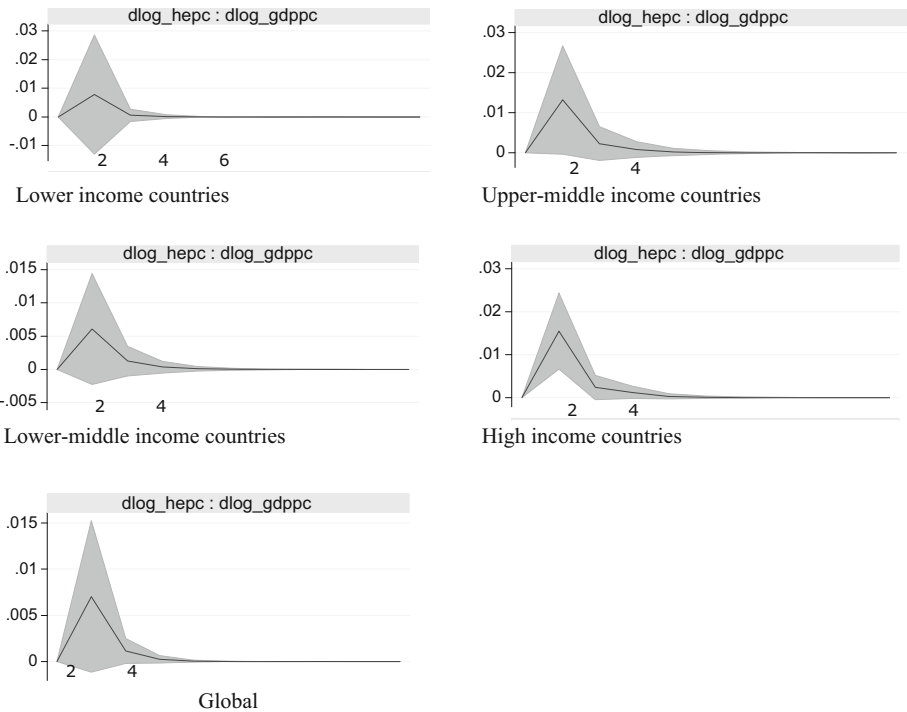**Fig. 1** Forecast-error variance decomposition results at 12th term

Fig. 2 Orthogonalised impulse response functions results

IRF

The orthogonalised IRF presented in Fig. 2 measures sensitivity of the growth in HE to shocks in the growth of GDP in the VAR model. The IRF plot for the global and other income groups' data shows that a positive shock in the growth of per capita GDP leads to a small decrease in the growth in per capita health expenditure. The shock is prominent for high income countries.

CCE mean group approach

Table 5 shows the results from the CCE mean group estimator for four income groups. The results are reported for the bivariate regression model where the log of per capita health

**Table 5** CCE mean group estimation for income classifications

Predictor variable	Low-income	Lower-middle	Upper-middle	High-income
LnGDP	0.65* (0.08) [0.07]	0.88* (0.09) [0.05]	0.93* (0.06) [0.07]	0.73* (0.09) [0.04]
LnGDP (pre GFC)	0.61* (0.11) [0.09]	0.87* (0.10) [0.09]	0.77* (0.21) [0.02]	0.63* (0.18) [0.01]
LnGDP (post GFC)	0.78* (0.28) [0.03]	1.06* (0.24) [0.03]	0.85* (0.09) [0.05]	0.89* (0.18) [0.01]
CDtest (*p* value)	0.000	0.000	0.000	0.000
CORRtest (*p* value)	0.347	0.265	0.769	0.810
HETtest (*p* value)	0.000	0.000	0.000	0.000

Standard errors in parentheses. *Significant at 95% confidence interval with the trend option. Root mean squared errors in the square brackets. GFC means global financial crisis. Pre GFC period (1995–2008) and post GFC period (2009–2014). CDtest indicates the cross-section dependence test developed by Pesaran (2004). The null hypothesis of the cross-section dependence test proposed by Pesaran (2004) is H0 = no cross-section dependence. The H0 is rejected for the global and all income levels. Hence, we conclude that the panel data are cross-sectional dependence in the disturbances. CORRtest indicates Wooldridge test for autocorrelation proposed by Wooldridge (2002) and the null hypothesis is no first order autocorrelation. HETtest is the modified Wald test for groupwise heteroscedasticity under the null hypothesis of homoscedasticity

expenditure is the dependent variable, and the log of per capita GDP is the predictor variable. The estimation produces coefficients on GDP of 0.65, 0.88, 0.93 and 0.73 for low-income, lower-middle, upper-middle and high-income countries, respectively. All the results are significant at the 5% confidence interval. Tests were also conducted for the pre (1995–2008) and post (2009–2014) global financial crisis (GFC). The findings indicate that the income elasticity of health expenditure (coefficient of per capita GDP) increased after the GFC. The table also illustrates the cross-section dependence test, along with test for slope homogeneity and autocorrelation tests in the panel data for all income groups. Lastly, low root mean squared errors indicate the goodness of fit of the estimated models.

Other diagnostic tests for global data

Pesaran's CD test was used to examine cross-sectional independence with H0 = no cross-sectional dependence. The results (*Z value* = 27.168 and *Prob* = 0.000) strongly reject the null hypothesis of no cross-section dependence. Detailed CD test results for each income group are reported in Table 11 in "Appendix". Wooldridge's test for autocorrelation in panel data with H0 = no first-order autocorrelation was then employed. The test for panel autocorrelation accepted the null hypothesis (*Z value* = 2.531; *Prob* = 0.1136). The Breusch–Pagan test was then used to examine heteroscedasticity. The result confirmed that the panel data contains unequal variances. Table 5 represents the autocorrelation and heteroscedasticity test results for individual income groups. Lastly, the stability condition of the estimated panel VAR was investigated using the VAR stability test (Abrigo and Love 2016; Lütkepohl and Krätzig 2004). All the eigenvalues in Fig. 3 are inside the unit circle. Thus, the panel VAR satisfies the condition of stability.

Global data

Lower income

Lower middle-income

Upper middle-income

High income
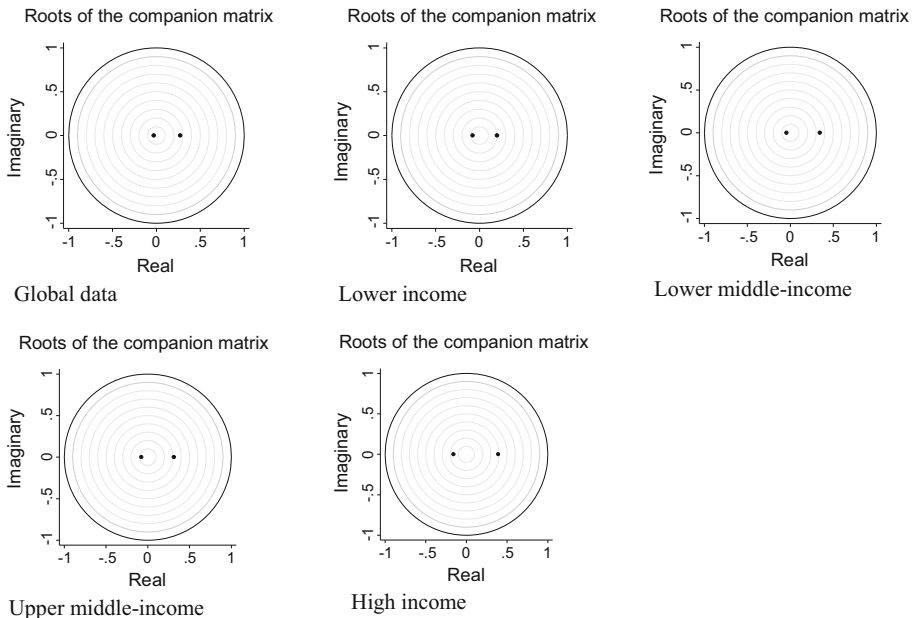
Fig. 3 Panel VAR stability test results

Discussion

Health expenditure and GDP data have unit root at levels except for the low-income countries

The results showed that health expenditure and GDP are mostly nonstationary for the pooled data of 161 countries. Although some country-specific tests might show stationarity, the weight of other countries pulled the grouped data towards the unit root. Similar results are found in other panel studies of MacDonald and Hopkins (2002) and Baltagi and Moscone (2010).

Health expenditure and GDP are not cointegrated in the long term when controlled for cross-section dependence

The results of no long-term cointegration matched the results of Hansen and King (1996) and Granados (2012) but contradicted the findings of Clemente et al. (2004) and Baltagi et al. (2017), who found that health expenditure and GDP have long-term cointegration at the global level. The variation in the cointegration test results may come from data differences. Although the time period is very similar, Baltagi et al. (2017) used constant (USD) 2005 PPP prices, and this study used constant (USD) 2011 PPP prices. Measuring the extent to which data characteristics have changed is difficult due to this six-year gap in the PPP prices.

In estimating the long-term relationship, the asymptotic p value results were not reliable. Therefore, the robust results from the bootstrap approach were preferred. In the health expenditure literature cross-section dependence is not normally accounted for (Moscone and Tosetti 2010), and it must be considered carefully whilst studying long-run relationships (Wang and

Rettenmaier 2007). The failure to reject the null hypothesis with bootstrapped *p* value may be due to the low power of the trace statistics when allowing for cross-section dependence. Thus, the study concluded that no long-run relationship exists between the variables at any income level.

Direction of causality varies for different levels of income

The Granger noncausality test has interesting results. Two-way causality is only established in the case of high-income countries. Other researchers also found two-way causality especially in (high income) OECD countries (Amiri and Ventelou 2012). Again, bidirectional causality is expected as GDP influences health expenditure, and the latter has a strong impact on GDP growth rate forecasting (Mladenović et al. 2016). Amiri and Ventelou (2012) indicated that rising health expenditure improves productivity and economic growth. In addition, healthy people work hard for long periods of time (Bloom et al. 2004). Therefore, a bidirectional causality in high-income countries is logical. All the causality results indicated that growth in GDP causes growth in health expenditure for lower-middle, upper-middle and high-income countries. The result was opposite for lower-middle income countries, where growth in health expenditure does not cause growth in GDP. This result matches with Erdil and Yetkiner (2009) and Baltagi and Moscone (2010) who found that not all countries have bidirectional causality between health expenditure and GDP.

The causality results were mixed for low income countries. The panel VAR Granger test and the TY approach concluded that GDP does not cause health expenditure. The reason may be because GDP growth in low income countries is not substantial enough to support significant health sector development (Chen et al. 2013). However, all three causality tests confirmed that growth in health expenditure causes GDP growth in low income countries. These countries receive substantial amounts of foreign aid for health which might explain this relationship to some extent. The DH Granger noncausality results showed that causality is bi-directional for low-income countries. The result is different because this method considers heterogeneity in the panel data. Thus, it is highly applicable for this study.

The Granger causality test had mixed results, as expected. Wang (2011) stated that the characteristics of a country's health care changes with levels of income. Therefore, the conclusion is that the direction of causality changes based on income levels.

Country-specific study complements the panel data analysis

The results in Table 3 demonstrated that the cointegration and causal relationship are not identical at different income levels. The results vastly differ from the findings of Erdil and Yetkiner (2009) who concluded that 61% of their study countries showed a causal relationship contrasting to only 38.5% in this study. The dissimilarity might be due to the difference in time period, contrasting methodologies and the large set of data used in this study. Nonetheless, the two studies provided the similar conclusion that the association between health expenditure and GDP differs across income levels.

A high proportion of variations in health expenditure growth can be explained by GDP

The FEVD estimation results showed that GDP is a significant predictor of future variation in the growth of health expenditure. The following conditions exist in FEVD: an explained variation in the predicted variable directly caused by the changes in the predictor variable and unexplained variations coming from factors other than predictor variable. More than 50% of the variation in upper-middle and high-income countries can be explained. However, only less than 40% can be explained for other countries. Thus, GDP is likely to explain variations in health expenditure for high income countries.

Any positive shock in GDP growth has a small short-lived negative impact on the growth of health expenditure

The IRF plots showed that, for all income levels, any positive shock in the growth of GDP leads to a very small decrease in the growth of health expenditure, and the effect dissipates roughly after three or four periods. However, the impact of the change in standard deviation or response to shock shows an increasing pattern for high income groups. Therefore, any negative shock to GDP growth has a great impact on rising health expenditure in high income countries.

Income elasticity of health expenditure is less than one for all income levels

The results of the CCE mean group approach provide strong evidence that growth in income has a positive and significant association with health expenditure. The estimated findings also indicate that health care is a necessity rather than luxury for all income levels. The income elasticity of health expenditure is between 0.65 and 0.93 according to the log-linear model. Therefore, increases in health care expenditure are less than the increase in per capita income over the period of 1995–2014, when heterogeneity and cross-section dependence in the panel data were accounted for. Similar results were found in early studies conducted by Baltagi and Moscone (2010); Ke et al. (2011) and Farag et al. (2013). Moreover, the income elasticity of demand increased considerably in the post-GFC period for all income groups. The results indicate that many countries' growth in health expenditure will outdistance growth in income in the near future.

Results of the study are subject to assumptions and estimation techniques

The choice of estimation approaches influences the results of any study. Firstly, the cointegration results changed entirely when the assumption of cross-section dependence was activated. Secondly, the Granger causality test results showed consistency except for low-income countries. According to the DH Granger noncausality test, GDP and health expenditure have bidirectional causality. However, the other causality tests rejected this result and concluded unidirectional causality. The DH approach accounted for heterogeneity in the panel data. Therefore, the results are inconsistent due to the adoption of the different estimation techniques. Lastly, despite using identical statistical estimation the results of this study were marginally different compared to Baltagi et al. (2017) due to variations in the length of panel

data and definition of the variables used. Thus, the panel analysis results related to health expenditure should be interpreted with caution.

The results indicated that growth in health expenditure health expenditure causes GDP growth in low-income countries. These countries often receive foreign aid targeted towards the health sector which may play an important role in 'increasing' income. Much of the literature has concluded that the efficient use of foreign aid promotes economic growth for recipient countries (Asteriou 2009). Moreover, countries proportionately spend their income on health care as the former increases. Therefore, as the proportion of health expenditure increases, middle-income countries should ensure that everyone gets the benefit of rising health care spending equally through sustainable health financing systems.

Conclusions

This study empirically investigated the causal relationship between growth in GDP and health expenditure using a large panel data set for 161 countries over a period of 20 years. These countries were divided into four different income groups. Stationarity, cointegration and causality of the data were investigated to understand the degree of association between GDP and health expenditure. The study used estimation techniques that controlled for cross-section dependence and unobserved heterogeneity.

Results show no long-term cointegration, and the causal relationship is strong at high-income levels once cross-section dependence in the panel is accounted for. The results from the Granger causality test show that the causal relationship can change due to changes in the level of income. Moreover, only 38% of countries showed a causal relationship after accounting for the assumption of heterogeneity in the panel data. FEVD estimates show that GDP strongly explains variations in health expenditure in general, and these variations are even stronger for high-income countries. The IRF results indicate that a positive shock to GDP has a negative effect on health expenditure for a short period of time across all income groups. However, the magnitude of impact is larger for high-income countries. Lastly, the income elasticity of health expenditure is less than one for all income levels, indicating health care is a necessity goods. Hence, instead of the market, governments should have an important role to play in ensuring the development and affordability of their health sector.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Appendix

See Tables 6, 7, 8, 9, 10 and 11.

Table 6 Unit root tests—Harris–Tsavalis and Im–Pesaran–Shin

Income group	Harris–Tsavalis test				Im–Pesaran–Shin test			
	No trend		Trend		No trend		Trend	
	Level	1st Diff	Level	1st Diff	Level	1st Diff	Level	1st Diff
GDP per capita								
Global	2.8895	− 67.9178*	3.2301	− 31.0364*	3.4196	− 22.0645*	− 0.5355	− 24.5810*
Low-income	− 0.7496	− 42.2831*	− 4.7392*	− 35.03488	− 0.9404	− 19.7044*	− 2.6126*	− 29.4570*
Lower-middle	2.1482	− 34.2325*	2.5724	− 14.5883*	0.9209	− 15.69.14*	0.1376	− 12.5455*
Upper-middle	0.6764	− 24.8593*	3.5214	− 10.1930*	− 0.6687	− 11.2830*	0.8994	− 8.6396*
High income	− 0.0375	− 31.9341*	0.5922	− 15.1356*	− 1.4150	− 14.8075*	− 1.5540	− 11.4213*
Health expenditure per capita								
Global	0.1137	− 66.1171*	1.4599	− 31.7085*	3.0318	− 23.2369*	− 1.2840	− 26.2624*
Low-income	− 2.9585*	− 70.2381*	− 2.5183*	42.4236*	− 1.8999*	− 25.649*	− 2.7954*	35.4291*
Lower-middle	0.0316	− 37.0010*	0.8715	− 18.0001*	− 0.7643	− 18.8382*	− 1.5868	− 15.1472*
Upper-middle	0.7697	− 26.2176*	1.8804	− 11.4772*	− 1.5285	− 11.7303*	− 0.0679	− 9.0500*
High income	0.6497	− 31.1717*	3.0072	− 14.8730*	0.0853	− 13.7274*	0.9593	− 11.7642*

*Significant at 95% confidence interval to reject the Ho = unit root in the panel data. AIC lag selection criteria have been used. For the HT test, a default lag was used. For all the tests the command 'demean' was used which can mitigate the impact of cross-section dependence (Levin et al. 2002). For the IPS and HT unit root tests, H0 = panel contains unit root

 Springer

Table 7 Westerlund cointegration test per capita health expenditure and per capita GDP

Group	Group mean test (z value)				Panel test (z value)			
	G_t	Robust p value	G_a	Robust p value	P_t	Robust p value	P_a	Robust p value
Global	− 69.045* (0.000)	0.310	− 15.696* (0.000)	0.960	− 16.889* (0.000)	0.263	− 25.178* (0.000)	1.000
Low-income	− 30.925* (0.000)*	0.130	− 6.739* (0.000)	0.913	− 8.362* (0.000)	0.200	− 11.931* (0.000)	0.970
Lower-middle	− 42.793 (0.000)*	0.185	− 9.451* (0.000)	0.770	− 14.792* (0.000)	0.218	− 14.837* (0.000)	0.802
Upper-middle	− 42.138* (0.000)	0.143	− 7.670* (0.000)	0.243	− 8.727* (0.000)	0.672	− 13.959* (0.000)	0.520
High income	− 22.570* (0.000)	0.768	− 7.410* (0.000)	0.938	− 7.183* (0.000)	0.693	− 12.868* (0.000)	0.900

*Significant at 95% confidence interval to reject the H0 = no cointegration. The asymptotic p value in the parenthesis, and robust p value represents bootstrapped p values in the separate columns. In the test, G_t and G_a represent group means test and P_t and P_a represent panel statistics (Persyn and Westerlund 2008)

Table 8 Dumitrescu and Hurlin (2012) Granger non-causality test results; Z-bar tilde value

Group	Ho = growth in per capita GDP does not Granger cause growth in per capita health expenditure	Ho = growth in per capita health expenditure does not Granger cause growth in per capita GDP	Direction of causality
Lags (1)			
Global	7.0107*	2.5875*	Bi-directional
Low-income	11.7573*	9.1133*	Bi-directional
Lower-middle income	5.2986*	0.0252	Uni-directional
Upper-middle income	2.6208*	0.2636	Uni-directional
High income	2.6979*	3.3327*	Bi-directional
Lags (2)			
Global	5.3473*	2.1974*	Bi-directional
Low-income	9.1088*	7.5052*	Bi-directional
Lower-middle income	3.1283*	− 0.9896	Uni-directional
Upper-middle income	1.9888*	− 0.5231	Uni-directional
High income	3.1264*	4.0943*	Bi-directional
Lags (3)			
Global	7.5096*	5.96226*	Bi-directional
Low-income	7.4078*	9.5415*	Bi-directional
Lower-middle income	5.2866*	1.5589	Uni-directional
Upper-middle income	2.2522*	− 0.1738	Uni-directional
High income	2.9428*	4.6111*	Bi-directional

*Significant at 95% confidence interval to reject the Ho = the explanatory variable does not granger cause the explained variable. Here the H_1 is the explanatory variable does granger cause the explained variable for at least one panel VAR

Table 9 Panel vector auto regressive Granger causality Wald test, Chi square-value

Granger causality Wald test

Group	Ho = growth in per capita GDP does not Granger cause growth in per capita health expenditure	Ho = growth in per capita health expenditure does not Granger cause growth in per capita GDP	Direction of causality
Global	12.967*	3.935*	Bi-directional
Low-income	2.004	6.187*	Uni-directional
Lower-middle income	18.763*	3.511	Uni-directional
Upper-middle income	12.658*	2.643	Uni-directional
High income	22.539*	13.453*	Bi-directional

*Significant at 95% confidence interval to reject the Ho = excluded variable does not Granger cause equation variable. Here the H_1 is excluded variable does Granger cause equation variable

Table 10 VAR Granger causality/block exogeneity Wald Tests, Chi square-value

Group	Ho = growth in per capita health expenditure does not Granger cause growth in per capita GDP	Ho = growth in per capita health expenditure does not Granger cause growth in per capita GDP	Direction of causality
Toda–Yamamoto approach to Granger causality tests			
Global	100.94*	38.268*	Bi-directional
Low-income	0.841	11.887*	Uni-directional
Lower-middle income	41.883*	3.372	Uni-directional
Upper-middle income	19.849*	13.362	Uni-directional
High income	46.073*	60.899*	Bi-directional

*Significant at 95% confidence interval to reject the Ho = excluded variable does not Granger cause equation variable. Here the H_1 is excluded variable does Granger cause equation variable

Table 11 Cross-sectional dependence test

	p value
Global	
Pesaran's test of cross sectional independence = 27.168	
Average absolute value of the off-diagonal elements = 0.427	0.000
Low-income	
Pesaran's test of cross sectional independence = 4.312	
Average absolute value of the off-diagonal elements = 0.358	0.000
Lower-middle income	
Pesaran's test of cross sectional independence = 7.951	
Average absolute value of the off-diagonal elements = 0.410	0.000
Upper-middle income	
Pesaran's test of cross sectional independence = 10.526	
Average absolute value of the off-diagonal elements = 0.408	0.000
High-income	
Pesaran's test of cross sectional independence = 35.398	
Average absolute value of the off-diagonal elements = 0.477	0.000

The null hypothesis of the cross-section dependence test proposed by Pesaran (2004) is H0 = no cross-section dependence. The H0 is rejected for the global and all income levels. Hence, we conclude that the panel data are cross-sectional dependence in the disturbances

References

Abrigo, M. R., & Love, I. (2016). Estimation of panel vector autoregression in Stata. *Stata Journal, 16,* 778–804.

Acemoglu, D., & Johnson, S. (2007). Disease and development: The effect of life expectancy on economic growth. *Journal of Political Economy, 115*(6), 925–985.

Adriana, D. (2014). Revisiting the relationship between unemployment rates and shadow economy. A Toda–Yamamoto approach for the case of Romania. *Procedia Economics and Finance, 10,* 227–236.

Amiri, A., & Ventelou, B. (2012). Granger causality between total expenditure on health and GDP in OECD: Evidence from the Toda–Yamamoto approach. *Economics Letters, 116*(3), 541–544. https://doi.org/10.1016/j.econlet.2012.04.040.

Asteriou, D. (2009). Foreign aid and economic growth: New evidence from a panel data approach for five South Asian countries. *Journal of Policy Modeling, 31*(1), 155–161.

Atilgan, E., Kilic, D., & Ertugrul, H. M. (2016). The dynamic relationship between health expenditure and economic growth: Is the health-led growth hypothesis valid for Turkey? *European Journal of Health Economics, 18*(5), 567–574. https://doi.org/10.1007/s10198-016-0810-5.

Baltagi, B. H., Lagravinese, R., Moscone, F., & Tosetti, E. (2017). Health care expenditure and income: A global perspective. *Health Economics, 26*(7), 863–874.

Baltagi, B. H., & Moscone, F. (2010). Health care expenditure and income in the OECD reconsidered: Evidence from panel data. *Economic Modelling, 27*(4), 804–811. https://doi.org/10.1016/j.econmod.2009.12.001.

Bloom, D. E., Canning, D., & Sevilla, J. (2004). The effect of health on economic growth: A production function approach. *World Development, 32*(1), 1–13.

Carrion-i-Silvestre, J. L. (2005). Health care expenditure and GDP: Are they broken stationary? *Journal of Health Economics, 24*(5), 839–854. https://doi.org/10.1016/j.jhealeco.2005.01.001.

Chen, W., Clarke, J. A., & Roy, N. (2013). Health and wealth: Short panel Granger causality tests for developing countries. *The Journal of International Trade and Economic Development, 23*(6), 755–784. https://doi.org/10.1080/09638199.2013.783093.

Chudik, A., Pesaran, M. H., & Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal, 14*(1), C45–C90. https://doi.org/10.1111/j.1368-423X.2010.00330.x.

Clarke, J. A., & Mirza, S. (2006). A comparison of some common methods for detecting Granger noncausality. *Journal of Statistical Computation and Simulation, 76*(3), 207–231.

Clemente, J., Marcuello, C., Montañés, A., & Pueyo, F. (2004). On the international stability of health care expenditure functions: Are government and private functions similar? *Journal of Health Economics, 23*(3), 589–613.

Dolado, J. J., & Lütkepohl, H. (1996). Making Wald tests work for cointegrated VAR systems. *Econometric Reviews, 15*(4), 369–386.

Dumitrescu, E. I., & Hurlin, C. (2012). Testing for Granger non-causality in heterogeneous panels. *Economic Modelling, 29*(4), 1450–1460.

Erdil, E., & Yetkiner, I. H. (2009). The Granger-causality between health care expenditure and output: A panel data approach. *Applied Economics, 41*(4), 511–518. https://doi.org/10.1080/00036840601019083.

Everaert, G., & De Groote, T. (2016). Common correlated effects estimation of dynamic panels with cross-sectional dependence. *Econometric Reviews, 35*(3), 428–463.

Farag, M., Nandakumar, A., Wallack, S., Hodgkin, D., Gaumer, G., & Erbil, C. (2013). Health expenditures, health outcomes and the role of good governance. *International Journal of Health Care Finance and Economics, 13*(1), 33–52.

Gengenbach, C., Palm, F. C., & Urbain, J. P. (2006). Cointegration testing in panels with common factors. *Oxford Bulletin of Economics and Statistics, 68*(1), 683–719.

Glied, S., & Smith, P. C. (2011). *The Oxford handbook of health economics*. Oxford: Oxford University Press.

Granados, J. A. T. (2012). Economic growth and health progress in England and Wales: 160 years of a changing relation. *Social Science and Medicine, 74*(5), 688–695.

Halici-Tuluce, N. S., Dogan, I., & Dumrul, C. (2016). Is income relevant for health expenditure and economic growth nexus? *International Journal of Health Economics and Management, 16*(1), 23–49. https://doi.org/10.1007/s10754-015-9179-8.

Hall, S. G., & Jones, C. I. (2007). The value of life and the rise in health spending. *The Quarterly Journal of Economics, 122*(2007), 39–72.

Hall, S. G., Swamy, P. A. V. B., & Tavlas, G. S. (2011). Generalized cointegration: A new concept with an application to health expenditure and health outcomes. *Empirical Economics, 42*(2), 603–618. https://doi.org/10.1007/s00181-011-0483-y.

Hansen, P., & King, A. (1996). The determinants of health care expenditure: A cointegration approach. *Journal of Health Economics, 15,* 127–137.

Harris, R. D., & Tzavalis, E. (1999). Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics, 91*(2), 201–226.

Hartwig, J. (2008). What drives health care expenditure? Baumol's model of 'Unbalanced growth' revisited. *Journal of Health Economics, 27,* 603–623.

Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics, 115*(1), 53–74.

Kapetanios, G., Pesaran, M. H., & Yamagata, T. (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics, 160*(2), 326–348.

Ke, X., Saksena, P., & Holly, A. (2011). The determinants of health expenditure: A country-level panel data analysis. *Working paper of the Results for Development Institute (R4D)*. Geneva: World Health Organization. www.resultsfordevelopment.org. Accessed on August 11th, 2017.

Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics, 74*(1), 119–147.

Lago-Peñas, S., Cantarero-Prieto, D., & Blázquez-Fernández, C. (2013). On the relationship between GDP and health care expenditure: A new look. *Economic Modelling, 32,* 124–129. https://doi.org/10.1016/j.econmod.2013.01.021.

Levin, A., Lin, C. F., & Chu, C. S. J. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics, 108*(1), 1–24.

Liddle, B., & Messinis, G. (2015). Which comes first-urbanization or economic growth? Evidence from heterogeneous panel causality tests. *Applied Economics Letters, 22*(5), 349–355.

Lütkepohl, H., & Krätzig, M. (2004). *Applied time series econometrics*. Cambridge: Cambridge University Press.

MacDonald, G., & Hopkins, S. (2002). Unit root properties of OECD health care expenditure and GDP data. *Health Economics, 11*(4), 371–376. https://doi.org/10.1002/hec.657.

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.

Menard, A.-R., & Weill, L. (2016). Understanding the link between aid and corruption: A causality analysis. *Economic Systems, 40*(2), 260–272.

Mladenović, I., Milovančević, M., Sokolov Mladenović, S., Marjanović, V., & Petković, B. (2016). Analyzing and management of health care expenditure and gross domestic product (GDP) growth rate by adaptive neuro-fuzzy technique. *Computers in Human Behavior, 64,* 524–530. https://doi.org/10.1016/j.chb.2016.07.052.

Moscone, F., & Tosetti, E. (2010). Health expenditure and income in the United States. *Health Economics, 19*(12), 1385–1403. https://doi.org/10.1002/hec.1552.

Okunade, A. A., & Karakus, M. C. (2001). Unit root and cointegration tests: Timeseries versus panel estimates for international health expenditure models. *Applied Economics, 33*(9), 1131–1137. https://doi.org/10.1080/00036840122612.

Panopoulou, E., & Pantelidis, T. (2012). Convergence in per capita health expenditures and health outcomes in the OECD countries. *Applied Economics, 44*(30), 3909–3920. https://doi.org/10.1080/00036846.2011.583222.

Persyn, D., & Westerlund, J. (2008). Error-correction-based cointegration tests for panel data. *Stata Journal, 8*(2), 232–241.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. *CESifo GmbH, CESifo working paper series: CESifo Working Paper No. 1229.* http://ezproxy.usq.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ecn&AN=0906171&site=ehost-live. http://www.cesifo.de/DocCIDL/1229.pdf. Accessed January 14th, 2018.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica, 74*(4), 967–1012.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics, 22*(2), 265–312.

Pesaran, M. H., & Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics, 161*(2), 182–202.

Rafiq, S., Salim, R., & Bloch, H. (2009). Impact of crude oil price volatility on economic activities: An empirical investigation in the Thai economy. *Resources Policy, 34*(3), 121–132.

Self, S., & Grabowski, R. (2003). How effective is public health expenditure in improving overall health? A cross-country analysis. *Applied Economics, 35*(7), 835–845. https://doi.org/10.1080/0003684032000056751.

Shahbaz, M. (2012). Does trade openness affect long run growth? Cointegration, causality and forecast error variance decomposition tests for Pakistan. *Economic Modelling, 29*(6), 2325–2339.

Shaw, J. W., Horrace, W. C., & Voge, R. J. (2005). The determinants of life expectancy: An analysis of the OECD health data. *Southern Economic Journal, 71*(4), 768–783.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, *48*(1), 1–48.

Swanson, N. R., & Granger, C. W. (1997). Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions. *Journal of the American Statistical Association, 92*(437), 357–367.

Tamakoshi, T., & Hamori, S. (2015). Testing cointegration between health care expenditure and GDP in Japan with the presence of a regime shift. *Applied Economics Letters, 23*(2), 151–155. https://doi.org/10.1080/13504851.2015.1058901.

Toda, H. Y., & Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics, 66*(1), 225–250.

van der Gaag, J., & Stimac, V. (2008). Towards a new paradigm for health sector development. *Amsterdam: Amsterdam Institute for International Development*. https://www.r4d.org/wpcontent/uploads/Toward-a-New-Paradigm-for-Health-Sector-Development.pdf. Accessed on November 7th, 2017.

Villaverde, J., Maza, A., & Hierro, M. (2014). Health care expenditure disparities in the European Union and underlying factors: A distribution dynamics approach. *International Journal of Health Care Finance and Economics, 14*(3), 251–268.

Wang, K. M. (2011). Health care expenditure and economic growth: Quantile panel-type analysis. *Economic Modelling, 28*(4), 1536–1549. https://doi.org/10.1016/j.econmod.2011.02.008.

Wang, Z. (2009). The determinants of health expenditures: Evidence from US state-level data. *Applied Economics, 41*(4), 429–435. https://doi.org/10.1080/00036840701704527.

Wang, Z., & Rettenmaier, A. J. (2007). A note on cointegration of health expenditures and income. *Health Economics, 16*(6), 559–578. https://doi.org/10.1002/hec.1182.

Westerlund, J. (2007). Testing for error correction in panel data. *Oxford Bulletin of Economics and Statistics, 69*(6), 709–748.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. Cambridge: MA: MIT Press.

World Bank. (2016). World development indicators. http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators. Accessed on May 21st, 2017.

World Health Organization. (2016). Global health observatory data repository. http://apps.who.int/gho/data/view.main.healthexpratioglobaL?lang=en. Accessed on May 21st, 2017.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.